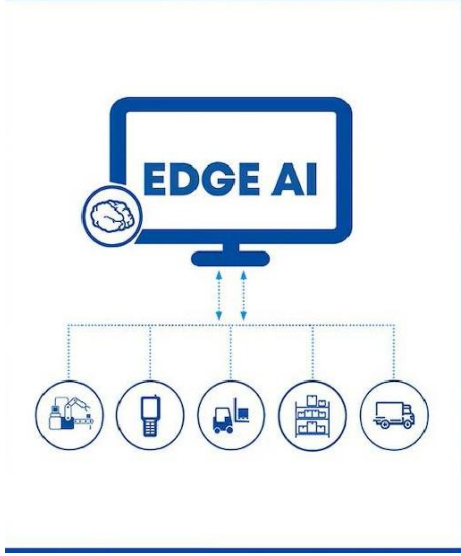
Edge AI (IA en el borde) para el Monitoreo en Tiempo Real como Asistente en la Participación Estudiantil en Aulas Híbridas: Optimización de Modelos Ligeros

Morales Ccasa, Geyson David
Grupo 22

Proyecto de Investigación I
2025-1

Resumen del Tema de investigación

La "Edge AI para el Monitoreo en Tiempo Real como asistente en la Participación Estudiantil en Aulas Híbridas: Optimización de Modelos Ligeros" aborda la necesidad crítica de integrar inteligencia artificial en entornos educativos con restricciones de conectividad y recursos. Se verán modelos de *Machine Learning* y *Deep Learning* optimizados para dispositivos de borde (*Edge AI*), capaces de operar en tiempo real sin dependencia de la nube. Existen distintas aplicaciones tradicionales (como monitoreo de salud, manufactura, vehículos autónomos, ciudades y hogares inteligentes, seguridad,etc), aquí se aplicará en la **detección de participación estudiantil** mediante análisis multimodal (voz, textos, expresiones faciales y postura), adaptándose a contextos híbridos (presenciales/virtuales) y posiblemente aplicable a zonas con infraestructura limitada (ej: áreas rurales del Perú). En la investigación se van a explorar técnicas avanzadas de optimización para reducir la carga computacional. Además, propone un marco ético para el uso de datos sensibles en educación, asegurando privacidad y cumplimiento normativo. Este enfoque no solo democratiza el *acceso a tecnologías educativas de vanguardia, sino que también ofrece una solución escalable para mejorar la retroalimentación docente y personalizar el aprendizaje en tiempo real*



Problema: Monitoreo de la participación estudiantil.
Técnica: Modelos de IA en el Borde (Edge AI), Optimización de Modelos Ligeros
Ámbito: Aulas Híbridas (presenciales/virtuales), Dispositivos Multifuncionales, escenarios con infraestructura limitada.
Aporte: Optimización de Modelos Ligeros para Dispositivos Multifuncionales

Problemática

La **falta de herramientas tecnológicas accesibles y eficientes para monitorear la participación estudiantil en aulas híbridas (presenciales/virtuales)** en escenarios con infraestructura limitada, lo que impide una retroalimentación docente en tiempo real y personalización del aprendizaje, debido a:

1. **Dependencia crítica de soluciones cloud**, que requieren conectividad estable y costosa, inviable en algunos escenarios o regiones.

1. **Modelos de IA no optimizados** para hardware de bajo costo, generando latencias inaceptables.

2. **Falta de enfoques éticos** en el manejo de datos sensibles (ej: expresiones faciales, voz), generando riesgos de privacidad y desconfianza institucional.

3. **Limitaciones en análisis multimodal**, donde sistemas tradicionales priorizan un solo tipo de dato (ej: video), ignorando la integración de voz, texto y postura para una evaluación integral.

Problema

Definición Ontológica: Las fuentes no proporcionan una definición ontológica explícita para el objeto de estudio específico de este proyecto. Aunque una fuente sugiere cómo obtener una definición ontológica para otro concepto (pishing), no se ofrece esta definición para el grupo de datos de sesiones híbridas o la participación estudiantil.

El **objeto de estudio real** se manifiesta en el comportamiento en el ~~medio ambiente natural de las~~ aulas híbridas de docentes y estudiantes. Esto incluye tanto los entornos presenciales como los virtuales o en línea. La situación que se desea analizar, la participación estudiantil, ocurre durante las sesiones de clase, utilizando herramientas de videoconferencia como Zoom o WebEx, o interactuando con plataformas LMS. Ejemplos concretos del objeto de estudio real son los flujos de video, audio y texto generados por los estudiantes y docentes durante una lección.

El **objeto de estudio modelado** se refiere a cómo este objeto real es representado para ser tratado digitalmente. Dado que el objeto de estudio son datos digitales (imágenes, videos, audio, textos) provenientes de sesiones en un computador, ya está en una forma representada. Para el análisis, estos datos se modelan mediante la extracción de características relevantes utilizando sensores (virtuales, como las cámaras o micrófonos de los dispositivos) y algoritmos de IA. La representación puede tomar la forma de características faciales (como relaciones de aspecto de ojos y boca), puntos clave (esqueleto) del cuerpo, o el conteo y coordenadas de objetos detectados (caras, manos, teléfonos).

El problema se mide y cuantifica a través de la inferencia de estados de participación o comportamientos a partir de los datos modelados. Esto puede implicar:

- ✓ La detección de niveles de compromiso, emoción o concentración.
- ✓ La identificación de clusters de estudiantes (e.g., altamente comprometidos vs. desinteresados).
- ✓ La generación de alertas si la participación cae por debajo de un umbral definido (e.g., < 50%).
- ✓ La evaluación de la atención del estudiante utilizando métricas derivadas del análisis de características. La precisión en la detección de la atención puede cuantificarse utilizando métricas de evaluación de modelos de clasificación como precisión, recall, F1-score y exactitud. Un ejemplo de rendimiento reportado es una precisión del 99.71% en la evaluación de niveles de atención

Problema (Cont.)

Algunos otros aspectos:

1.**Modalidades**: El análisis es multimodal, integrando voz, textos, expresiones faciales y postura. Esto implica el procesamiento de datos visuales (cámaras), auditivos (micrófonos) y contextuales (interacción con plataformas LMS).

2.**Características**: El objeto de estudio exhibe características que permiten inferir el estado y la participación del estudiante, tales como expresiones faciales, movimientos oculares, postura de la cabeza, rotación en ejes, cantidad de manos detectadas, o detección de teléfono móvil.

3.**Estados**: La participación puede manifestarse en diferentes estados o niveles, como altamente comprometido, desinteresado, o fatigado, Los estados afectivos, como la fatiga o la confusión, también forman parte de estos aspectos subyacentes.

4.**Dinámica** y Ciclo de Vida: La dinámica del fenómeno ocurre a lo largo de una sesión de clase, donde el monitoreo se realiza en tiempo real y de forma continua. El proceso implica la captura continua de datos, la detección de características, su análisis y la generación de información

5.**Taxonomías**: Aunque no se presenta una taxonomía formal del objeto de estudio per se, las fuentes implican una categorización de los datos y características analizadas (e.g., por modalidad: visual, auditivo, textual; por tipo de comportamiento: facial, postural, manipulación de objetos). En trabajos relacionados, se tipifican comportamientos no verbales (e.g., en docentes: gestos, movimientos de cabeza, expresiones faciales, posición espacial) que son pertinentes para el análisis de participación.

Objetivos

Técnico

- Desarrollar un modelo multimodal (voz o texto o imagen o video).
- Implementar cuantización sin pérdida significativa de precisión.

Educativo

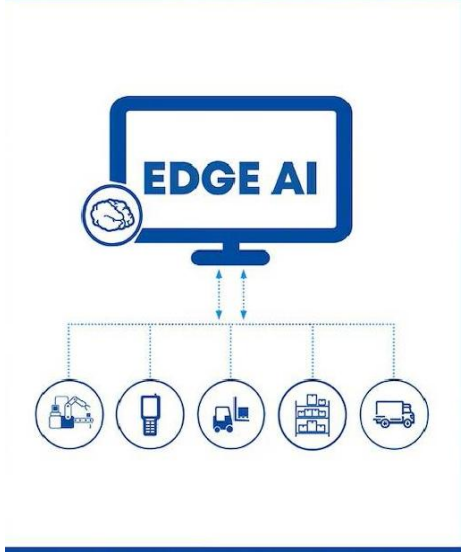
- Generar alertas automáticas para cuando la participación disminuya (por ejem. < 50%).
- Adaptar contenido según el estado emocional detectado (ej: fatiga, confusión).

Social

- Reducir la brecha tecnológica en sesiones de clases bajo dificultades técnicas o de recursos.

Concepto

La **Edge AI** (*IA de borde*) es un campo emergente que combina la inteligencia artificial con la computación de borde, lo que permite el procesamiento de IA directamente en dispositivos locales de borde. Facilita el procesamiento y análisis de datos en tiempo real sin depender constantemente de la infraestructura en la nube. Este enfoque ha cobrado impulso a medida que el panorama tecnológico experimenta una importante transformación, impulsada por el aumento exponencial de los datos generados en el borde, gracias a los dispositivos IoT, sensores y otras tecnologías conectadas.



Comportamiento y tipificación de un comportamiento:

Comportamientos faciales: Detección de caras, reconocimiento facial para autenticación, extracción de características faciales, análisis de emociones ("neutral", "fear", "disgust", "angry", "surprised", "happy", "sad") [53, 63, 78, 92, 96, RB0101], estimación de la mirada (eye gaze) [63, 69, 71, 79, 83, 96, RB0101], análisis de relaciones de aspecto de ojos y boca (EAR, MAR) como indicadores de atención/fatiga [31, 173, 181, 183, 189, 190, 194, 198, RB0104].

▪ **Comportamientos posturales/corporales:** Estimación de pose (esqueleto, puntos clave del cuerpo) [34, 215, 222, 247, 256, 257, 335, RB0106, RB0107], postura de la cabeza [23, 219, 222, 247, 248, 256, 318, RB0106].

▪ Comportamientos de interacción con objetos: Detección de manos [34, 215, 222, 246, 258, RB0106], detección de teléfono móvil [34, 23, 215, 222, 246, 255, 258, RB0106].

◦ Se puede tipificar estos comportamientos como indicadores de diferentes estados de participación/atención/afectivos (ej. enfocado, distraído, comprometido, desinteresado, fatigado) [53, 86, 215, 222, RB0101, RB0104, RB0106]. La fuente [RB0107] proporciona un ejemplo de categorización de comportamientos no verbales (aunque para profesores), que podría servir de inspiración para una tipificación más detallada de los comportamientos de los estudiantes si fuera necesario

.

String de Búsqueda

("Edge AI" OR "On-Device Machine Learning" OR "Edge Computing") AND ("Real-Time Monitoring" OR "Student Engagement" OR "Classroom Analytics" OR "Learner Attention" OR "Student Participation" OR "Student Attentiveness") AND ("Hybrid Classrooms" OR "Hybrid Learning" OR "Low-Resource Settings“) AND ("Model Optimization" OR "Lightweight Models" OR "Efficient Neural Networks" OR "Model Compression")

Los conceptos clave de tu investigación son:

1. IA en el Borde (Edge AI)
2. Monitoreo en Tiempo Real
3. Participación Estudiantil / Engagement
4. Aulas Híbridas / Entornos de Bajos Recursos
5. Optimización / Modelos Ligeros

Papers de investigación

- **A Real-Time Learning Analytics Dashboard for Automatic Detection of Online Learner's Affective States**: Este artículo describe un panel de análisis de aprendizaje en tiempo real diseñado para la detección automática de estados afectivos de estudiantes en línea. Emplea la Red Neuronal Convolucional en Cascada Multitarea (MTCNN) para la detección facial.

- **Challenges and Applications of LLM**: Es un artículo de revisión que aborda los principales desafíos y aplicaciones de los Grandes Modelos de Lenguaje (LLMs). Discute problemas como los costos de pre-entrenamiento y la longitud limitada del contexto, y explora sus aplicaciones en diversos campos como el derecho, la medicina, educación y las ciencias sociales.

- **Edge-Cloud Architectures for Hybrid Energy Management Systems: A Comprehensive Review**: Es una revisión exhaustiva de las arquitecturas edge-cloud para Sistemas de Gestión de Energía Híbrida (HEMS). Analiza en detalle una arquitectura de cinco capas, discute métricas de rendimiento y métodos de soporte para la adopción, y revisa arquitecturas de tres y cuatro capas, así como plataformas IoT en la nube.

- **Modelos Tecnológicos de Aprendizaje Adaptativo Aplicados a la Educación**: Es un estudio de revisión sistemática sobre modelos tecnológicos de aprendizaje adaptativo aplicados a la educación. Analizó 50 estudios publicados entre 2012 y 2021, identificando que la mayoría de los modelos se estructuran en torno a la inteligencia artificial, el objeto de aprendizaje y la técnica de adaptabilidad.

- **Optimizing student engagement in edge-based online learning with advanced analytics**: Propone un sistema de análisis de atención del estudiante basado en Edge (borde - EBSAAS). Este sistema utiliza la detección facial mediante el algoritmo MTCNN y el reconocimiento de atención extrayendo características faciales (como las relaciones de aspecto de ojos y boca con DLIB) e ingresándolas a una red VGG-19. Está diseñado para conectarse a dispositivos edge y a un servidor edge.

1. An adaptive teacher-student learning algorithm with decomposed knowledge distillation for on-edge intelligence

Planteamiento del Problema:

Objeto de Estudio: Algoritmos adaptativos de aprendizaje teacher-student con enfoque en la inteligencia en el borde (on-edge intelligence).

Problema Abordado: Optimizar el rendimiento de modelos de Machine Learning en dispositivos de borde con recursos limitados, lo cual a menudo implica transferir conocimiento (Knowledge Distillation - KD) de modelos grandes ("teacher") a modelos más pequeños y eficientes ("student")

Abordaje del Problema y Método/Técnica:
Se introduce un algoritmo de aprendizaje teacher-student adaptativo que utiliza destilación de conocimiento descompuesta.
El método se evalúa comparando el rendimiento de un modelo Teacher (ResNet Teacher) con un modelo Student (ResNet Student).
La evaluación incluye comparativas con métodos de destilación de conocimiento de vanguardia (state-of-the-art KD methods) para demostrar su efectividad.
Los resultados mostrados en la tabla de la fuente indican que la técnica se aplica a tareas de clasificación, específicamente mencionando categorías como ductal_carcinoma y lobular_carcinoma, sugiriendo una aplicación en el ámbito médico

sensors MDPI

Article

A Real-Time Learning Analytics Dashboard for Automatic Detection of Online Learners' Affective States

Mohammad Nehal Hasnine [1,*], Ho Tan Nguyen [1], Thuy Thi Thu Tran [1], Huyen T. T. Bui [1], Gökhan Akçapınar [2] and Hiroshi Ueda [1]

[1] Research Center for Computing and Multimedia Studies, Hosei University, Tokyo 102-8160, Japan; ho.nguyen.56@hosei.ac.jp (H.T.N.)
[2] Department of Computer Education and Instructional Technology, Hacettepe University, 06230 Ankara, Türkiye
* Correspondence: nehal.hasnine.79@hosei.ac.jp; Tel.: +81-(0)42-387-6070

Abstract: Students' affective states describe their engagement, concentration, attitude, motivation, happiness, sadness, frustration, off-task behavior, and confusion level in learning. In online learning, students' affective states are determinative of the learning quality. However, measuring various affective states and what influences them is exceedingly challenging for the lecturer without having real interaction with the students. Existing studies primarily use self-reported data to understand students' affective states, while this paper presents a novel learning analytics system called MOEMO (Motion and Emotion) that could measure online learners' affective states of engagement and concentration using emotion data. Therefore, the novelty of this research is to visualize online learners' affective states on lecturers' screens in real-time using an automated emotion detection process. In real-time and offline, the system extracts emotion data by analyzing facial features from the lecture videos captured by the typical built-in web camera of a laptop computer. The system determines online learners' five types of engagement ("strong engagement", "high engagement", "medium engagement", "low engagement", and "disengagement") and two types of concentration levels ("focused" and "distracted"). Furthermore, the dashboard is designed to provide insight into students' emotional states, the clusters of engaged and disengaged students', assistance with intervention, create an after-class summary report, and configure the automation parameters to adapt to the study environment.

Keywords: AI in education; affective states detection; dashboard; learning analytics framework; emotion; lecture video analysis

check for updates

Citation: Hasnine, M.N.; Nguyen, H.T.; Tran, T.T.T.; Bui, H.T.T.; Akçapınar, G.; Ueda, H. A Real-Time Learning Analytics Dashboard for Automatic Detection of Online Learners' Affective States. *Sensors* **2023**, *23*, 4243. https://doi.org/10.3390/s23094243

Academic Editor: Zahir M. Hussain

Received: 2 March 2023
Revised: 18 April 2023
Accepted: 18 April 2023
Published: 24 April 2023

Copyright © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://

1. Introduction

With the rapid growth of learning technologies, online learning, or virtual education, has become popular. It opened new doors to learners with physical challenges or those who were busy caring for a family that prevented them from learning at school [1]. E-learning has become a frontline approach to supporting education during a crisis such as the COVID–19 pandemic. It is because online learning systems are convenient for students, cost-efficient, flexible ways of learning, scalable, allow for better repetition, and have a higher degree of freedom [2]. With a compelling e-learning system and a highly motivated student, one can achieve great success in a short period of time [2]. On the contrary, online learning environments have several drawbacks, including the fact that they lack the face-to-face interactions that students would receive in traditional classroom settings, and real-time interaction can be frustrating. In addition, it is hard for the lecturer to understand the students' affective states, such as whether they are feeling confused, motivated to learn, happy with the lesson, taking the lecture with enthusiasm, engaged with the learning

2. Paper: Attention monitoring of students during online classes using XGBoost classifier

Problemática:

Un desafío significativo en la educación online es la evaluación efectiva del involucramiento, el compromiso y la atención de los estudiantes durante las clases virtuales.

La falta de interacción presencial y cara a cara en las aulas virtuales presenta desafíos especiales en tiempo real.

Los sistemas de monitoreo de atención existentes a menudo tienen un enfoque limitado, pudiendo pasar por alto la totalidad del involucramiento estudiantil más allá de las señales visuales (como movimientos de ojos o tiempo frente a la pantalla). No integran aspectos como el análisis de audio o la participación activa

Objetivo:

Introducir un sistema especializado para identificar y comprender la compleja gama de comportamientos estudiantiles durante las clases online y monitorear sus niveles de atención.

Buscar emular las clases convencionales presenciales, garantizando una ejecución fluida y fiable.

Abordar las limitaciones de los sistemas actuales mediante un enfoque multimodal

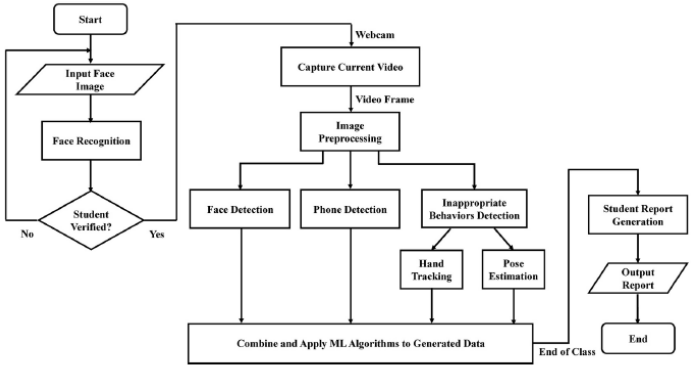


Fig. 1. The workflow diagram of the online class monitoring system.

Computers and Education: Artificial Intelligence 5 (2023) https://doi.org/10.1016/j.caeai.2023.100191

2. Paper: Attention monitoring of students during online classes using XGBoost classifier

Estrategia de Solución (Sistema Propuesto - Multimodal):

Utiliza diversas técnicas para detectar una amplia gama de comportamientos estudiantiles, incluye autenticación de usuario a través de reconocimiento facial e integra componentes clave: detección facial, seguimiento de manos, detección de teléfonos móviles y estimación de pose. (Estos módulos fundamentales generan características de alto nivel). Estas características se combinan para crear un dataset.

Se entrena un clasificador de Machine Learning (seleccionando **XGBoost** por su alta precisión) para ***determinar si los estudiantes están atentos***.

El procesamiento de video ocurre directamente en las computadoras individuales de los estudiantes, ofreciendo ventajas en privacidad y optimización del ancho de banda.

Todo se integra en una interfaz web (desarrollada con Django) para profesores y estudiantes

Objeto de Estudio

El comportamiento de los estudiantes durante las clases online, manifestado a través de señales visuales como expresiones faciales, movimientos de manos, pose corporal y el uso de dispositivos (teléfono móvil). Los datos generados por los módulos de detección de comportamiento y las características extraídas para el entrenamiento del modelo.

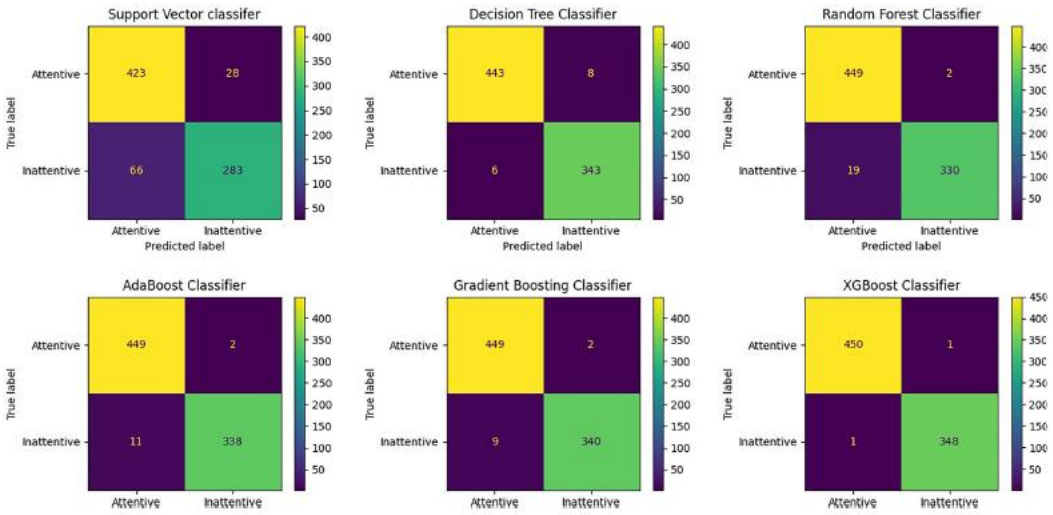


Fig. 14. Confusion matrices of machine learning classifiers on test data.

Contribuciones:

Desarrollo de un sistema de monitoreo de atención basado en técnicas de machine learning.
Y que permite la identificación de los comportamientos disruptivos más comunes en clases online mediante la integración de varios módulos.

Logro de una alta precisión (99.71%) con el método XGBoost para la evaluación de los niveles de atención.

Generación de reportes anónimos y completos sobre el nivel de atención de los estudiantes para los profesores

Computers and Education: Artificial Intelligence 5 (2023) https://doi.org/10.1016/j.caeai.2023.100191

3. Paper: Graph convolutional network for automatic detection of teachers' nonverbal behavior

Problemática:

La evaluación precisa de los comportamientos no verbales de los profesores es esencial para su desarrollo profesional. Sin embargo, los métodos manuales tradicionales son muy laboriosos y consumen mucho tiempo.

Existe investigación limitada sobre la detección automatizada y precisa en videos de aula. Además, hay una falta de marcos estandarizados para categorizar estos comportamientos y escasez de conjuntos de datos públicos.

Los métodos tradicionales de machine learning requieren extracción manual de características y los métodos que analizan imágenes directamente pueden verse afectados por el ruido de fondo en aulas complejas

Objetivo:

Investigar y proponer métodos para la detección automática de los comportamientos no verbales de los profesores, abordando las limitaciones existentes. Se buscan mejorar la precisión y fiabilidad de la detección automatizada.

Fig. 7. Samples of real classroom dataset.

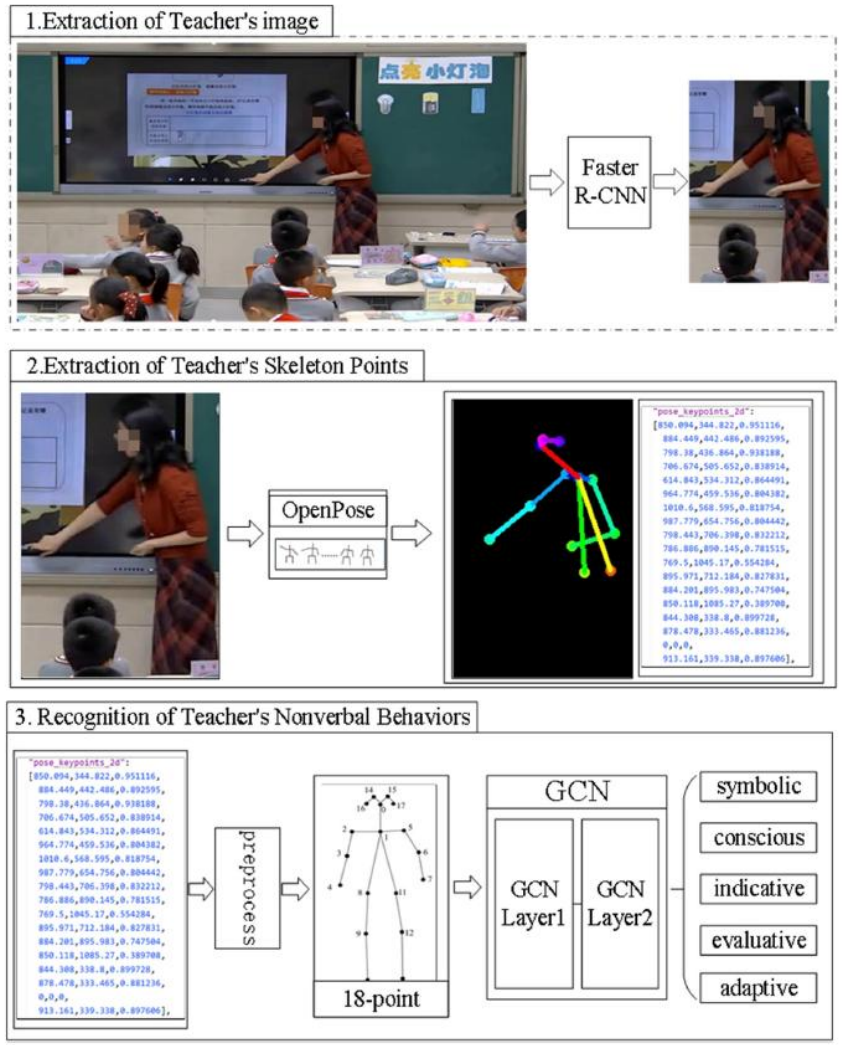


Fig. 1. Flowchart of teachers' nonverbal behavior detection

Computers and Education: Artificial Intelligence 5 (2023) https://doi.org/10.1016/j.caeai.2023.100174

3. Paper: Graph convolutional network for automatic detection of teachers' nonverbal behavior

Objeto de Estudio

El **comportamiento no verbal de los profesores en videos de aula.**
Se centra en los comportamientos visualmente discernibles, excluyendo expresiones vocales o distancia espacial.

Se definen cinco categorías comunes:
- acción simbólica
- acción consciente
- acción indicativa
- acción evaluativa
- acción adaptativa

El objeto de estudio real (el comportamiento del profesor) *se representa digitalmente mediante la extracción de puntos esqueléticos del cuerpo* del profesor a partir de las imágenes y las relaciones entre estos puntos.

Metodología Clave:

Un enfoque multi-paso basado en técnicas de visión por computadora y Deep Learning:

Detección del Profesor: Se usa el algoritmo Faster R-CNN para localizar la posición del profesor y extraer su imagen del video.

Extracción de Puntos Esqueléticos: Se aplica el algoritmo OpenPose a la imagen del profesor para extraer las coordenadas de sus puntos esqueléticos (representación de las articulaciones) Esto ayuda a eliminar el ruido de fondo,.

Clasificación del Comportamiento: Se construye un modelo de **Graph Convolutional Network (GCN)** que utiliza las coordenadas de los puntos esqueléticos y sus conexiones espaciales para clasificar el comportamiento no verbal. La novedad reside en el diseño de la matriz de adyacencia del GCN, que distingue entre conexiones físicas (articulaciones directas) y conexiones lógicas (importantes para el comportamiento pero no físicamente conectadas),,,,.

Datasets: Se construyeron dos datasets específicos (*simulado y real) con videos de profesores*, anotando la ubicación del profesor

Contribuciones y Resultados:

Propone un enfoque para la detección automática del comportamiento no verbal del profesorado usando GCN. Logra una alta precisión. El uso de puntos esqueléticos y GCN con matriz de adyacencia optimizada mitiga la interferencia del fondo y representa mejor las características de la acción en comparación con métodos previos.
El sistema puede ser una herramienta valiosa para la reflexión y el desarrollo profesional de los profesores al automatizar el análisis de video. Se inserta en el campo de la IA aplicada a la educación y el análisis de comportamiento

Fig. 6. Samples of simulated classroom dataset.

Computers and Education: Artificial Intelligence 5 (2023) https://doi.org/10.1016/j.caeai.2023.100174

sensors

MDPI

Article

A Real-Time Learning Analytics Dashboard for Automatic Detection of Online Learners' Affective States

Mohammad Nehal Hasnine [1,*], Ho Tan Nguyen [1], Thuy Thi Thu Tran [1], Huyen T. T. Bui [1], Gökhan Akçapınar [2] and Hiroshi Ueda [1]

[1] Research Center for Computing and Multimedia Studies, Hosei University, Tokyo 102-8160, Japan; ho.nguyen.56@hosei.ac.jp (H.T.N.)
[2] Department of Computer Education and Instructional Technology, Hacettepe University, 06230 Ankara, Türkiye
* Correspondence: nehal.hasnine.79@hosei.ac.jp; Tel.: +81-(0)42-387-6070

Abstract: Students' affective states describe their engagement, concentration, attitude, motivation, happiness, sadness, frustration, off-task behavior, and confusion level in learning. In online learning, students' affective states are determinative of the learning quality. However, measuring various affective states and what influences them is exceedingly challenging for the lecturer without having real interaction with the students. Existing studies primarily use self-reported data to understand students' affective states, while this paper presents a novel learning analytics system called MOEMO (Motion and Emotion) that could measure online learners' affective states of engagement and concentration using emotion data. Therefore, the novelty of this research is to visualize online learners' affective states on lecturers' screens in real-time using an automated emotion detection process. In real-time and offline, the system extracts emotion data by analyzing facial features from the lecture videos captured by the typical built-in web camera of a laptop computer. The system determines online learners' five types of engagement ("strong engagement", "high engagement", "medium engagement", "low engagement", and "disengagement") and two types of concentration levels ("focused" and "distracted"). Furthermore, the dashboard is designed to provide insight into students' emotional states, the clusters of engaged and disengaged students', assistance with intervention, create an after-class summary report, and configure the automation parameters to adapt to the study environment.

Keywords: AI in education; affective states detection; dashboard; learning analytics framework; emotion; lecture video analysis

check for updates

Citation: Hasnine, M.N.; Nguyen, H.T.; Tran, T.T.T.; Bui, H.T.T.; Akçapınar, G.; Ueda, H. A Real-Time Learning Analytics Dashboard for Automatic Detection of Online Learners' Affective States. *Sensors* **2023**, *23*, 4243. https://doi.org/10.3390/s23094243

Academic Editor: Zahir M. Hussain

Received: 2 March 2023
Revised: 18 April 2023
Accepted: 18 April 2023
Published: 24 April 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the rapid growth of learning technologies, online learning, or virtual education, has become popular. It opened new doors to learners with physical challenges or those who were busy caring for a family that prevented them from learning at school [1]. E-learning has become a frontline approach to supporting education during a crisis such as the COVID–19 pandemic. It is because online learning systems are convenient for students, cost-efficient, flexible ways of learning, scalable, allow for better repetition, and have a higher degree of freedom [2]. With a compelling e-learning system and a highly motivated student, one can achieve great success in a short period of time [2]. On the contrary, online learning environments have several drawbacks, including the fact that they lack the face-to-face interactions that students would receive in traditional classroom settings, and real-time interaction can be frustrating. In addition, it is hard for the lecturer to understand the students' affective states, such as whether they are feeling confused, motivated to learn, happy with the lesson, taking the lecture with enthusiasm, engaged with the learning materials, or how well the lecture is delivered.

In online learning, the problem of student disengagement and poor concentration is gaining attention. In the classroom, students' engagements consist of their behavioral, cognitive, and emotional elements [3]. For example, student academic effort, persistence, attention, concentration, and a lack of conduct problems are associated with behavioral engagement; thoughtfulness and willingness to make the efforts necessary to understand complex ideas and master difficult skills are related to cognitive engagement; and the presence of interest, enthusiasm, the absence of anger, anxiety, and boredom are connected with emotional engagement [3]. It is also evident that students' engagement is directly proportional to their achievement [4]. Due to poor concentration and a low level of engagement during the lecture, many students fail to achieve the learning goal. This also raises the concern of not receiving a high-quality education. This is because it is challenging for an instructor to keep a close eye on the entire class and regulate teaching. Therefore, for an educational system, it is essential to monitor students' engagement and concentration frequently, as learning occurs when the students are meaningfully involved in the learning environment.

To address the above-mentioned issues, this article presents a novel educational application called the MOEMO system with a real-time dashboard. The key contributions of this paper are:

1. To develop an LMS (Learning Management System)-independent learning analytics platform; however, integrable into a chosen e-learning platform such as Moodle, Sakai, and Canvas;
2. The platform could understand students' motions to detect various affective aspects such as engagement and concentration from the emotional data;
3. To make the platform capable of analyzing the data for valuable educational insights for the lecturers;
4. Visualize the results on the lecturers' screen in real-time so that they can intervene at the right time while the lecture is in progress;
5. To make the platform suitable for various educational scenarios such as problem-based learning, group discussion, and collaborative learning.

The rest of this paper is organized as follows: In the following Background section (Section 2), we discuss the background of this study, covering why emotion is vital in education, the association between emotion and learning, and learning theories associated with emotion and learning. In Section 3, we provide a literature review covering emotion and learning, emotion detection and ways to detect emotions from logs, sensors, and cameras; and learning analytics dashboards including teacher-centric, student-centric, and human-centric dashboards. Section 4 describes the materials and methods used to build this learning analytics system. Section 5 presents the dashboard and discusses its main components. In Section 6, we provide details on the use case conducted to test the system and understand its usefulness. Finally, the conclusions drawn from this work, along with possible future work lines, are described in Section 7.

2. Background

2.1. Emotion and Learning: Why Emotions Are Important in Education?

Kop et al. [5] stated that emotions are conceptual entities that arise from brain-to-body-to-environment interactions. Emotions are inherently linked to and influence cognitive skills such as attention, memory, executive function, decision-making, critical thinking, problem-solving, and regulation, all of which play a crucial role in learning [6]. Generally speaking, we have positive emotions and negative emotions that play powerful roles in learning. Positive emotions such as curiosity, passion, interest, wonder, creativity, and joy make our learning experience more desirable and aid in enhancing our focus and attention. These types of emotions enable learners to broaden their perspective and respond effectively to critical situations. On the other hand, negative emotions including sadness, disinterest, disengagement, anxiety, stress, worry, and fear can slow learning processes. These types of emotions negatively impact a learner's learning experience. While positive

emotions help learners stay engaged for longer, negative emotions could be disturbing. Yet negative emotions are not always bad for learning. For example, negative emotions such as confusion can increase students' interest levels. Additionally, negative emotions are important for learning complex concepts. Therefore, emotions, either positive or negative, are important factors that influence our learning process. Emotions influence our memory, reasoning capability, perception, and logical thinking, and therefore, in education, emotion is essential as it is highly associated with students' attention [7].

2.2. Theories Associated with Emotion and Learning

In order to find the relation between emotion and learning, psychologists and neurologists have evaluated many theories. For example, one study [8] evaluated a model in terms of how emotions influence students' learning and achievement. According to the findings of this study, the influence of emotions can be mediated by several mechanisms with cumulative or contradictory effects on predicting overall effects on performance. The second example is Pekrun's control-value theory of emotions [9]. The findings of this study indicated that the influence of emotions can be mediated by a few mechanisms with cumulative or contradictory effects for predicting overall effects on performance.

In addition, affective models are proposed that aim to establish a link between academic emotions and engagement levels. For example, an affective model to identify the level of student engagement depending on their emotions is proposed by Khawlah Altuwairqi [10]. This study conducted a series of experiments to find the relation between engagement levels and their emotions. A study conducted by CR Seal [11] proposed social and emotional development (SED). SED is the integration of theory, emotional intelligence, and competence development applied to educational practice. This study suggests that sustainable enhancement of personal capacity to utilize emotional information, behaviors, and traits could facilitate social outcomes.

Following these theories, in academics, understanding students' emotional states is crucial to understanding the learning process. However, despite the importance of recognizing learners' emotions in e-learning platforms, only a few platforms that recognize learners' emotions could be found. Hence, new frameworks need to be developed to uncover the unsolved learning problems.

2.3. Measurements of Emotion in E-Learing

Online learning has become the core of research and practices in learning analytics, educational data mining, artificial intelligence in education, intelligent tutoring systems, and educational recommendation systems. The research and practices of learning analytics (LA) are defined as "the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" by the Society for Learning Analytics Research (SOLAR). Multimodal learning analytics (MMLA), a subbranch of learning analytics, is focusing research on how critical aspects of the learning process could be revealed using multimodal data such as facial, emotional, gesture, and cognitive data [12].

With the help of LA and MMLA, many complex issues, including disengagement in online learning, are explored. Disengagement is regarded as one of the main challenges in online learning. Due to disengagement and poor concentration, many learners fail to achieve their learning goals. To understand more about why disengagement and poor concentration happen, LA and MMLA approaches measure the student's work context, actions, utterances, facial expressions, body language, and interactions with teachers or fellow students. As of now, many machine learning models, learning analytics systems, and learning analytics dashboards are being developed to regulate learning and teaching. However, measuring emotions and affective states during the class has been a challenge for teachers without LA or MMLA technologies.

3. Literature Review

Emotion and affect are associated with cognitive processes and learning [13]. (Craig et al., 2004); Mayer's affective cognitive model explains this relationship as follows: the learning episode causes an emotional reaction in the learner that affects cognitive processing during learning and leads to a learning outcome [14]. Studies reveal that different emotional states are associated with task completion and academic success in learning environments. For instance, an intelligent learning environment called AutoTutor [15] has been developed for educational purposes using emotional data. This study found a link between learning and the affective states of confusion, flow, and boredom. In addition, this study also revealed that learning has a positive correlation with confusion and flow but a negative correlation with boredom. In the same line, Whitehill et al. found that students' engagement levels correlate with their task performance [16]. Sharma et al. also found a correlation between the emotions expressed during the lesson and the concentration level of the students [17]. However, many potentials and aspects of emotional data are yet to come to light. Therefore, researchers are encouraged to collect larger sample sizes. Additionally, more research needs to be conducted about interventions with learning technologies. More precise timing and transitions between emotions are needed to uncover emotion dynamics, and interventions designed to regulate and productively respond to learner emotions are needed [18].

So far, studies conducted on emotion measurement are divided into two groups. The studies in the first group are based on a person's self-report. The other group of studies relies on biometric measures and other multimodal data sources. Self-report-based studies ask learners to rate how strongly they are experiencing individual emotions, such as anxiety or enjoyment, at a given point in time in the context of a specific activity at hand [14]. For example, the Achievement Emotion Questionnaire, developed by Pekrun et al. [19], is one of the most commonly used emotion surveys used in educational settings. The Geneva Emotion Wheel (GEW) is another self-report instrument used to measure emotion. However, instead of Likert-type questions, a visual card in the form of a wheel is used here, where students can mark their emotional states [20]. The second type of study is non-invasive, and automatic methods based on biometric measurements are used. For example, Sharma et al. proposed a prototype computer vision-based system to measure students' levels of concentration (e.g., low, medium, and high) in real-time from the expressed facial emotions during a lesson [17]. In another study, Krithika and Lakshmi Priya proposed a system for identifying and monitoring students' emotions in an e-learning environment by eye and head movements [21]. On the other hand, Whitehill et al. developed a face-recognition-based system that recognizes students' engagement levels [16].

In the literature, many measurement techniques are employed in learning analytics and psychology to understand why students disengage during class. Although self-report is widely used to measure emotions, there are criticisms of the use of self-reported data. For example, honesty, introspective ability, interpretation, rating scale, validation process, response bias, and sampling bias are the issues for self-reported data. For these reasons, it is clear that there is a need for non-invasive and automatic methods to measure students' emotional states during their learning processes. Therefore, in this paper, we introduce a new learning analytics framework that could automate the emotion detection process using computer vision and deep learning methods. This learning analytics system uses a computer's built-in camera function as the source to record lecture videos. These days, having a built-in/web camera is affordable for any student with a laptop or desktop computer; therefore, the system does not rely on any other sensors. The system reads real-time information about the students' eye movements (eye gazing) and various facial emotions to determine concentration levels. The system reads real-time information about the students' eye movements (eye gazing) and various facial emotions to determine concentration levels by which an instructor monitors how engaged (or not) the students are. The MOEMO system, in its dashboard, also provides insights on the learners' emotional states, such as how "neutral", "fearful", "disgust", "angry", "surprised", "happy", and "sad" they are. In

real-time, the dashboard shows the instructor the top-3 clusters of engaged and disengaged students based on their engagement score. To regulate learning, it also notifies the instructor that the students have been disengaged for a long period of time so that the instructor can intervene. The use cases of this system could be typical e-learning environments, including collaborative learning, discussion in Zoom's breakout rooms, peer programming, and lectures related to foreign cultures in cross-cultural language learning scenarios.

4. Materials and Methods

4.1. The Architecture

The aim of building this framework is to allow the lecturers to understand their students' affective states based on their emotions during the lecture. Emotions are captured from students' facial expressions and eye movements. The key stakeholders of this platform are the lecturers who are delivering lectures using web conferencing tools such as Zoom or WebEx. The MOEMO platform reads raw data (i.e., lecture videos) from web conferencing tools such as Zoom or WebEx to extract students' emotions. The emotion detection module is the core of the platform, as it provides the system with the capability of collecting, storing, processing, analyzing, and visualizing all the data regarding students' emotions. The emotion detection module dotes on the system with emotion, engagement, and concentration detection. In addition, this module identifies the clusters of highly engaged and highly disengaged students in a class. Based on this module's analytics, the lectures know the appropriate intervention time for those students who are highly disengaged and have low concentration. Figure 1 shows the overall architecture of the system.

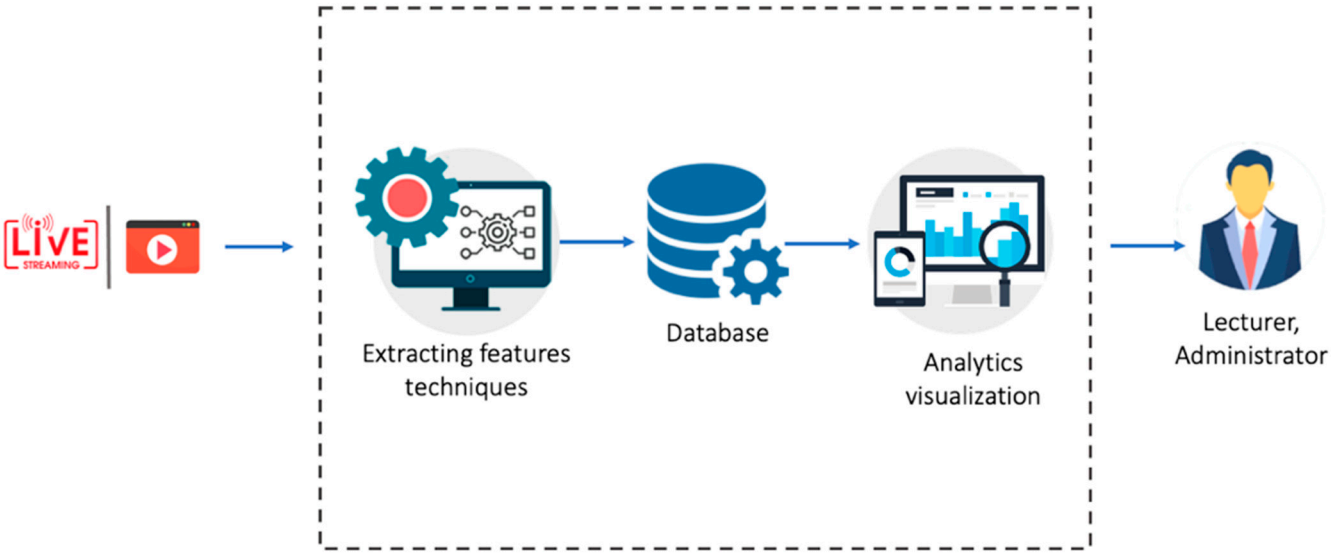


Figure 1. Overall architecture of the platform.

The technical architecture of the developed platform is shown in Figure 2. First, the system processes the video from the streaming URL or from the recorded video. Then, the system uses computer vision and deep learning techniques to extract emotional information from the entire video. The extracted information, such as face position information, student identification, emotional state, eye gaze, concentration level, and engagement level, is stored in the database. After processing the video, the system proceeds to load data from the database and analyze it to generate analytical data and statistical charts and display them on the web application for lecturers.

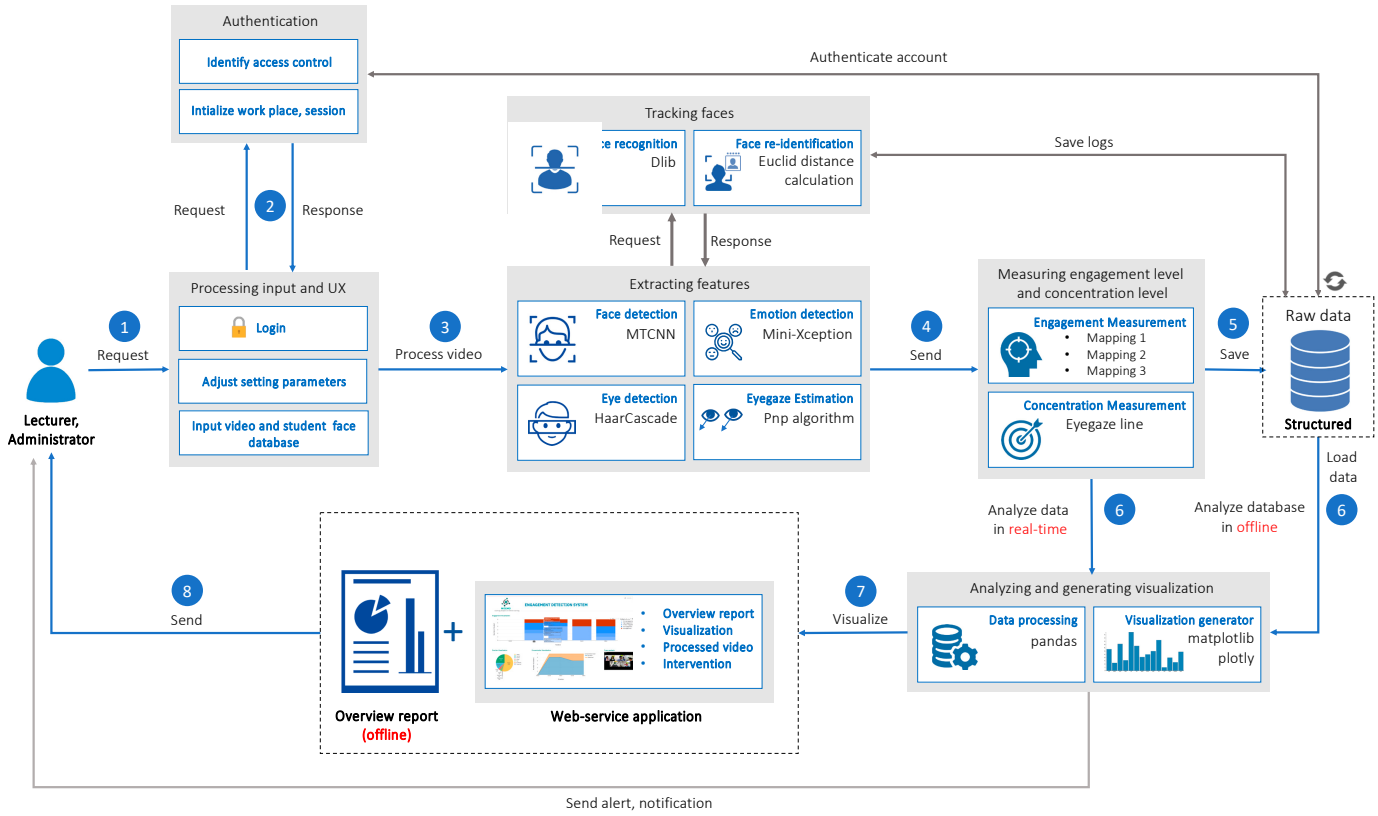


Figure 2. Detailed architecture of the platform.

In the backend, a general-purpose cross-platform software library called Dlib is used for face recognition tasks. Euclidean distance calculation was performed for the face re-identification task. Face re-identification occurs when a student leaves the web camera and returns after a while. For face detection, MTCNN (Multi-Task Cascaded Convolutional Neural Networks) is used. For various emotion extractions from facial features, the Mini Xception algorithm is used. For eye detection and eye gaze estimation, the HaarCascade and PNP algorithms are used, respectively. For data processing, a Python data analysis library called Pandas is used. To visualize the result on the dashboard, Matplotlib and Plotly's libraries are used.

4.2. Steps

As stated earlier, the platform can take recorded video and streaming video from the online class as input. The output of this platform is analytic visualizations. From taking recorded video and streaming video as the input to producing results, there are eight steps involved:

Step 1. The lecturer needs to input the following information for generating and processing the video. Figure 3 shows the operation of Step 1 in the system.

- First, to upload a recorded video or enter a streaming URL;
- Second, to upload a database for face recognition;
- Third, to customize the setting parameters.

Step 2. Authenticate the account (by the lecturer).

Step 3. Extract behavior features such as the face region, emotion, eye region, and eye gaze line. The methods employed for face detection, face recognition, face re-identification, and emotion detection.

Step 4. Measure the engagement levels and concentration levels. The methods for measuring concentration level and engagement level are articulated in 4.3.5. and 4.3.6., respectively.

Step 5. Save the extracted data to the database.

Step 6. Analyze the data in real-time or analyze the data from the database.
Step 7. Generate the visualization and send it to the web-based service.
Step 8. Visualize the results in the responsive dashboard and send a report to the lecturer.

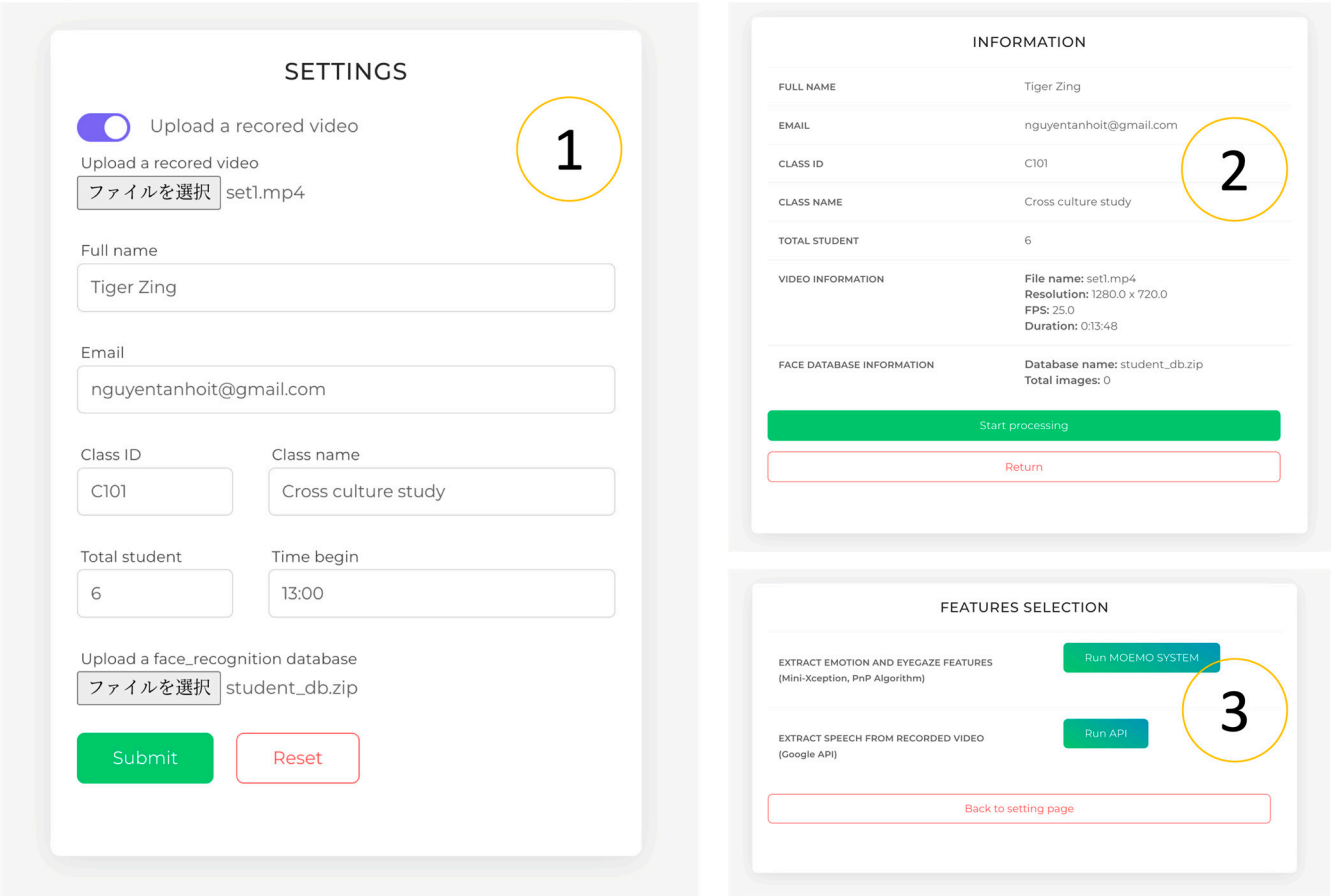


Figure 3. Parameter settings (1), video quality check (2), and customizable feature (3).

4.3. Methods

4.3.1. Face Detection Method

The Multi-task Cascade Convolutional Neural Network (MTCNN) [22] is employed as the face detection method. This deep learning method is one of the state-of-the-art recent methods that have achieved state-of-the-art results on standard benchmark face detection datasets. The MTCNNs consist of three stages. In the first stage, it produces candidate windows quickly through a shallow convolutional neural network. Then, it refines the windows to reject a large number of non-face windows through a more complex convolutional neural network. Finally, it uses a more powerful convolutional neural network to refine the result and output facial landmark positions. This pre-trained model is trained on a dataset including the face detection dataset and benchmarks FDDB [23], WIDER FACE [24], and the annotated facial landmarks in the wild (AFLW) benchmark [25]. Table 1 shows the datasets used to train the model.

Table 1. Datasets.

No	Name	Description
1	FDDB	Contains the annotations for 5171 faces in a set of 2845 images.
2	WIDER FACE	Consists of 393,703 labeled face bounding boxes in 32,203 images.
3	AFLW	Contains the facial landmarks annotations for 24,386 faces.

4.3.2. Face Recognition Method

Dlib library is used for face recognition. The Dlib model has achieved an accuracy of 99.38% on the standard benchmarks for face recognition. This accuracy means that, when the Dlib model is asked to correctly identify a pair of faces of the same person, it would detect the right face 99.38% of the time. The model transforms a human face into a 128–dimensional vector space where the images of the same person are near each other and images from different people are far apart. Therefore, performing face recognition by mapping faces to the 128–dimensional vector space and then checking if their Euclidean distance is small enough (using a distance threshold of 0.6) is possible. However, it requires a database that includes all students' faces.

4.3.3. Face Re-Identification Method

The main idea behind implementing the face re-identification model lies in estimating the small distances (Euclidean distance) from a face position in the current frame compared to the face positions in the previous frame. In short, the pair of similar faces will have the shortest Euclidean distance. The Euclidean distance between two points in Euclidean space is the length of a line segment between the two points. It can be calculated from the Cartesian coordinates of the points using the Pythagorean theorem, therefore occasionally being called the Pythagorean distance. The distance between (q_1,q_2) and (p_1,p_2) is given by:

$$d(q,p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2} \quad (1)$$

The condition is that the movement of the face between the two adjacent frames is tiny. If the movement is significant, we use face recognition to identify all the faces in the current frame. If there is any face that cannot be identified, we set the face as a new face.

For the experimental purpose, we first saved the face position as a central point (Face_x, Face_y). Then, we calculate the Euclidean distance between the face position in the current frame and all other face positions in the previous frame. The matching face will have the smallest distance. Note that the distance calculation is only suitable in cases where the movement of a face across two frames is very small. For that reason, a threshold is necessary to determine the magnitude of the minimum distance. The concept of a "threshold" in the context of face movement is pertinent to the two factors that influence such movement: the size of the face and the distance of the frame for processing. The purpose of this threshold is to decide when we need to use face recognition (refer to 4.3.2) or face re-identification methods. This threshold will depend on the size of the face and the distance between the two frames. In the real-world scenario, a user is able to adjust the "gap" between two frames for processing in the system. For example, if we obtain 1 frame every 3 seconds to process in a video with 25 fps, the "gap ratio" will be (75-1)/3*25. In short, the large size and the long distance between the two adjacent frames cause significant motion. This means if the gap ratio is bigger than the threshold, we will apply face recognition because the gap between the two frames is too long for us to apply face identification. If the gap ratio is smaller than the threshold, we can apply face identification due to the short gap between the two frames. For example, if we process 1 frame every 3 seconds, it leads to a large motion.

$$\text{threshold} = 1.5 * \max(\text{width}, \text{height}) + 2 * \Delta f \quad (2)$$

where (width, height) is the size of the face and Δf is the distance between two frames, with Δf < (3*fps) meaning that distance calculation cannot be used if we process the frame every 3 seconds because of unidentified motions. The face re-identification method saves much more time than the face recognition method. Therefore, we need to use this threshold to save the processing time as much as possible.

4.3.4. Emotion Detection Method

Mini–Xception [26], which is one of the state-of-the-art methods for extracting emotion from facial expressions, was used for this study. The Mini–Xception model has achieved

an accuracy of around 95.60% on the FER-2013 dataset [27]. This convolutional neural network (CNN) model has been used to build various real-time systems. However, we faced several challenges related to the performance of the Mini–Xception model that may affect the performance of emotion detection. For example, (i) the detected emotion may not be accurate when a student is speaking without facing the camera, (ii) when a student is doing something off-task, such as playing with a smartphone, and (iii) the orientation of the face may affect the accuracy of emotion detection. At present, the MOEMO system extracts seven types of basic emotions, namely "angry", "disgust", "fear", "happy", "sad", "surprise", and "neutral". Here, our challenges included when the face is covered by objects, emotion detection under bad light, and when speaking or engaging in off-task behavior.

4.3.5. Concentration Level Measurement Method

In concentration level measurement, we leveraged eye gaze information to determine the concentration level. In particular, we calculate the angle of the human eye with respect to the camera and the length of the eye gaze line (as shown in Figure 4). We implemented the flowing pseudo-code for determining if the student is looking at the screen or looking away from the screen.

```
if (degree >= 35 and degree <= 135) and length_eyegaze < length_threshold:
    return focused_state
else:
    return distracted_state
```
(3)

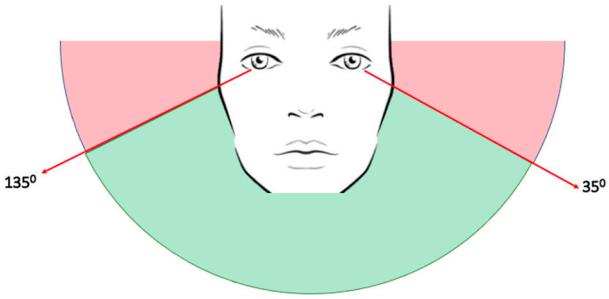


Figure 4. The focused_state and distracted_state detection method for identifying concentration level.

We employed the Perspective-n-Point (PnP) algorithm—a widely used technique in computer vision and computer graphics for determining the position and orientation of a three-dimensional (3D) object in a scene based on its 2D projection in an image. Eye detection and feature extraction methods are first applied to the image to obtain the eye_center point, eye_gaze line, and eye_view using the PnP algorithm. Once the eyes are localized, landmarks are extracted using techniques such as the Active Shape Model (ASM) or the Active Appearance Model (AAM), which enable estimation of the geometric center of the eye or the eye-center point. It is worth mentioning that other methods, such as Haar cascades, Viola-Jones, or deep learning-based methods, can be used for this purpose. The eye_gaze line, which denotes the direction of the gaze, is computed by drawing a line from the eye_center point in the direction of the gaze. This is achieved by estimating the rotation and translation parameters of the camera using the PnP algorithm and projecting the eye_center point onto the image plane. The direction of the eye_gaze line is then obtained by calculating the vector connecting the projected eye_center point with the image center. By transforming the eye position into the coordinate system of the camera using the estimated head pose and projecting it onto the image plane using the camera's intrinsic parameters, it is possible to calculate the eye view. In the above pseudo-code, we use the

math functions math.atan2 and math.degrees to calculate the degree between two points: the eye_center point and the eye_view point.

$$degree = math.degrees\left(math.atan2\left(P_{eye_center}, P_{eye_view}\right)\right) \quad (4)$$

where P_{eye_center} is the eye center point $\left(x_{eye_center}, y_{eye_center}\right)$ and P_{eye_view} is the eye view point $\left(x_{eye_view}, y_{eye_view}\right)$.

The eye gaze line is a way of showing where the eye is looking in 3D space. The longer the eye's gaze line, the farther the eye can see. After conducting many experiments, we have found a threshold for the eye gaze line that can help us determine the quality of someone's vision, which means the person is looking at the screen or not. To understand how well someone can see between their eye and a computer/laptop, we set a benchmark or a standard (length_threshold). This threshold helps us decide whether their visual range is good or not. After testing the system using various intervals, we set this interval [35°–135°] as an allowed threshold for determining if the student is looking at the screen or looking away from the screen. Furthermore, the length of the eye gaze line is considered when measuring concentration level. The length of the eye gaze line indicates the distance between the eye and the point of view. In short, the shorter the length, the closer the student looks to the camera. In this system, we set the length_threshold as:

4.3.6. Engagement Detection Method using Emotion Data

In this research, students' engagements are detected using the mapping method described in the related literature [7]. The calculation of engagement detection is based on the concentration index calculation method proposed by Sharma et al. [17]. The mapping method uses simple assumptions and mathematical grounds to identify student engagement from emotional data.

5. The Dashboard

5.1. Design of the Dashboard

The dashboard of the MOEMO system provides a detailed insight into the student's affective states on a lecturer's screen while the class is in progress (that is, in real–time). The dashboard can generate an after–class report. It is designed in a way that the instructors can interpret the graphs and other visualizations as efficiently as possible. The analytics engine is designed to run in real time and update related tables in the database. Figure 5 shows the dashboard.

5.2. Components of the Dashboard

The current version of the dashboard has eleven elements to support instructors in understanding affective states, concentration levels, and engagement levels. Table 2 summarizes the elements, update intervals, and descriptions of the dashboard elements.

5.2.1. Engagement Prediction Panel and Engagement Graph

The dashboard contains an engagement prediction panel where instructors can see the prediction made by the system on their students' engagement rate. The standardized scores between 0 and 100 are generated from their emotional data. The dashboard also has an engagement graph that visualizes online learners' five types of engagement, namely "strong engagement", "high engagement", "medium engagement", "low engagement", and "disengagement". This way, an instructor could see the overall engagement rate in the classroom.

5.2.2. Concentration Prediction Panel and Concentration Graph

The dashboard is supported by a concentration prediction panel where instructors can see the prediction made by the system on their students' concentration rate. As with the engagement rate, the standardized scores between 0 and 100 are generated from

students' emotional data. The concentration graph on the dashboard visualizes two types of concentration, which are "focused" or "distracted".

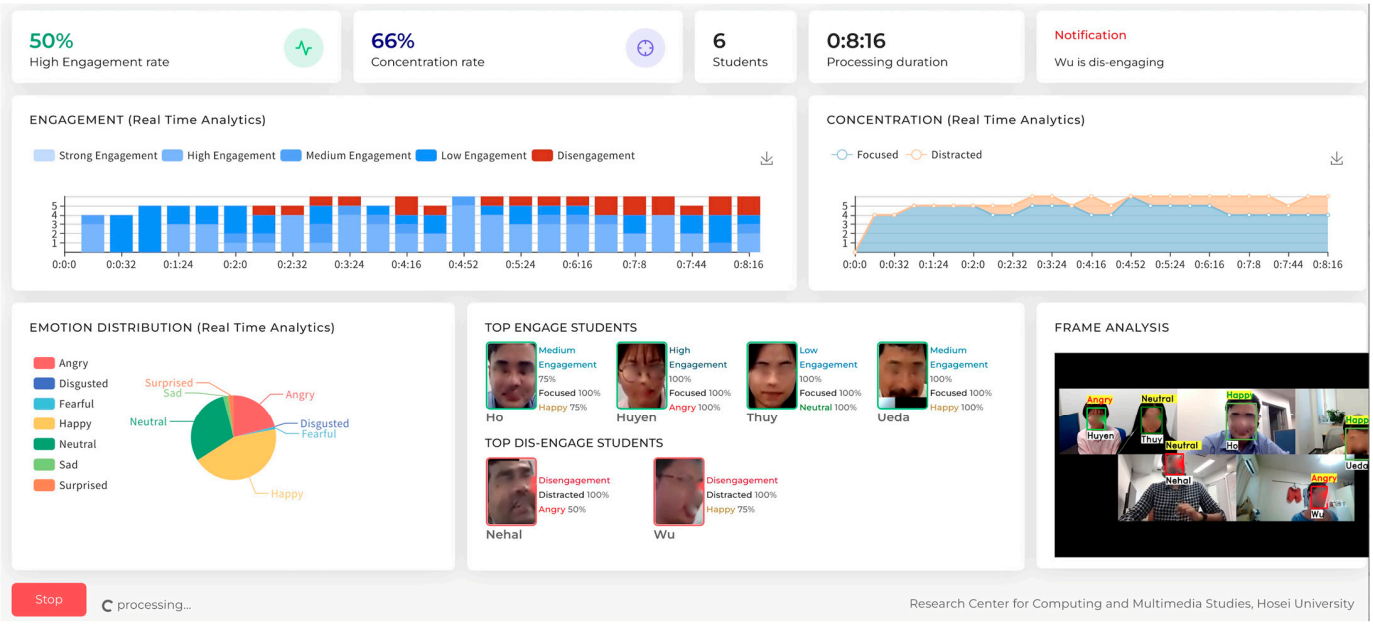


Figure 5. The dashboard.

Table 2. Components of the dashboard.

Dashboard Function	Update Interval	Description
Engagement prediction panel	Realtime	Overall engagement rate (range 0 to 100%)
Concentration prediction panel	Realtime	Overall concentration rate (range 0 to 100%)
Classroom overview	Once (before class)	Number of students in the class
Processing duration	Realtime	Total processing time of the lecture video
Notification panel	Realtime	Intervention
Engagement graph	Realtime	Engagement level in each minute
Concentration graph	Realtime	Concentration details
Cluster panel	Realtime	Top engaged and disengaged students
Emotion distribution	Realtime	Overall emotional rate of the class
Frame analysis	Realtime	Shows the affective states
After class report	Once (after class)	Report provides a detailed insight of the class's affective states

5.2.3. Notification Panel

The notification panel is the intervention system designed to inform instructors about students who have been disengaged for a long time. When a student is disengaged for more than 30 seconds, a notification appears on the notification panel. By looking at the notifications, the instructor can intervene to increase engagement. The time to set intervention notifications could be easily customized. For instance, if a teacher is unhappy with a 30–second intervention, they could select their intervention time.

5.2.4. Cluster Panel

The cluster panel, in real–time, shows the clusters of top-engaged and top–disengaged students. These clusters are generated based on student emotional data. The cluster panel is updated every second.

5.2.5. Emotion Distribution

In the pie chart on the dashboard, the distribution of overall emotions is shown to the instructor. Seven types of emotional states, such as "neutral", "fearful", "disgust", "angry",

"surprised", "happy", and "sad", are shown on the dashboard. By seeing this pie chart, the instructor can quickly understand whether students enjoy the class and how their affective states are at a particular class time.

5.2.6. Frame Analysis

The dashboard contains a video player with frame analysis capability. This video player shows, in real-time, students' affective states, such as how "neutral", "fearful", "disgusted", "angry", "surprised", "happy", and "sad" they are.

5.2.7. After-Class Report Generation

The dashboard is supported by an after-class report generation function. Using this function, an instructor could generate a report after the class. This report comprehensively details students' affective states, concentration, engagement, and intervention time. Figure 6 shows the insight of the after-class report.

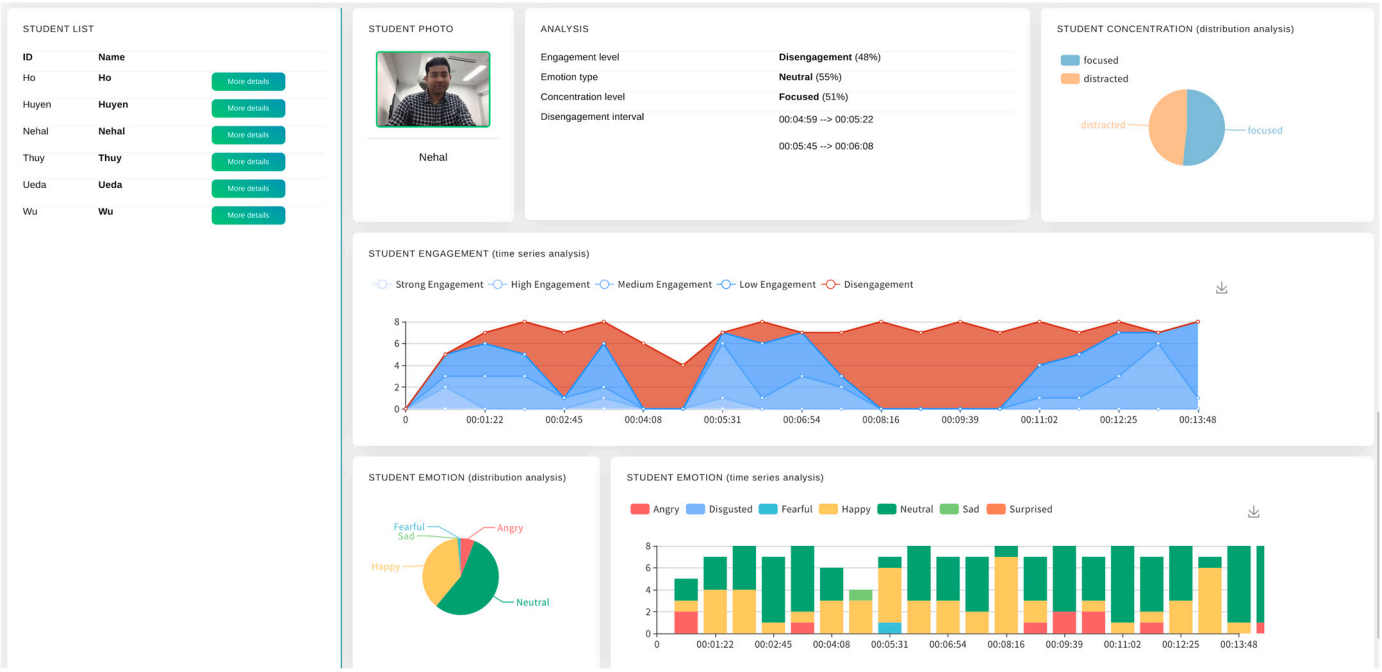


Figure 6. After-class report generation.

6. Case Study

In order to test the system and better understand the usefulness of the dashboard, we conducted a case study. We created a class called "*Cultural Exchange on Kanji*" in this case study. The class was formed with one teacher and five students. The students have the cultural backgrounds of Vietnam (N = 3), China (N = 1), and Bangladesh (N = 1). The teacher was native Japanese. Before the case study, we explained the purpose of this case study to the teacher and the students, describing how they allow us to extract their emotional data to visualize their affective states, engagement level, and concentration level. The topic of discussion was "How is Kanji used in your culture?". The discussion covered the history and revolution of Kanji in China and Vietnam. The class was conducted using the Zoom video conferencing tool. The class was 9 minutes and 9 seconds long. We visualized the result on the teacher's display during the class. We also generated an after-class report for the teacher. The result of this case study is presented in Figures 5 and 6.

7. Conclusions and Future Works

The emotions of online learners can be identified by the patterns in their ability to think, respond, communicate, or behave in an educational context. Human beings are highly influenced by emotion when making decisions and exhibiting behaviors with their surroundings and other aspects of their well-being [28]. In education, emotions are very important. To describe the importance of emotion in learning, Damasio elaborates, "emotions are inseparable from the ideas of reward and punishment, pleasure and pain, approach and withdrawal, personal advantage and disadvantage. In organisms equipped to sense emotions—that is, to have feelings—emotions have an effect on the mind, as they occur, in the here and now. Emotion is devoted to an organism's survival (quoted) [29]".

In online learning, disengagement and low concentration levels are critical issues that teachers, researchers, and educational institutions often face. Concurrently, understanding students' affective states is challenging in online education. Therefore, a teacher-support tool is essential to monitor students' engagement and concentration frequently, as learning occurs when the students are meaningfully involved in the learning environment. In this study, we designed and developed a framework to analyze lecture videos to understand students' affective states, engagement levels, and concentration levels from their facial features. We employed computer vision and deep learning methods to extract emotion from students' faces. After that, we constructed a database for handling emotional data. In real-time, the analytics engine of the proposed MOEMO system provides insight into engagement prediction, concentration prediction, classroom overview, processing duration, notification of intervention, engagement graph, concentration graph, student list, clusters of engaged and disengaged students, emotion analytics, frame analysis per second, and after-class report generation. The dashboard of the system could be operated in real-time or offline. Therefore, this work's main experimental contribution was designing and implementing a real-time learning analytics dashboard that could automatically detect online learners' affective states using deep learning and computer vision techniques.

However, this study has several limitations that will be addressed in future work. The first limitation is a proper evaluation of the system in terms of academic performance, and the accuracy of the data has yet to be determined. To address this limitation, a set of hypotheses was raised: (H1) Being aware of online learners' affective states helps an instructor guide them to better academic performance; (H2) the dashboard's quality as a learning analytics tool to monitor online learners' engagement and concentration; and (H3) the types of engagement (i.e., five types) and concentration (i.e., two types) presented on the dashboard are standardized. Therefore, the extended work would test these hypotheses with the dashboard's usage data. The second limitation is the validation of the emotional data. To overcome this limitation, we aim to validate the data by collecting feedback from online learners. One considered approach is to validate the data by implementing the Geneva Emotion Wheel (GEW: https://www.unige.ch/cisa/gew/ (accessed on 19 April 2022)). The third limitation is that we could not ensure to what extent the dashboard can provide relevant information about student behaviors so that the lecturers can monitor, observe, and measure students' engagement, concentration, and intervention time. Therefore, the extended work will undoubtedly explore this.

Author Contributions: Conceptualization, M.N.H. and G.A.; methodology, M.N.H.; software, M.N.H., H.T.N., T.T.T.T. and H.T.T.B.; validation, M.N.H., H.T.N., T.T.T.T. and G.A.; formal analysis, M.N.H., H.T.N., T.T.T.T. and H.T.T.B.; investigation, M.N.H., H.U. and G.A.; resources, M.N.H. and H.U.; data curation, M.N.H. and H.U.; writing—original draft preparation, M.N.H.; writing—review and editing, M.N.H., H.T.N., and G.A.; visualization, H.T.N., T.T.T.T. and H.T.T.B.; supervision, M.N.H.; project administration, M.N.H. and H.U.; funding acquisition, M.N.H. and H.U. All authors have read and agreed to the published version of the manuscript.

Funding: The APC is covered by the fund of Research Center for Computing and Multimedia Studies, Hosei University.

Institutional Review Board Statement: Ethical review was waived for this study due to- 'there will be no invasion of the research subjects, and there is no possibility of greater risk than that which would occur in their normal daily lives' (translated from Japanese using DeepL). It was approved by the Institutional Review Board (or Ethics Committee) of Hosei University (protocol code 2021-1 and date of approval: 8 July 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study. Necessary precautions to protect personal information and measures for information leakage were maintained strictly.

Acknowledgments: The authors would like to thank the participants of the experiment.

Conflicts of Interest: The authors have no conflict of interest to declare.

References

1. Wang, Y.; Liu, Q.; Stein, D.; Xia, Q. Measuring students affective states through online learning logs—an application of learning analytics. *Int. J. Inf. Educ. Technol.* **2019**, *9*, 356–361. [CrossRef]
2. Guragain, N. E-learning benefits and applications. *Psychol*. 2016. Available online: https://urn.fi/URN:NBN:fi:amk-201602122192 (accessed on 19 April 2022).
3. Silvola, A.; Näykki, P.; Kaveri, A.; Muukkonen, H. Expectations for supporting student engagement with learning analytics: An academic path perspective. *Comput. Educ.* **2021**, *168*, 104192. [CrossRef]
4. Skinner, E.A.; Zimmer-Gembeck, M.J.; Connell, J.P.; Eccles, J.S.; Wellborn, J.G. Individual Differences and the Development of Perceived Control. *Monogr. Soc. Res. Child Dev.* **1998**, *63*, i-231. [CrossRef]
5. Kop, R.; Fournier, H.; Durand, G. *A Critical Perspective on Learning Analytics and Educational Data Mining*; Society for Learning Analytics Research Publishing: Canada, 2017.
6. Immordino-Yang, M.H.; Damasio, A. We Feel, Therefore We Learn: The Relevance of Affective and Social Neuroscience to Education. *Mind Brain Educ.* **2007**, *1*, 3–10. [CrossRef]
7. Hasnine, M.N.; Bui, H.T.; Tran, T.T.T.; Nguyen, H.T.; Akçapınar, G.; Ueda, H. Students' emotion extraction and visualization for engagement detection in online learning. *Procedia Comput. Sci.* **2021**, *192*, 3423–3431. [CrossRef]
8. Pekrun, R. The Impact of Emotions on Learning and Achievement: Towards a Theory of Cognitive/Motivational Mediators. *Appl. Psychol.* **1992**, *41*, 359–376. [CrossRef]
9. Ranelluci, J.; Hall, N.C.; Goetz, T. Achievement goals, emotions, learning, and performance: A process model. *Motiv. Sci.* **2015**, *1*, 98–120. [CrossRef]
10. Altuwairqi, K.; Jarraya, S.; Allinjawi, A.; Hammami, M. A new emotion–based affective model to detect student's engagement. *J. King Saud Univ. Comput. Inf. Sci.* **2018**, *33*, 99–109. [CrossRef]
11. Seal, C.R.; Naumann, S.E.; Scott, A.N.; Royce-Davis, J. Social emotional development: A new model of student learning in higher education. *Res. High. Educ. J.* **2011**, *10*, 114–115.
12. SOLAR. What Is Learning Analytics. Available online: https://www.solaresearch.org/about/what-is-learning-analytics/ (accessed on 19 April 2022).
13. Tyng, C.M.; Amin, H.U.; Saad, M.N.M.; Malik, A.S. The Influences of Emotion on Learning and Memory. *Front. Psychol.* **2017**, *8*, 1454. [CrossRef] [PubMed]
14. Mayer, R.E. Searching for the role of emotions in e-learning. *Learn. Instr.* **2019**, *70*, 101213. [CrossRef]
15. Craig, S.; Graesser, A.; Sullins, J.; Gholson, B. Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *J. Educ. Media* **2004**, *29*, 241–250. [CrossRef]
16. Whitehill, J.; Serpell, Z.; Lin, Y.C.; Foster, A.; Movellan, J.R. The faces of engagement: Automatic recognition of student engagementfrom facial expressions. *IEEE Trans. Affect. Comput.* **2014**, *5*, 86–98. [CrossRef]
17. Sharma, P.; Joshi, S.; Gautam, S.; Filipe, V.; Reis, M.J. Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning. *arXiv* **2019**, arXiv:1909.12913.
18. Graesser, A.C. Emotions are the experiential glue of learning environments in the 21st century. *Learn. Instr.* **2019**, *70*, 101212. [CrossRef]
19. Pekrun, R.; Goetz, T.; Frenzel, A.C.; Barchfeld, P.; Perry, R.P. Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemp. Educ. Psychol.* **2011**, *36*, 36–48. [CrossRef]
20. Scherer, K.R.; Shuman, V.; Fontaine, J.R.; Soriano, C. The GRID meets the Wheel: Assessing emotional feeling via self-report. In *Components of Emotional Meaning: A Sourcebook*; Fontaine, J.R.J., Scherer, K.R., Soriano, C., Eds.; Oxford University Press: Oxford, UK, 2013; pp. 281–298.
21. Krithika, L.B.; Gg, L.P. Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric. *Procedia Comput. Sci.* **2016**, *85*, 767–776. [CrossRef]
22. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]

23. Jain, V.; Learned-Miller, E. *FDDB: A Benchmark for Face Detection in Unconstrained Setting*; University of Massachusetts: Amherst, MA, USA, 2010.
24. Yang, S.; Luo, P.; Loy, C.-C.; Tang, X. Wider face: A face detection benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 5525–5533.
25. Koestinger, M.; Wohlhart, P.; Roth, P.M.; Bischof, H. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV workshops), Barcelona, Spain, 6–13 November 2011; Volume 9.
26. Fatima, S.A.; Kumar, A.; Raoof, S.S. Real Time Emotion Detection of Humans Using Mini-Xception Algorithm. *IOP Conf. Series Mater. Sci. Eng.* **2021**, *1042*, 012027. [CrossRef]
27. Panagiotis, G.; Perikos, I.; Hatzilygeroudis, I. Deep Learning Approaches for Facial Emotion Recognition: A Case Study on FER-2013. In *Advances in Hybridization of Intelligent Methods*; Springer: Cham, Switzerland, 2018; pp. 1–16.
28. Revanth, A.; Prathibamol, C.P. Graph Emotion Distribution Learning Using EmotionGCN. In Proceedings of the Future Technologies Conference (FTC) 2022, Vancouver, BC, Canada, 20–21 October 2022; pp. 217–229.
29. Damasio, A.R. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*; Harcourt Brace Publishing: Boston, MA, USA, 1999.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Measuring Students Affective States through Online Learning Logs — An Application of Learning Analytics

Yang Wang, Qingtang Liu, David Stein, and Qihui Xia

***Abstract*—The affective state is determinate to online learning quality. It is related to students' attitude, learning motivation, and learning engagement. Learning affective states consists of engagement, frustration, confusion, and off-task state in this study. Different affective states are associated with different online learning behavior features. Affective states analysis consists of data collecting, data processing, affective states analyzing, evaluating, and intervening. Students' affective states can be analyzed by the affective state detectors. 12,912 online learning operations from an online course are used to construct the affective state detectors and examine the effectiveness. The consistency of the analysis results with the self-reports supported that the model is reliable to analyze students' affective state.**

***Index Terms*—Research and application, online affective state; analysis model.**

I. Introduction

With the rapid development of information technology, online learning has become popular as a new learning environment. It opens a new door for learners who are busy with their work and family or those with physical challenges that prevent them from learning at schools far from them. However, there are several challenges faced by online learning. For example, there are lots of online learners but few of them insisted on learning and obtained a certification [1]. There are abundant learning resources, but few satisfy learners' needs. There are numerous online learning platforms but few of them engaged learners. The lack of in-depth analysis of students' affective states is an essential factor that influences their online learning [2]. The state of students' engagement, the form of their participation, and their learning attitude are becoming the starting and driving force of online learning. In-depth analysis of the affective states has become an important topic for the popularity of online learning.

Affective states refer to a learner's physical and mental concentration on learning. Studies have shown that different affective states can produce different learning effects [3]. An intelligent tutoring system is aimed at improving students' learning based on their affective states which is a combination of cognition and emotion. Therefore, affective states are the starting and destination of online learning. Being actively engaged in the learning process is necessary for deep learning. The concentration state allows learners to have higher and lasting emotional, cognitive, and behavioral engagement in learning content [4]. How to use students' learning log data to analyze their affective states and support their online learning is essential to improve the online learning quality. To this end, Gartner announced a project "Reducing Gap: Transforming Data into Actions" to analyze students' affective states through educational data to achieve effective teaching [5]. Furthermore, the eCART system can provide real-time affective states assessment reports through monitoring and connecting learners' past and current learning performance. It is useful to monitor learners' progress and generate learners' reports to achieve effective teaching feedback [6]. The Chinese Ministry of Education announced the 13th Five-Year Plan for Education Informatization which states that it is necessary to actively use online educational data to enhance the quality of online education [7]. It indicated that the data-driven affective states analysis can enable instructors to know students' affective states and find out the learning problems to intervene in time. It is a guide for the formative assessment.

Studies on affective states focused on the affective states analysis, the influencing factors of the affective state, and the corresponding strategies. Moreover, the affective state analysis mainly uses the self-assessment method [8]. As for the influencing factors of affective states, the questionnaires and investigation are used. Then, the correlation analysis is performed to determine the influencing factors of affective states. With the development of learning analytics, the affective state analysis attracted researchers' attention. These researchers focused on learners' behaviors, physiological indicators, and other objective data to perform the affective state analysis [9]. As for the learners' affective states analysis based on students' learning log data need further study. This study constructed a data-driven affective states analysis model and examined its effectiveness based on 12,912 online learning data of an online course. Corresponding teaching strategies for those learners in different states were proposed through the interview.

II. Related Work

A. The Affective State

The affective state describes a learner's mind when he or she is learning. The affective state is caused by human emotions. For example, the learning affective state is relatively relaxed when a learner is in a good mood, and his

Manuscript received November 11, 2018; revised March 1, 2019.

Yang Wang and Qingtang Liu are with the Central China Normal University, China (e-mail: wangyang0519@mails.ccnu.edu.cn, liuqtang@mail.ccnu.edu.cn).

David Stein is with the Ohio State University, USA (e-mail: stein.1@osu.edu).

Qihui Xia is with the Nanjing Normal University, China (e-mail: qihuixia1993@163.com).

doi: 10.18178/ijiet.2019.9.5.1226

learning affective state is negative when the mood is irritated. The cognitive state influences the learning affective state by promoting the learner's understanding of the learning content [10]. In addition, an engagement state can also enable students to produce positive emotional states. The affective state is influenced by personal and environmental factors. The personal factor mainly refers to the individual's intrinsic desire for understanding a specific topic. Namely, the personal factor is a learner's internal cognitive and emotional characteristics. However, the environmental factor is the structural feature related to task organization [11]. The stimulation of an active affective state requires the combination of personal and environmental factors. Many researchers take affective states as learners' engagement [12]-[14]. They proposed that the degree of learning engagement is an important factor in students' learning experience which can affect their learning efficiency [15].

At present, there are many studies on online learning affective states, but different scholars have different understandings of affective states. Schmitz, Klug and Schmidt [16] proposed that the learning affective state is the learning behavior state measured at a particular time in a specific situation. Ansari, Montoya and Netzer [17] defined the learning affective state as the learner's involvement level and proposed that the learning affective state is a stable and adjustable psychological state. The learning affective state is the level of a learner's engagement which determined his learning efficiency. Yang, Cheng and Shih [18] supported that the learning affective state is essential to improve students' learning efficiency. The current studies on the online learning affective state are mainly focused on the following two aspects: 1) analysing a learner's knowledge construction process and learning ability based on the learner's online learning logs [19]; 2) analysing a learner's engagement according to the learner's physiological data, such as brain waves, eye movements, and skin electricity [20]. However, the learning affective states analysis based on online learning data is still in an infancy stage.

Studies indicated that the learning affective state can make a difference in learning [3], [21], [22]. However, studies on this topic tend to use the investigation methods, which are subjective and can only be applied in a relatively long period with a large population. This indicated that an analysis method based on students' online learning log may be more helpful to analyze students' learning affective states.

B. Indicators of Learning Affective States

The learning affective state is the reflection of a learner's cognitive and emotional states. In the current studies of the online learning affective state, learners' online operations are often overlooked [23]. However, these behaviors are important indicators of students' learning affective state. Statistical analysis of each learner's online learning duration, resource access, and the unit quiz can provide data for the quantitative analysis of their online learning affective states. The learning behavior is a learner's external manifestation of emotion and cognition. Although the learning affective state is below the surface of the iceberg which is difficult to observe directly, learning behavior data can be detected and reported to scale the learning affective state. The indicators for the concentration, frustration, confusion and off-task states are shown in Table I which is based on the study carried out by Pardos, Baker, San Pedro, Gowda and Gowda [24].

TABLE I: DESCRIPTION AND DATA INDICATORS OF EACH LEARNING AFFECTIVE STATE

Dimension	Description
Concentration	The concentration is a state within a period of time. For example, if the learner continuously answered correctly, or the learner spends a lot of time after the first wrong answer and finally answers this question correctly.
Frustration	Frustration is an emotional state that occurs when the learner's wrong answer rate is high. A student who misses many times after viewing multiple tips is considered to be frustrated.
Confusion	Confusion is a state in which a learner answered incorrectly first and still answered incorrectly after spending a long time on the problem.
Off-Task	Off-Task is a state in which learners seldom answer questions or answer incorrectly but take a relatively long time.

III. METHODS

A. The Online Learning Affective State Analysis Model

The learning affective state analysis is a process of evaluating learners' cognitive and emotional engagement level during the learning process to interpret the data generated by learners and detect potential problems. It is an important aspect of learning analytics. The existing learning affective state analysis models mainly consist of the following four types: the linear learning analysis model proposed by Siemens [25]; the continuous improvement model represented by Elias [26]; the circinate learning analytics framework supported by Ifenthaler and Tracey [27]; and a multi-factor learning analysis model proposed by SOLAR [28]. Siemens's linear model of learning analysis contains the basic elements of learning analysis. Elias' continuous improvement model reflects the circularity of the learning analysis process from the perspective of technical resources. A data-driven online learning affective states analysis model (as shown in Fig. 1) based on Siemens's learning analytical process model and Elias' continuous improvement model was constructed aiming at analyzing the learning affective state.

The data-driven learning affective states analysis model mainly consists of the online learning data acquisition, the data processing and affective states analysis, and the assessing, forecasting, and intervention of learning. The data collected during the data collecting are learners' online operations, including the correct answer rate, the number of operations, the interval between two adjacent operations, questions continuously answered correctly, previous operations involved the same problem, incorrect answers on a topic, and the number of past helps request. These data are analyzed to evaluate students' learning affective states. There are many redundant data in a large amount of online learning data collection. The goal of the data processing and learning

affective states analysis is to mine useful information for analysis and visualization from the complex and varied online learning data. The data cleaning is the first step to check data consistency and process invalid values using the data standard of Table I and the method shown in Fig. 1. The learning affective state of each learner is determined according to their online learning data based on the online learning affective states' detector.

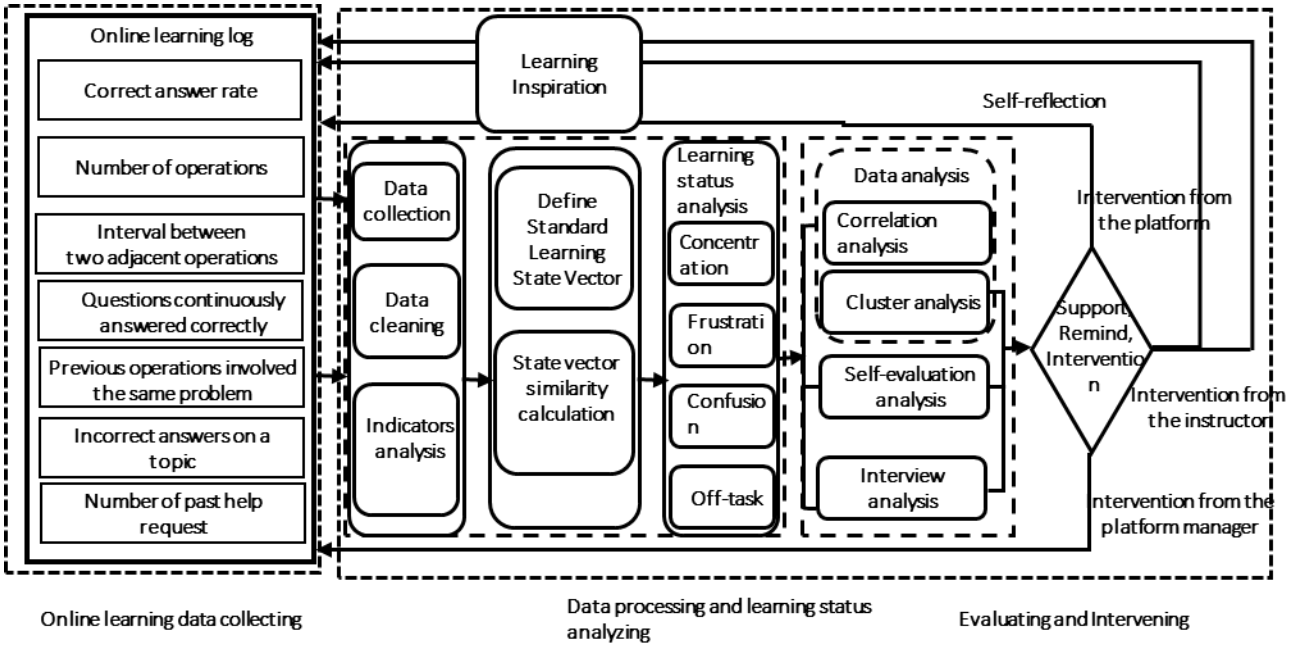


Fig. 1. Online learning affective states analysis model driven by the online learning data.

The purpose of assessing, forecasting and intervention is to engage learners in learning and maintain their active learning affective states. In this stage, the instructor is expected to intervene when learners demonstrate a negative learning affective states, such as frustration, confusion, and off-task states. For example, learners who show frustration and confusion need support in time. The presentation of corresponding knowledge and learning materials can help them to get out of frustration and confusion. For those learners who demonstrated off-task state, the appropriate monitoring, tips, and interventions are helpful. Platforms, learners, and instructors can take corresponding measures based on the results presented in the data processing and learning affective states analysis phase. For example, learners can reflect their learning behaviors, take corresponding measures, and adjust their learning affective states in time to achieve effective learning based on the presented results of the evaluation. The platform can be improved based on the analytical results. Instructors can improve teaching strategies, teaching content to optimize learning methods and stimulate learners' active learning affective states aiming at achieving adaptive teaching based on the learning affective state results. Managers can allocate resources rationally and affect the learning environment to further realize adaptive learning based on the analysis results. Through a series of interventions by learners, platforms, instructors, and managers, the online learning data, including learner-related data and learning environment-related data, will change. This can result in a cyclic process of online learning data collection, data processing, learning affective states analysis, evaluation, and intervention. It is aimed at stimulating the learner's positive learning affective states and improving their learning efficiency.

A. *Online Learning Affective States Analysis Method*

In the learning affective state analysis, the learning affective state detectors were used which take many indicators into account including 1) the correct answer rate; 2) number of operations; 3) time interval between two adjacent operations; 4) questions continuously answered correctly; 5) previous operations involving the same problem; 6) incorrect answers on a topic; 7) and the number of past help request et al. That is, the platform data was used to analyze the learning affective state. The correct answer rate is calculated according to the score of each learner in a test. The test consists of 10 questions and the full score is 100. The number of operations is calculated according to the number of behaviors of each learner. The time interval is calculated by the learning duration divided by the number of operations. The questions continuously answered correctly is calculated by the number of questions continuously answered correctly. The learner's previous operations involved the same problem, incorrect answers on a topic, and the total number of help request is calculated according to each learner's online learning logs. The online learning platform used in this study is the Clou*d Classroom* developed by us which can automatically record student's learning logs including every interaction with the system.

Two hundred sixty-nine learners' 12,912 online learning operations and their learning affective states self-report results from the course "Educational information processing" from 2016 to 2018 were collected. The data was used to train and examine the effectiveness of the learning affective states analysis model. The learners are Chinese undergraduates majoring in educational technology. The "Educational information processing" is a required course for them. The

data was shown in Fig. 2.

The self-report questionnaire for learners' online learning was constructed based on Wiebe's Learning affective states Assessment Scale [29]. There are 20 items in this scale to measure students' learning affective states. Moreover, there are 5 items in each of the following four dimensions, concentration, frustration, confusion, and off-task state [30]. Each item uses the Likert five-point scale, 1 represents strongly disagree and 5 represents strongly agree. The scale was examined to be reliable according to three online learning experts. The dimension with the highest score is the learner's main affective state. The scale was administrated to all the learners through email after the second class.

Half number (n=135) of students' learning log data were used to train the model, and the other half number (n=134) of students' learning data was used to examine the effectiveness of the model.

	userid	useraction	actiontime	mycharpterid	myunitid	mycoursewareid	status
2198001	2016[redacted]01	CourseToolsIDocViewShowAct	2017/4/1 7:33	2	1	1	online
2198002	201[redacted]01	login	2017/4/1 7:34	2	1	1	online
2198003	201[redacted]01	Assign_check	2017/4/1 7:34	2	1	1	online
2198004	20[redacted]01	CourseToolsIDocViewShowAct	2017/4/1 7:36	2	1	1	online
2198005	20[redacted]01	Assign_check	2017/4/1 7:37	2	1	1	online
2198006	20[redacted]01	AnswerQuestion1	2017/4/1 7:37	2	1	1	online
2198007	20[redacted]01	AnswerQuestion2	2017/4/1 7:39	2	1	1	online
2198008	20[redacted]01	AnswerQuestion3	2017/4/1 7:43	2	1	1	online
2198009	2[redacted]201	AnswerQuestion3	2017/4/1 7:47	2	1	1	online
2198010	2[redacted]201	AnswerQuestion3	2017/4/1 7:49	2	1	1	online
2198011	2[redacted]201	AnswerQuestion4	2017/4/1 7:52	2	1	1	online
2198001	2[redacted]201	AnswerQuestion4	2017/4/1 7:53	2	1	1	online
2198002	2[redacted]201	AnswerQuestion5	2017/4/1 7:55	2	1	1	online
2198003	2[redacted]201	AnswerQuestion6	2017/4/1 7:56	2	1	1	online
2198004	2[redacted]201	AnswerQuestion6	2017/4/1 7:59	2	1	1	online
2198005	2[redacted]201	AnswerQuestion7	2017/4/1 8:01	2	1	1	online
2198006	2[redacted]201	AnswerQuestion8	2017/4/1 8:04	2	1	1	online
2198007	2[redacted]201	AnswerQuestion9	2017/4/1 8:05	2	1	1	online
2198008	2[redacted]201	AnswerQuestion10	2017/4/1 8:06	2	1	1	online
2198009	2[redacted]201	AnswerQuestion10	2017/4/1 8:09	2	1	1	online
2198010	2[redacted]201	AnswerQuestion10	2017/4/1 8:11	2	1	1	online
2198011	2[redacted]201	AnswerQuestion10	2017/4/1 8:13	2	1	1	online
2198011	2[redacted]201	ViewAnswer10	2017/4/1 8:16	2	1	1	online
2198002	2[redacted]202	login	2017/4/1 7:34	2	1	1	online
2198004	2[redacted]202	CourseToolsIDocViewShowAct	2017/4/1 7:36	2	1	1	online
2198003	2[redacted]202	Assign_check	2017/4/1 7:34	2	1	1	online
2198005	2[redacted]202	Assign_check	2017/4/1 7:37	2	1	1	online
2198006	20[redacted]02	AnswerQuestion1	2017/4/1 7:37	2	1	1	online
2198007	2016[redacted]2	AnswerQuestion2	2017/4/1 7:39	2	1	1	online

Fig. 2. Learning log data.

The detectors for the four learning affective states were developed separately (concentration- no concentration; frustration- no frustration; confusion- no confusion; and off-task-not off task). The forward selection method was used to select the appropriate features for machine learning. The features that can improve the goodness of the model will be added until no more features can be added. Finally, 24 features were included in the learning affective states detector. Cohen's Kappa was used to evaluate the detectors in which the 0 represents chance level and 1 represents perfectly. The Kappa of 0.34 means that the detector is 34% better than chance at selecting the factor. The F measure indicates the model accuracy in which the 0 is the worst and 1 is the best [31]. All the indicators are better than the chance as shown in Table 2. The best concentration detector was found with the Naïve Bayes which demonstrating a Kappa of .682 and an F-Measure of .570. The best frustration detector was detected using K* algorithm with a Kappa of .368 and an F-Measure of .740. the detector of confusion was achieved with the Naive Bayes algorithm indicating a Kappa of .513 and F-Measure of .441. The detector of off-task was reached with the K* algorithm with a Kappa of .691 and an F-Measure of .754. these indicators supported that the model is informative. The 135 students' data were used to train the learning affective states analysis model. The classifier of Naïve Bayes and K* were used to build the model. The seven indicators of each learning affective state were collected. Results demonstrated that students in the concentration state demonstrated the following features: 1) a higher correct answer rate; 2) more operations; 3) a shorter time interval; 4) more questions continuously answered correctly; 5) more previous operations involved the same problem; 6) fewer incorrect answers on a topic; 7) and fewer past help requests. Students in frustration demonstrated the following features: 1) a lower correct answer rate; 2) more operations; 3) a shorter time interval; 4) fewer questions continuously answered correctly; 5) more previous operations involved the same problem; 6) more incorrect answers on a topic; 7) and more past help requests. Those in confusion demonstrated the following features: 1) a lower correct answer rate; 2) more operations; 3) a more time interval; 4) fewer questions continuously answered correctly; 5) more previous operations involved the same problem; 6) more incorrect answers on a topic; 7) and fewer past help requests. Students in off-task showed the following features: 1) a lower correct answer rate; 2) fewer operations; 3) more time interval; 4) fewer questions continuously answered correctly; 5) fewer previous operations involved the same problem; 6) fewer incorrect answers on a topic; 7) and fewer past help requests.

TABLE II: GOODNESS OF LEARNING AFFECTIVE STATES MODELS

Affective States	Algorithm	Kappa	F-Measure
Concentration	Naïve Bayes	0.682	0.57
Frustration	K*	0.368	0.74
Confusion	Naïve Bayes	0.513	0.441
Off-Task	K*	0.691	0.752

The other half number (n=134) of students' learning log data was used to examine the affective state detectors. These learners' operating data was analyzed using the method above. The symmetric measures results of the learning affective state self-reports and learning log analysis demonstrated that the overall results of the two methods are consistent (Kappa=0.871, p<.001). Though there are 22 inconsistencies between the analysis results and the self-report results, the consistency between the affective state analysis and the self-report reached 83.6% which demonstrated that the data analysis model is valid.

B. Interview

Two students in each learning affective state (concentration, frustration, confusion, and off-task state) were selected randomly for an interview. Namely, eight students were interviewed. The interview was mainly conducted from the following four aspects, engagement, satisfaction, perceived difficulty and expected support. Students in concentration showed a high level of engagement, satisfaction, and a low perceived difficulty. They thought that resources in the learning platform were comprehensive. Although the basic functions of the system can satisfy students' needs, they proposed that the resources can be refined which would help them find useful learning materials quickly. Students in the frustration and confusion states demonstrated a moderate level of engagement and learning satisfaction, a higher level of perceived difficulty compared with those in the concentration state. They suggested that the tips could be more detailed. In terms of expected support,

they expressed that the analysis of topics could be more detailed and instructor's timely help was beneficial for them. Besides, they expected that the real-time learning affective states visualization would help them realize the problem and adjust in time. Students in the off-task state showed a low level of learning engagement and satisfaction. Their perceived difficulty is moderate. They argued that it is easily distracted when faced with computers.

IV. Discussion and Conclusion

The analysis results can inform online instructors about students learning affective states. It can also provide new ideas for online learning affective states analysis. That is, making full used of students' online learning log data to analyze their learning affective states and give them corresponding feedback. It is very valuable to explore the influencing factors and data indicators of different online learning affective states. Compared with existing studies, this study analyzed students learning affective states through their learning log data. The leaning state analysis model was examined to be reliable.

There are some implications in this study for the teaching strategies for different learning affective states students. For example, one student whose learning affective state is concentration proposed that the resources of the learning management system are difficult to find. It indicates that these resources could be refined and categorized clearly. One male student whose learning affective state is frustration said that he is thirsty for the teachers' feedback when he cannot understand the tips. Another male student in confusion expressed a similar opinion. He thought if the confusion cannot be solved in time, he will lose his interest to move on. As for the male student in the off-task state, he wants to play the game when he faced with the computer. "There is no teacher here, I thought I can play first", he said. The stduents in the four learning affective states can represent many online learners situation. Their feedback can give us more suggestions on how to improve our online teaching.

In addition to the correct answer rate, the number of operations, and the items mentioned above, there are many other valuable learning logs such as the number of postings and reply content. These indicators are also valuable for further learning analytics. In the future study, more online learning operations can be collected to examine the effectiveness of the learning analysis model. The relationship of the learning affective states with their learning effectiveness can be further explored to inform instructors about the necessity of appropriate intervene for students' learning. It is a guide for adaptive online teaching and an important force in facilitating personalized teaching.

Acknowledgment

This study was funded by Ministry of education - Chinese mobile research fund project, "Research and experiment on Regional Teaching and research supported by information technology"(No: MCM20170502) and Hubei Province Technology Innovation special projects "Key technologies and demonstration applications of Internet + Precision Education"(No.2017ACA105).

Reference

[1] K. Ng, "Implementation of new communication tools to an online chemistry course," *Journal of Educators Online*, vol. 15, no. 1, p. n1, 2018.

[2] M.-S. Jian and J.-L. Chen, "Individual learner big data based cloud role player game learning," *Intelligent Data Analysis*, vol. 21, no. S1, pp. S41-S53, 2017.

[3] M. M. T. Rodrigo, R. S. Baker, M. C. Jadud, A. C. M. Amarra, T. Dy, M. B. V. Espejo-Lahoz, S. A. L. Lim, S. A. Pascua, J. O. Sugay, and E. S. Tabanao, "Affective and behavioral predictors of novice programmer achievement," pp. 156-160.

[4] Y. Li and R. M. Lerner, "Interrelations of behavioral, emotional, and cognitive school engagement in high school students," *Journal of Youth and Adolescence*, vol. 42, no. 1, pp. 20-32, 2013.

[5] J. R. Evans, and C. H. Lindner, "Business analytics: The next frontier for decision sciences," *Decision Line*, vol. 43, no. 2, pp. 4-6, 2012.

[6] E. M. Queiroga, C. Cechinel, R. M. Araújo, and G. C. Bretanha, "Generating models to predict at-risk students in technical e-learning courses," pp. 1-8.

[7] W. Pang and J. A. Plucker, "Recent transformations in China's economic, social, and education policies for promoting innovation and creativity," *The Journal of Creative Behavior*, vol. 46, no. 4, pp. 247-273, 2012.

[8] T. M. Miller and L. Geraci, "Training metacognition in the classroom: The influence of incentives and feedback on exam predictions," *Metacognition and Learning*, vol. 6, no. 3, pp. 303-314, 2011.

[9] J. Qiu, J. Tang, T. X. Liu, J. Gong, C. Zhang, Q. Zhang, and Y. Xue, "Modeling and predicting learning behavior in MOOCs," pp. 93-102.

[10] W. Kintsch, "Learning from text, levels of comprehension, or: Why anyone would read a story anyway," *Poetics*, vol. 9, no. 1-3, pp. 87-98, 1980.

[11] Z. Wang, C. Bergin, and D. A. Bergin, "Measuring engagement in fourth to twelfth grade classrooms: The classroom engagement inventory," *School Psychology Quarterly*, vol. 29, no. 4, pp. 517, 2014.

[12] T. Dragon, I. Arroyo, B. P. Woolf, W. Burleson, R. El Kaliouby, and H. Eydgahi, "Viewing student affect and learning through classroom observation and physical sensors," pp. 29-39.

[13] Y. L. Hu, G. S. Ching, and P. C. Chao, "Taiwan student engagement model: Conceptual framework and overview of psychometric properties," *International Journal of Research Studies in Education*, vol. 1, no. 1, pp. 69-90, 2012.

[14] M. T. Wang and J. Degol, "Staying engaged: Knowledge and research needs in student engagement," *Child Development Perspectives*, vol. 8, no. 3, pp. 137-143, 2014.

[15] D. J. Shernoff, *Optimal Learning Environments to Promote Student Engagement*, Springer, 2013.

[16] B. Schmitz, J. Klug, and M. Schmidt, "Assessing self-regulated learning using diary measures with university students," *Handbook of Self-regulation of Learning and Performance*, pp. 251-266, 2011.

[17] A. Ansari, R. Montoya, and O. Netzer, "Dynamic learning in behavioral games: A hidden Markov mixture of experts approach," *Quantitative Marketing and Economics*, vol. 10, no. 4, pp. 475-503, 2012.

[18] M. T. Yang, Y. J. Cheng, and Y. C. Shih, *Facial Expression Recognition for Learning Status Analysis*, pp. 131-138.

[19] Y. Charband and N. J. Navimipour, "Online knowledge sharing mechanisms: A systematic review of the state of the art literature and recommendations for future research," *Information Systems Frontiers*, vol. 18, no. 6, pp. 1131-1151, 2016.

[20] C.-H. Wu, B.-C. Kuo, and G.-H. Tzeng, "Investigation of time interval size effect on SVM model in emotion norm database," *New Trends in Intelligent Information and Database Systems*, pp. 121-128, Springer, 2015.

[21] R. S. J. D. Baker, "Modeling and understanding students' off-task behavior in intelligent tutoring systems," pp. 1059-1068.

[22] M. Cocea, A. Hershkovitz, and R. S. Baker, "The impact of off-task and gaming behaviors on learning: Immediate or aggregate"?

[23] M. A. Conde, R. Colomo-Palacios, F. J. García-Peñalvo, and X. Larrucea, "Teamwork assessment in the educational web of data: A learning analytics approach towards ISO 10018," *Telematics and Informatics*, vol. 35, no. 3, pp. 551-563, 2018.

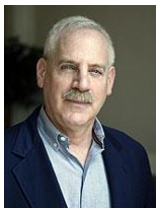
[24] Z. A. Pardos, R. S. Baker, M. O. San Pedro, S. M. Gowda, and S. M. Gowda, "Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes," pp. 117-124.

[25] G. Siemens, "1st international conference on learning analytics and knowledge 2011 (LAK'11)," 2011.
[26] T. Elias, "Learning analytics," *Learning*, 2011.
[27] D. Ifenthaler and M. W. Tracey, "Exploring the relationship of ethics and privacy in learning analytics and design: Implications for the field of educational technology," *Educational Technology Research and Development*, vol. 64, no. 5, pp. 877-880, 2016.
[28] G. Siemens, D. Gasevic, C. Haythornthwaite, S. P. Dawson, S. Shum, R. Ferguson, E. Duval, K. Verbert, and R. Baker, "Open Learning Analytics: An integrated & modularized platform," 2011.
[29] E. N. Wiebe, A. Lamb, M. Hardy, and D. Sharek, "Measuring engagement in video game-based environments: Investigation of the User Engagement Scale," *Computers in Human Behavior*, vol. 32, pp. 123-132, 2014.
[30] Author, *Details Removed for Peer Review*, 2016.
[31] C. J. Van Rijsbergen, "Foundation of evaluation," *Journal of Documentation*, vol. 30, no. 4, pp. 365-373, 1974.

Yang Wang was born in China in May 1991. Yang earned her bachelor degree in educational technology in Hubei Normal University in 2014. Yang is a Ph.D. candidate at the School of Educational Information Technology at Central China Normal University (CCNU) and now a visiting scholar at the Ohio State University (OSU). Her research interests include learning analytics, technology-supported teaching and learning. She has published more than 7 papers in this field including in the British Journal of Educational Technology, 2018, Distance Education, 2017, and IEEE Consumer Electronics Magazine 2018.

Qingtang Liu was born in China in February 1969. He is a professor at the School of Educational Information Technology at CCNU. He earned the doctoral degree in Huazhong University of Science and Technology in 2005. His recent research interests include technology-enhanced teaching and learning. He has published more than 60 papers in these fields including the Computers and Education, 2017; Journal of Educational Computing Research, 2015; the British Journal of Educational Technology, 2018.

Dr. Liu is currently a member of ISO/IEC, a member of AVS Standards Organization, a member of Chinese E-Learning Technology Standardization Committee (CELTSC); a member of Chinese Higher Education Technology Association, a member of Chinese Education Management Information Committee.



David Stein is an associate professor of College of Education and Human Ecology at the Ohio State University. He earned his doctor degree in the University of Michigan in 1976. His work centers on online teaching and learning. His recent research has focused on how learners make decisions concerning participation, choice, and use of technology in learning online settings. He has published more than 131 research papers and books including a book in 2015, the Internet and Higher Education (2013), American Journal of Distance Education (2011), and Distance Learning (2013).

Dr. Stein is currently an associate professor in the Ohio State University (OSU). He is a chair of the American Association for Adult and Continuing Education from 2014 to 2016. He earned an Imogene Oakes Award for Outstanding Research in Adult Education in 2013. He was a co-chair of the Commission on Distance Learning and Technology of the American Association for Adult and Continuing Education from 2014 to 2016.

Qihui Xia was born in China in January 1993. Qihui gets her master degree of psychology at Nanjing Normal University in 2018. Her research interests include the development and educational psychology, children's theory of mind development. Her paper "Just World Belief: Conception, measurement and related research" obtained the second prize in the 17th National Forum on the Applied Research of Psychological Techniques and the 2016 Academic Conference of Jiangsu Psychological Society in November 2016. She obtained the third prize in 2013 Jiangxi Province College Student Technology Innovation and Vocational Skills Competition in December 2013.

REVISTA INTERNACIONAL
TECNOLÓGICA - EDUCATIVA DOCENTES 2.0
ISSN: 2665-0266 / ISNI: 0000 0005 0409 1664 / RINGGOLD ID 608948.
VOL.15 (1), JUNIO 2022 | https://doi.org/10.37843/rted.v15i1.308

Modelos Tecnológicos de Aprendizaje Adaptativo Aplicados a la Educación

Adaptive Learning Technology Models Applied to Education

Raúl Quintanar-Casillas[1] y Ma. Sandra Hernández-López[2]

EDICIÓN: RTED Resiliencia Paradigmática

Recibido: 24/febrero/2022
Aceptado: 01/junio/2022
Publicado: 29/junio/2022

Páginas: 41-58

País
[1]México
[2]México

Institución
[1]Facultad de Informática. Universidad Autónoma de Querétaro
[2]Facultad de Ingeniería. Universidad Autónoma de Querétaro

Correo Eletrónico
[1]rquintanar02@alumnos.uaq.mx
[2]ma.sandra.hernandez@uaq.mx

ORCID
[1]https://orcid.org/0000-0002-5943-3885
[2]https://orcid.org/0000-0002-0768-87802

Citar así: APA / IEEE

Quintanar-Casillas, R. & Hernández-López, Ma. S. (2022). Modelos Tecnológicos de Aprendizaje Adaptativo Aplicados a la Educación: Revisión Sistemática de la Literatura. *Revista Tecnológica-Educativa Docentes 2.0*, 15(1), 41-58. https://doi.org/10.37843/rted.v15i1.308

R. Quintanar-Casillas y Ma. S. Hernández-López, "Modelos Tecnológicos de Aprendizaje Adaptativo Aplicados a la Educación: Revisión Sistemática de la Literatura", *RTED*, vol. 15, n.° 1, pp. 41-58, jun. 2022.

Resumen

En los últimos diez años, se han publicado diversas investigaciones que proponen modelos o sistemas de aprendizaje adaptativo o personalizado tomando como base el estudio de los estilos de aprendizaje, las habilidades cognitivas o la interacción del estudiante con los objetos de aprendizaje. El objetivo de esta investigación fue efectuar una revisión sistemática de los modelos o sistemas de aprendizaje adaptativo propuestos durante el periodo 2012-2021 tomando en cuenta autores de origen hispanoamericano. Para su realización, se empleó el método analítico, bajo el paradigma positivista, con enfoque cuantitativo de tipo descriptivo apoyado en el metaanálisis y con un diseño experimental de orden transversal. Se aplicó la metodología de análisis de contenido apoyada en las directrices de la declaración *Preferred Reporting Items for Systematic reviews and Meta-Analyses* para la elaboración de revisiones sobre una muestra de 50 estudios seleccionados de diferentes bases de datos científicas que propusieran modelos o sistemas de aprendizaje adaptativo. Se utilizaron dos categorías de análisis: objeto de estudio y estructura del sistema o modelo propuesto. A través de una matriz bibliográfica, como, una matriz analítica se registró y organizó la información recabada. Dentro de los resultados obtenidos se observó que la mayor parte de los estudios tienen estructuras basadas en tres componentes, siendo la inteligencia artificial la técnica de adaptabilidad más empleada; de igual forma, se observó una relativa exigüidad de estudios elaborados en Hispanoamérica. Se concluyó mencionando los nichos de investigación sobre aprendizaje adaptativo aplicado a la Educación y una serie de sugerencias en torno a trabajos futuras.

Palabras Clave: Modelos tecnológicos, aprendizaje adaptativo, educación.

Abstract

In the last ten years, several investigations have been published that propose adaptive or personalized learning models or systems based on the study of learning styles, cognitive abilities, or student interaction with learning objects. The objective of this research was to carry out a systematic review of the adaptive learning models or systems proposed during the 2012-2021 period, taking into account authors of Hispanic-American origin. For its realization, the analytical method was used, under the positivist paradigm, with a descriptive quantitative approach supported by meta-analysis and a cross-sectional experimental design. The content analysis methodology was applied based on the guidelines of the Preferred Reporting Items for Systematic reviews and Meta-Analyses declaration for the elaboration of thoughts on a sample of 50 studies selected from different scientific databases that proposed adaptive learning models or systems. Two categories of analysis were used: the object of research and the structure of the proposed approach or model. Like an analytical matrix, a bibliographic matrix was broken and organized the information collected. The results showed that most of the studies have structures based on three components, with artificial intelligence being the most used adaptability technique; likewise, a relative scarcity of studies carried out in Latin America and Spain was observed. It was concluded by mentioning the research niches on adaptive learning applied to Education and a series of suggestions regarding future work.

Keywords: Adaptive learning, technology models, education.

D.L:LA2017000128 / ISSN: 2665-0266 / Ringgold ID:608948
ISNI: 0000 0005 0409 1664
Junio 2022 - Volumen 15, Número 1, Ejemplar No. 32.

Las obras se publican en DOCENTES 2.0 bajo una licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 International (CC BY-NC-ND)

Introducción

En los últimos diez años, se han publicado diversas investigaciones que proponen modelos o sistemas de aprendizaje adaptativo o personalizado tomando como base el estudio de los estilos de aprendizaje (EA), las habilidades cognitivas (HC) o la interacción del estudiante con los objetos de aprendizaje (OA); en algunos casos se han incluido técnicas de inteligencia artificial (IA) como el aprendizaje automático (*machine learning*) o el aprendizaje por reforzamiento (Rothman, 20218). A nivel global, la región de Hispanoamérica cuenta con poca producción científica en cuanto a propuestas de aprendizaje adaptativo. Dadas las premisas antes mencionadas resulta de interés para los investigadores y académicos que enfrenten problemáticas en las que la adaptabilidad del aprendizaje se perciba como una solución, conocer el estado del arte de los modelos y sistemas de aprendizaje adaptativo y aprendizaje personalizado, identificando los objetos de estudio y las estructuras propuestas en estos modelos.

Durante el proceso de búsqueda se ubicaron tres revisiones sistemáticas sobre el aprendizaje adaptativo y el aprendizaje personalizado: a) Xie et al. (2019) realizan un análisis descriptivo de 70 diferentes trabajos sobre aprendizaje adaptativo o personalizado clasificándolos por año de publicación, edad, nivel de estudios de los sujetos de estudio, temática del contenido educativo, tipo de enfoque, hardware utilizado y tipo de resultados obtenidos; por su parte; b) Afini et al. (2018) exponen una revisión sistemática de 78 estudios enfocados a la investigación de los rasgos individuales en ambientes de aprendizaje adaptativo; c) por su parte, Zhai et al. (2021) consideran 100 diferentes estudios que utilicen a la inteligencia artificial (IA) aplicada en la educación.

Estas tres revisiones exponen la construcción del estado del arte de los modelos y sistemas de aprendizajes adaptativo y personalizado desde diferentes aristas, en diferentes niveles de detalle, aunque que no consideran dentro de su análisis las estructuras de los modelos planteados, las técnicas de IA empleadas; además no se incluyen los trabajos realizados en la región hispanoamericana.

De acuerdo con Alvarado et al. (2013), con la concepción del estado del arte es posible avanzar en la producción de nuevo conocimiento. Tomando en consideración esta aseveración, la presente investigación tuvo como objetivo aplicar una revisión sistemática a los modelos o sistemas de aprendizaje adaptativo propuestos durante el periodo 2012-2021 mediante la metodología *Preferred Reporting Items for Systematic reviews and Meta-Analyses* (PRISMA) y considerando los estudios realizados en Hispanoamérica, con el fin de conocer el estado del arte en la materia. Este cuerpo de conocimiento resulta útil al fungir como una base teórica en el desarrollo de nuevos modelos y sistemas de aprendizaje adaptativo.

Metodología

La presente investigación se desarrolló bajo el método analítico, bajo el paradigma positivista, con enfoque cuantitativo de tipo descriptivo apoyado en el metaanálisis y con un diseño experimental de orden transversal. El método analítico que es un camino para llegar a un resultado mediante la descomposición de un fenómeno en sus elementos constitutivos (Lopera et. al, 2010), en este caso, se realizó la descomposición de los modelos y sistemas considerados para el estudio en sus elementos estructurales. De igual forma, esta investigación se fundamenta bajo el paradigma constructivista que permitió realizar una representación teórica de la realidad de los modelos y sistemas de aprendizaje adaptativo (Ramos, 2015).

El enfoque desde el que se abordó el presente estudio es de tipo cuantitativo apoyado en los procesos de codificación y cuantificación de las variables o categorías de estudio, y de esta forma, seguir un orden lógico e inconmutable (Hernández-Sampieri, 2018). La investigación se basó en las directrices establecidas en la declaración PRISMA en lo referente a la realización de revisiones sistemáticas (Page et al., 2021). El trabajo desarrollado es de tipo descriptivo a apoyado en el metaanálisis, identificando las estructuras de los modelos propuestos en los estudios analizados y comparando los resultados obtenidos con

investigaciones previas (Page et al., 2021). La investigación es de tipo descriptiva con un diseño no experimental de orden transversal, con la finalidad llevar a cabo una revisión sistemática de la literatura científica referente al aprendizaje adaptativo aplicado en la educación (Hernández-Sampieri & Mendoza, 2018).

Para desarrollar la revisión sistemática, en primera instancia se consideró como universo de estudio o población a los trabajos enfocados al aprendizaje adaptativo A partir de esta población, se seleccionó una muestra integrada por 50 trabajos de investigación sobre aprendizaje adaptativo. De acuerdo con Arias-Gómez et al. (2016), la población es un conjunto de casos, definido, limitado y accesible, que forma el referente para la elección de la muestra que cumple con una serie de criterios predeterminados; por su parte, la muestra es un subconjunto representativo de la población sobre el cual se ejecuta el proceso investigativo (Otzen & Manterola, 2017).

Durante el proceso de selección de trabajos, se emplearon como instrumentos a la matriz bibliográfica y a la matriz analítica para el registro de la información recabada. Este proceso se apoyó en la técnica del análisis documental; la técnica es la forma en la que se siguen los pasos del método, mientras que el instrumento es la aplicación de la técnica (González, 2020). Con la información registrada en la matriz bibliográfica, se realizó la codificación, para su posterior análisis estadístico, en el cual se cuantifican los datos y se realizan cálculos estadísticos a fin de entender e interpretar los resultados de dichos cálculos (Hernández-Sampieri & Mendoza, 2018).

Con respecto a la metodología PRISMA, el autor Page et al. determina dos fases investigativas:

1) Planeación y diseño, en la que se establece los criterios de elegibilidad, las fuentes de información y la estrategia de búsqueda; y

2) Gestión, Análisis, elaboración y formalización en la que se establecen a) el proceso de selección de los estudios dividido en tres etapas: identificación, cribado e inclusión y de cuyo resultado se determina el universo, la muestra y las categorías de análisis. Con esta información se realiza una lectura lineal apoyada en una matriz bibliográfica en la que se inventarían los textos que conforman el universo y sobre la cual se aplican los filtros de selección, y en una matriz analítica de contenido en la que se observen las categorías de análisis empleadas para el estudio; y b) proceso de extracción y análisis de los datos, en la que se realice la lectura horizontal sobre una matriz analítica de contenidos utilizada como método de síntesis a través de dos categorías de análisis: objeto de estudio y estructura propuesta (2020, p.794).

Primera Fase

Se realizó una búsqueda a través de diferentes motores como lo son Academia.edu, Google Académico, la Red de Revistas Científicas de América Latina y el Caribe, España y Portugal (REDALYC), Scientific Electronic Library Online (SciELO), Sciencedirect y Scopus, de trabajos que propongan Modelos de Aprendizaje Adaptativo o Personalizado utilizando las frases exactas "aprendizaje adaptativo" o "aprendizaje personalizado" y que incluyeran ya sea la palabra "sistema" o "modelo". Cabe señalar que se priorizó la búsqueda y recuperación de trabajos originados en la región de Hispanoamérica.

En idioma inglés, se emplearon "*adaptive learning*", "*personalized learning*" con la combinación booleana AND con los vocablos "*model*" o "*system*", ya que las búsquedas arrojaban un número de resultados mucho más amplio. En la Figura 1 se puede observar el flujo de trabajo para la búsqueda, recuperación y selección de estudios siguiendo la metodología PRISMA.

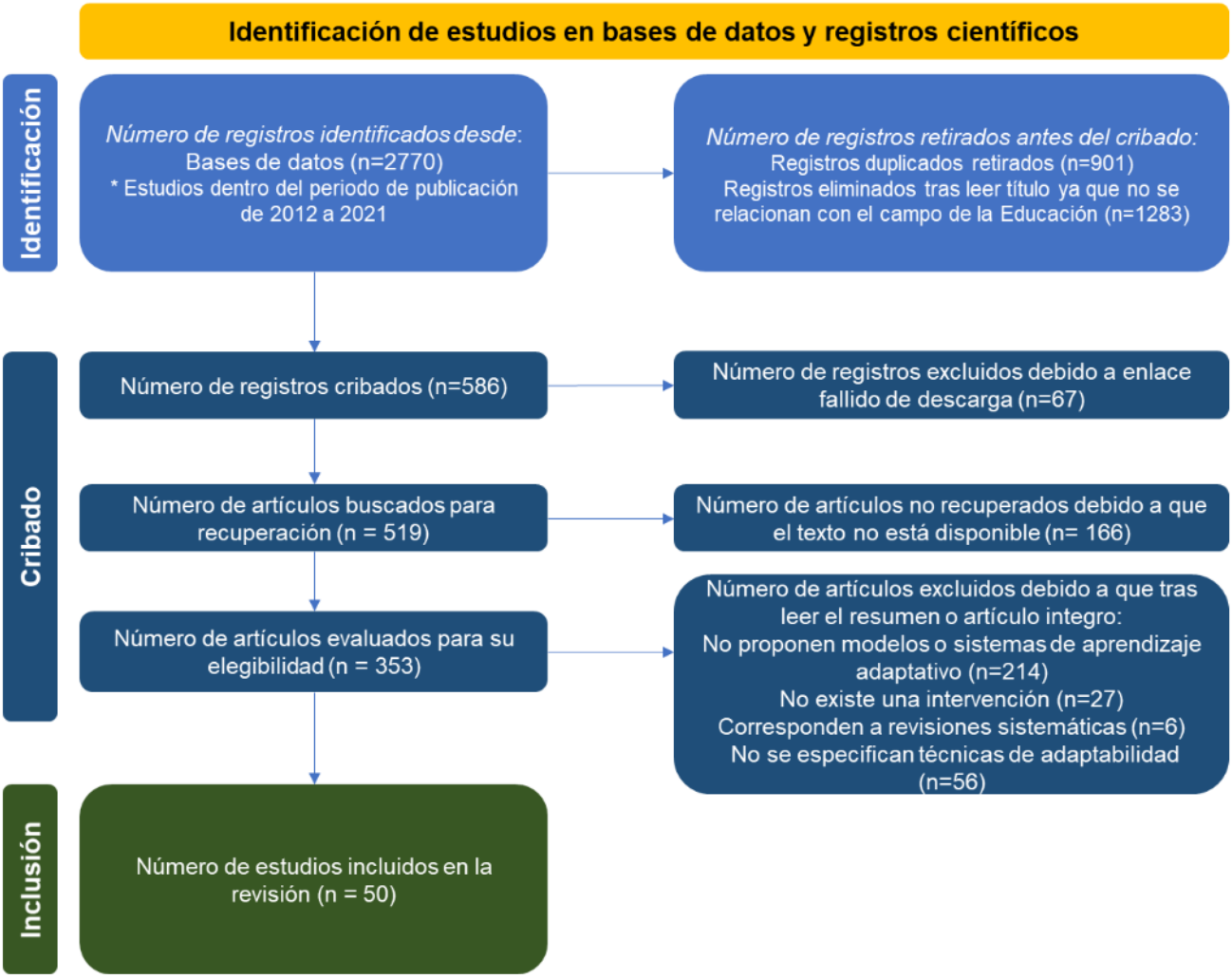
Como criterios de inclusión se estableció que para los artículos considerados: a) su fecha de publicación estuviera dentro del periodo comprendido entre los años 2012 a 2021 y b) la investigación subyacente se relacionara con el campo de la Educación. Como criterios de exclusión se estableció que los estudios no considerados fueran aquellos que: a) no propusieran modelos o sistemas de aprendizaje adaptativo; b) no tuvieran un proceso de intervención, en otras palabras, que el estudio no contemplara la aplicación de instrumentos de investigación a un conjunto determinado de sujetos de estudio; c) correspondieran a revisiones sistemáticas; y d) no especificaran técnicas de

Quintanar-Casillas, R. & Hernández-López, Ma.S. (2022). Modelos Tecnológicos de Aprendizaje Adaptativo Aplicados a la Educación: Revisión Sistemática de la Literatura. *Revista Tecnológica-Educativa Docentes 2.0*, 15(1), 41-58. https://doi.org/10.37843/rted.v15i1.308

adaptabilidad. A través de una lectura lineal como la sugerida por Gómez et. al (2015) se seleccionaron del universo de estudio inventariado en la matriz bibliográfica y después de aplicar los criterios de inclusión y exclusión, 50 trabajos de investigación sobre aprendizaje adaptativo.

Figura 1

Diagrama de Flujo PRISMA para la Búsqueda, Recuperación y Selección de Estudios



Nota. Proceso elaborado con base en la declaración PRISMA del año 2020, elaboración propia (2022).

Segunda Etapa

Los resultados obtenidos en la primera etapa se utilizaron para conformar una matriz analítica con la que se desarrolló un análisis de contenido mediante dos categorías de análisis: objeto de estudio y estructura. Para cuantificar los resultados de la segunda etapa, la categoría "Objeto de estudio" se define como una dimensión de la cual se desprenden los siguientes indicadores: a) el tipo de propuesta (variable dependiente) con la que se trabajó en cada estudio, verificando si se trata de un modelo o sistema de aprendizaje adaptativo (MAA o SAA) o de un modelo o sistema de aprendizaje personalizado (MAP o SAP); b) la estrategia utilizada en la construcción del objeto de estudio como pueden ser las habilidades cognitivas (HC), las competencias (CA), o los estilos de aprendizaje (EA) de los sujetos de estudio, o bien, los objetos de aprendizaje (OA); c) si los sujetos de estudio pertenecen a un ambiente escolar o laboral (E/L).

La segunda categoría de análisis es la "estructura propuesta" en cada uno de los estudios revisados. Como parte de la lectura vertical realizada en un inicio, se identificaron tres submodelos componentes que son los más utilizados: a) Modelo de Aprendizaje, b) Modelo del Dominio, y c) Modelo de Adaptación y se codificaron para su posterior cuantificación de acuerdo con la siguiente nomenclatura respectivamente: MA. MDOM y MAD. Otro componente identificado en la "estructura propuesta" son las técnicas de adaptabilidad utilizadas en los módulos o modelos de adaptación, identificadas con las siglas TA. Dentro del conjunto de estudios analizados se encontró un gran número de herramientas de Inteligencia Artificial empleados en el MAD, de las cuales se observa su clasificación en la Figura 2.

Figura 2

Clasificación de las Técnicas de Aprendizaje Automático para la Inteligencia Artificial

Nota. Clasificación de las diferentes técnicas de Aprendizaje Automático en la IA utilizadas en los Modelos y Sistemas de Aprendizaje Adaptativo (MAA/SAA), elaboración propia con base en Sabita (2021).

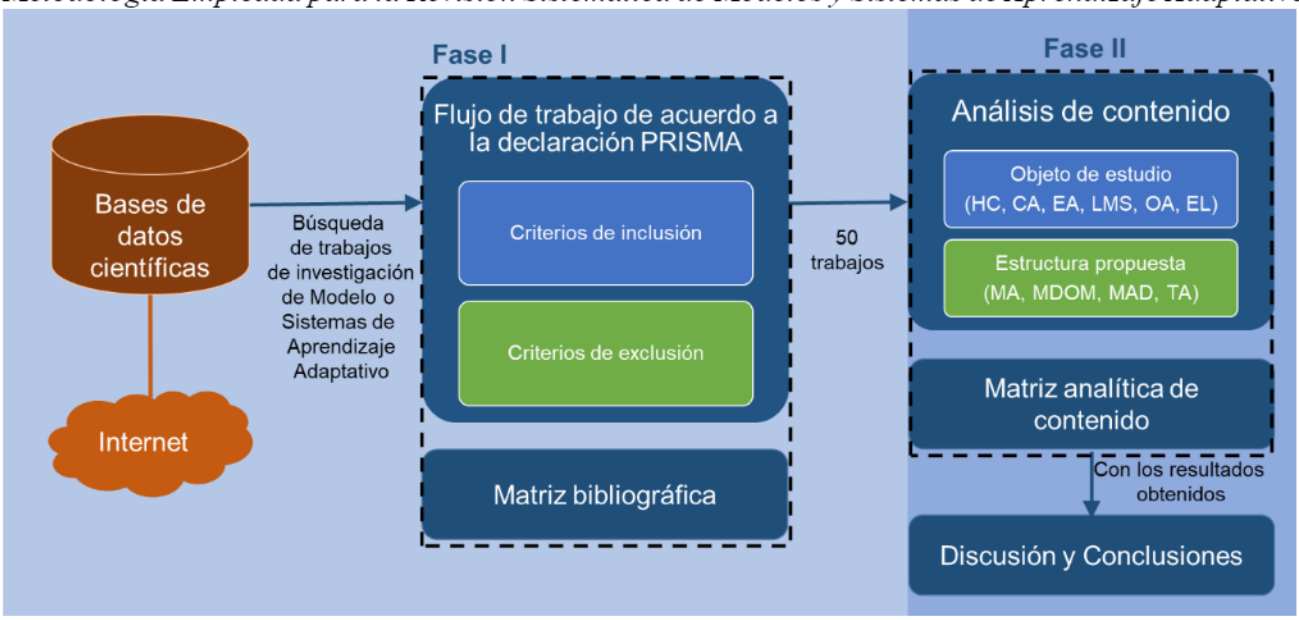
De esta forma la "estructura propuesta" se concibe como una dimensión cuyos indicadores son MA, MDOM, MAD y TA. A través de una lectura horizontal de la matriz analítica de contenido, se selecciona uno por uno los indicadores anteriormente definidos y se contrasta la información obtenida para cada una de las investigaciones revisadas. La codificación y cuantificación de las dimensiones e indicadores de la investigación permitieron exponer en los resultados una serie de gráficas que muestran las tendencias en materia de aprendizaje adaptativo.

Los resultados de la investigación son analizados en la sección "Discusión" y finalmente en la sección "Conclusiones" se exponen la importancia y utilidad del *corpus* definido, y las recomendaciones para propuestas futuras. En la Figura 3 se muestra un esquema que resume la metodología empleada para la revisión sistemática objeto de la presente investigación.

Quintanar-Casillas, R. & Hernández-López, Ma.S. (2022). Modelos Tecnológicos de Aprendizaje Adaptativo Aplicados a la Educación: Revisión Sistemática de la Literatura. *Revista Tecnológica-Educativa Docentes 2.0*, 15(1), 41-58. https://doi.org/10.37843/rted.v15i1.308

Figura 3

Metodología Empleada para la Revisión Sistemática de Modelos y Sistemas de Aprendizaje Adaptativo



Nota. Diagrama resumen del proceso completo que se llevó a cabo durante la investigación, elaboración propia (2022).

Resultados

Con base en el análisis de los estudios seleccionados, en lo que refiere a la dimensión de "Estructura", se observó que los módulos o componentes de los modelos o sistemas propuestos de aprendizaje adaptativo que más frecuentemente se utilizaban dentro de sus estructuras fueron el Módulo del Aprendiz (MAP), el Módulo de Dominio u Objeto de Aprendizaje (MDOM) y el Módulo de Adaptación (MAD). Por otro lado, para la dimensión del "Objeto de Estudio", se observó que un gran número de trabajos se enfocaron en el análisis de los objetos de aprendizaje y en proponer versiones mejoradas de los *LMS* incorporando elementos de adaptabilidad; la investigación de las habilidades cognitivas en el marco del aprendizaje adaptativo es escasa.

Primera Etapa

En la etapa de identificación se ubicaron 2770 estudios, de los cuales se retiraron 901 por estar duplicados y 1283 trabajos por tener un enfoque relacionado al estudio de sistemas de inteligencia artificial automatizados que utilizan algoritmos de aprendizaje adaptativo que no se aplicaban al terreno de la Educación. De este primer filtro se obtuvieron un total de 586 registros para la etapa de cribado de los cuales 67 presentaron un enlace de descarga incorrecto; de los 519 artículos restantes, no fue posible recuperar el texto completo de 166 de ellos, dando como resultado un total de 353 artículos que conformaron el universo de estudio, mismos que fueron evaluados para su elegibilidad e inventariados en una matriz bibliográfica. Los criterios de inclusión y de exclusión expuestos en líneas previas, se aplicaron al universo de artículos seleccionados dando como resultado en la etapa de inclusión un total de 50 trabajos que conformaron la muestra de estudio, mismos que fueron tomados en cuenta para efectuar la revisión sistemática.

Segunda Etapa

En esta etapa se utilizó una matriz analítica de contenido, la cual se puede observar en la Tabla 1, ordenada de forma cronológica, en la que se compendian los 50 trabajos de investigación elegidos para realizar la revisión sistemática. Se emplearon dos categorías de análisis o dimensiones: Objeto de Estudio y Estructura.

Tabla 1

Lista de Trabajos Enfocados en el Estudio del Aprendizaje Adaptativo

		Objeto de Estudio							Estructura				
ID	**Autor(es)**	**O**	**Prop.**	**HC**	**CA**	**EA**	**LMS**	**OA**	**E / L**	**MAP**	**MDOM**	**MAD**	**TA**
1	Gonçalves y García (2012)	H	SAA				X	X	E	X	X	X	IA - Sistemas de Recomendación
2	Ahmad, Tasirb, Kasimc, & Sahata (2013)	A	SAA			X	X		E	X	X	X	Modelo basado en literatura - FSLSM
3	Leris, Vea y Velamazán (2015)	H	SAA				X		E		X		Itinerario de aprendizaje
4	Peña, González y Mendoza (2015)	H	MAA			X	X	X	E	X			IA - Redes bayesianas
5	Roque-Alayón, Sánchez-Díaz, y López-Padrón (2015)	H	SAP				X		E		X		N/A
6	Velandia (2015)	H	SAA	X				X	E	X	X	X	IA - Arboles de decisión, redes neuronales, regresión lineal, SVM y algoritmos genéticos
7	May (2016)	E	MAA			X	X		E	N/A	N/A	N/A	IA - Reglas, redes bayesianas, redes neuronales y árboles de decisión
8	Qodad et al. (2016)	F	SAA		X	X	X		L	X	X	X	Reglas de extracción y FSLSM
9	Battou (2017)	F	SAA			X	X		E	X			Reglas Pedagógicas
10	Fernández, Ramírez y Blanco (2017)	H	MAA	X		X	X	X	E			X	Métodos centrados en profesorado, descubrimiento y resolución de problemas
11	González, Benchoff, Huapaya y Remon (2017)	H	SAA			X	X		E				Felder and Silverman (FSLSM)
12	Liu, McKelroy, Corliss & Carrigan (2017)	U	SAA				X		E		X	X	Brightspace LeaP™
13	Mira-Giménez (2017)	H	SAA		X		X	X	E			X	Itinerario de aprendizaje
14	Real-Fernández, Molina-Carmona y Llorens-Largo (2017)	H	MAA		X		X	X	E		X		IA - Redes neuronales gráficas (GNN)
15	Salazar-Ospina et al. (2017)	H	SAA			X		X	E	X	X	X	Sistema multiagente BROA
16	Shawky & Badawi (2018)	F	SAA					X	E			X	IA- Aprendizaje por refuerzo
17	Balasubramanian & Margret (2018)	A	MAA	X		X		X	E	X	X	X	IA - Aprendizaje por refuerzo
18	El Asame, Wakrim & Battou (2018)	F	MAA		X		X	X	E		X		Competencias requeridas
19	González, Becerra y Olmos (2018)	H	SAA			X		X	E			X	FSLSM
20	Kellman, P. & Krasne, S (2018)	U	SAA	X		X			E	X			VARK (Enfoque Perceptivo)
21	Serrao-Neumann, S., Cox. M. & Low, D. (2018)	O	MAA		X			X	L			X	Itinerario de aprendizaje
22	Sfenrianto, S., Hartarto, Y. & Akbar. H. (2018)	A	SAA		X		X	X	E			X	Itinerario de aprendizaje y árbol de decisión
23	Zulfiani, Suwarna, & Miranto (2018)	A	SAA			X	X	X	E	X	X		VARK
24	Brateanu et al. (2019)	U	SAA				X	X	E	X		X	Itinerario de aprendizaje
25	Chávez (2019)	H	SAP				X	X	E		X		N/A
26	Firdausiah, Yusofa, & Hoirul (2019)	A	MAA		X			X	E	X	X	X	IA – Redes neuronales

Quintanar-Casillas, R. & Hernández-López, Ma.S. (2022). Modelos Tecnológicos de Aprendizaje Adaptativo Aplicados a la Educación: Revisión Sistemática de la Literatura. *Revista Tecnológica-Educativa Docentes 2.0*, 15(1), 41-58. https://doi.org/10.37843/rted.v15i1.308

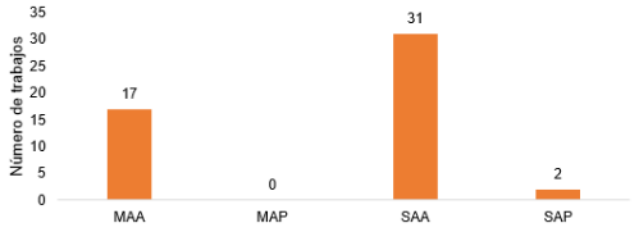
27	**Guevara y Aguilar (1) (2019)**	**H**	**MAA**					**X**	**E**		**X**		**Objeto de Aprendizaje Adaptativo**
28	Guevara y Aguilar (2) (2019)	H	MAA					X	E		X		Objeto de Aprendizaje Adaptativo
29	Hassan, Habiba, Majeed & Shoaib (2019)	A	MAA			X		X	E	X	X	X	FSLSM
30	Hünicken, González, Haag y Ruppel (2019)	H	SAA		X		X	X	E		X		Itinerario de aprendizaje
31	Otero, Rivera, Pedraza, y Canay (2019)	H	SAA					X	E			X	IA – Sistemas de Recomendación
32	Sheeba y Krishnan (2019)	A	MAA			X	X		E	X			FSLSM
33	Yaplea & Yu (2019)	A	MAA	X					E	X			Aprendizaje reverso
34	Duque, Ovalle & Carrillo (2020)	H	MAA			X		X	E			X	Reglas, VARK, FSLSM
35	Jiménez, Fernández y Almenárez (2020)	H	MAA		X			X	L		X		Itinerario de aprendizaje
36	Koch, de Azevedo, Sebastiany y Cortina	B	SAA				X	X	E	X	X	X	IA - AIEd
37	Koshova (2020)	E	MAA				X		L				N/A
38	Krechetov & Romanenko (2020)	E	SAA				X	X	E			X	IA - Algoritmo genético
39	Kuniyoshi & Kurahashi (2020)	A	SAA				X		E			X	IA - Red Compleja (Bayesiana)
40	Renn, Arean, Raue, Aisenberg, Friedman & Popovic (2020)	U	SAA	X					E			X	ITS - Intelligent Tutorial System
41	Sokol, Bronin, Karnaukh, & Bilov (2020)	E	SAA		X		X	X	L	X	X		Algoritmo de búsqueda
42	Tabares, Duque & Fabregat (2020)	H	MAA			X		X	E	X	X		VARK y FSLSM
43	Troussasa, Krouskaa & Sgouropouloub (2020)	E	SAA		X			X	E	X	X	X	Generador de consejos dinámicos (DAG)
44	Wang, Christensen & Cui (2021)	U	SAA				X	X	E		X		Squirrel Ai Learning ™
45	Lwande, Muchemi, & Oboko (2021)	F	SAA	X		X	X		E	X		X	Estimación del comportamiento por patrones
46	Rasheed & Wahid (2021)	A	SAA			X	X		E	X			IA - Arboles de decisión, SVM, Regresión Logística y FSLSM
47	Ríos-Rodríguez, Román-Cao y Pérez-Medinilla (2021)	H	SAA		X			X	E	X	X	X	Sistema multiagente APA Prolog ©
48	Véliz, A., Madrigal, O., Correa, V. (2021)	H	SAA					X	E	X			IA - Aprendizaje por refuerzo, retroalimentación y optimización
49	Yu, J. et al. (2021)	A	SAA		X			X	E		X		IA – Redes neuronales
50	Zhikharev, A., Deeney, A., Igrunova, S., Klyuchnikov. D. & Frolova, A. (2021)	E	SAA		X			X	E		X		IA – Redes neuronales

Nota. Origen (O): Asia (A), Brasil (B), Europa excepto España (E), África (F), Hispanoamérica (H), Estados Unidos (U); Propuesta de Sistema o Modelo (Prop), Habilidades cognitivas (HC), Competencias de Aprendizaje (CA), Estilos de Aprendizaje (EA), Sistema de Gestión del Aprendizaje (LMS), Objeto de Aprendizaje (OA), Contexto escolar o laboral (E/L), Módulo de Aprendiz (MAP), Módulo de Dominio (MDOM), Módulo de Adaptación (MAD), Técnica de adaptabilidad (TA), elaboración propia (2022).

En lo que refiere a la dimensión Objeto de Estudio, para cada uno de los indicadores se encontraron los siguientes hallazgos: a) en cuanto al tipo de propuesta de los 50 estudios analizados, 31 correspondieron a sistemas de aprendizaje adaptativo, 16 a modelos de aprendizaje adaptativo, 2 a sistemas de aprendizaje personalizado y 1 a un modelo de aprendizaje personalizado (Figura 4); b) para la construcción de su objeto de estudio, 34 trabajos consideraron el estudio de los objetos de aprendizaje, 28 tomaron en cuenta un LMS, 18 se enfocaron en el estudio de los estilos de aprendizaje, mientras que 14 consideraron estudiar las competencias del aprendiz y 7 las habilidades cognitivas (Figura 5); c) en cuanto al contexto al que se aplicaron los modelos o sistemas propuestos, 45 se enfocaron a un ambiente escolar y 5 a un ambiente laboral (Figura 6).

Figura 4

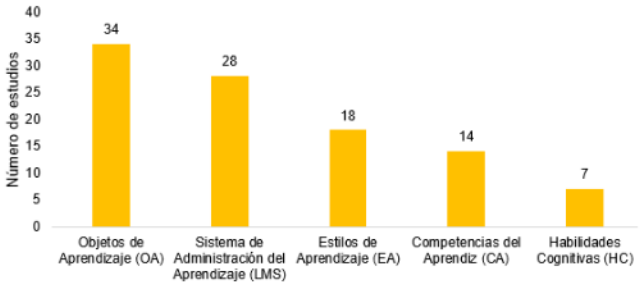
Distribución de Trabajos por Tipo de Propuesta



Nota. Dimensión "Objeto de Estudio", Modelo de Sistema Adaptativo (MAA), Modelo de Aprendizaje Personalizado (MAP), Sistema de Aprendizaje Adaptativo (SAA) o Sistema de Aprendizaje Personalizado (SAP), elaboración propia (2022).

Figura 5

Estrategias Utilizadas en la Construcción del Objeto de Estudio



Nota. Dimensión "Objeto de Estudio", distribución de los trabajos analizados con base en el objeto de estudio, elaboración propia (2022).

Figura 6

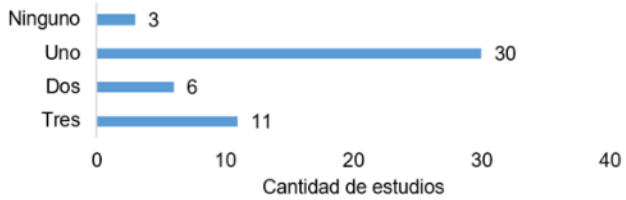
Distribución de Estudios de acuerdo con el Contexto de la Investigación

Nota. Dimensión "Objeto de Estudio". Los contextos en los que se llevaron a cabo los estudios analizados se clasifican en Escolar (E) y Laboral (L), elaboración propia (2022).

Para la dimensión "Estructura", en cuanto a los módulos (tratándose de un sistema) o submodelos (tratándose de un modelo) MA, MDOM y MAD, los resultados fueron los siguientes: a) 30 estudios consideraron un solo componente, de los cuales 7 se enfocaron únicamente en MA, 12 exclusivamente en MDOM, mientras que 11 solo en MAD; b) seis estudios consideraron dos componentes, 1 MDOM y MAD, 2 MA y MAD, y 3 MA y MDOM; y por último c) 11 estudios consideraron los tres componentes mientras que 3 no utilizaron ninguno (Figuras 7 y Figura 8). Cabe señalar que los Modelos y Sistemas de Aprendizaje Personalizado no disponen de un Módulo de Adaptabilidad.

Figura 7

Número de Trabajos Revisados de acuerdo con la Cantidad de Módulos o Submodelos Empleados

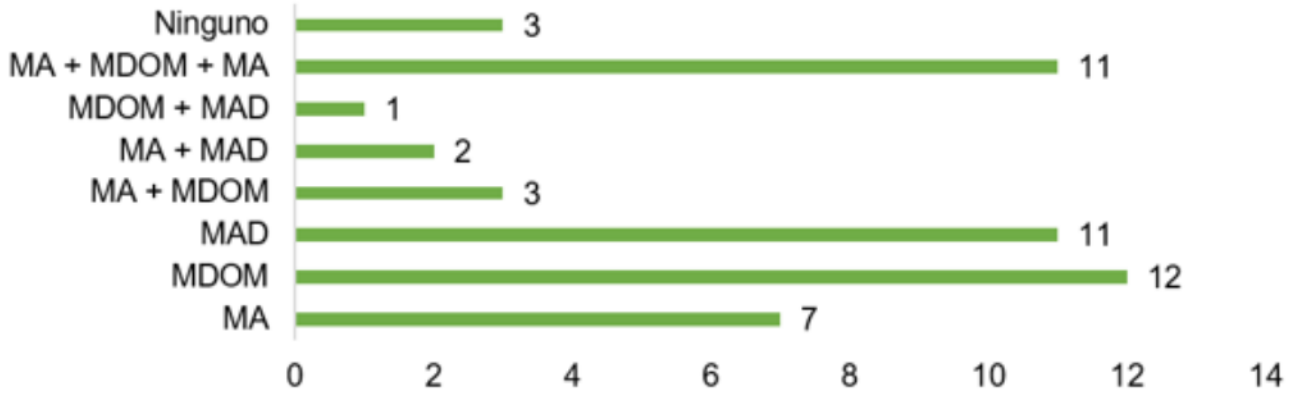


Nota. Dimensión "Estructura". Distribución de los trabajos revisados con respecto al número de componentes de estructura considerados, elaboración propia (2022).

Quintanar-Casillas, R. & Hernández-López, Ma.S. (2022). Modelos Tecnológicos de Aprendizaje Adaptativo Aplicados a la Educación: Revisión Sistemática de la Literatura. *Revista Tecnológica-Educativa Docentes 2.0*, 15(1), 41-58. https://doi.org/10.37843/rted.v15i1.308

Figura 8

Número de submodelos (MAA/MAP) o módulos (SAA/SAP) empleados en los trabajos revisados



Nota. Dimensión "Estructura". Distribución de los trabajos revisados con base en la estructura considerada en sus propuestas, elaboración propia (2022).

En cuanto a las técnicas de adaptabilidad empleadas, en el conjunto de estudios analizados las más utilizadas fueron las técnicas de inteligencia artificial con un total de 27. En 10 ocasiones se utilizó el modelo Felder- Silverman (FSLSM) para la detección del estilo de aprendizaje, mientras que el modelo VARK se utilizó en 5 estudios. La técnica de itinerario de aprendizaje apareció en siete trabajos, mientras que técnicas como las reglas, los objetos de aprendizaje adaptativo, la minería de datos y el uso de competencias requeridas aparecieron en dos trabajos cada una (Figura 9).

Figura 9

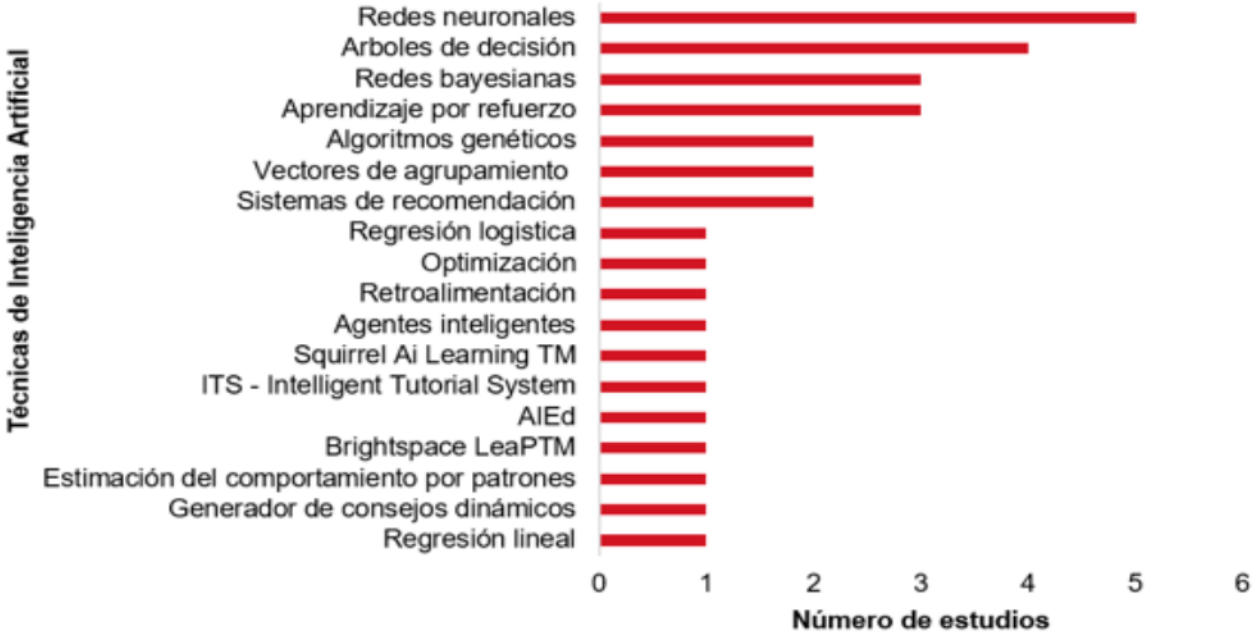
Técnicas de Adaptabilidad Empleadas en los Trabajos de Investigación Analizados

Nota. Dimensión "Estructura". Distribución de los estudios analizados de acuerdo con las técnicas de adaptabilidad utilizadas, elaboración propia (2022).

En específico, las técnicas de inteligencia artificial mayormente empleadas fueron las redes neuronales que aparecieron en 5 trabajos, seguidas de los árboles de decisión utilizados en cuatro trabajos. En un tercer peldaño se ubicaron el aprendizaje por refuerzo y las redes bayesianas, mismas que aparecieron cada una en tres trabajos de investigación revisados (Figura 10).

Figura 10

Técnicas de Inteligencia Artificial Empleadas en los Trabajos de Investigación Analizados



Nota. Dimensión "Estructura". Hay estudios que aplican más una técnica de IA, elaboración propia (2022).

De acuerdo con el análisis realizado a las estructuras de los trabajos revisados, 11 de estos estudios integran en sus propuestas las estructuras más robustas al considerar los tres componentes definidos, es decir los submodelos o módulos de aprendiz, dominio y adaptación. A continuación, se describen de forma breve las estructuras utilizadas en estas propuestas:

1. Gonçalves & García (2012) proponen un Sistema e-learning adaptativo hipermedia el cual emplea un sistema de recomendación basado en contenidos e integrado por cuatro subsistemas diferentes: subsistema de gestión de objetos de aprendizaje y diseño de aprendizaje; subsistema de gestión del conocimiento; subsistema de adaptación; y subsistema de visualización y presentación. El subsistema de adaptación toma en consideración las características del estudiante por lo que en este se integra el módulo del aprendiz. Los objetos de aprendizaje son buscados y seleccionados de diferentes sitios de Internet a través de un proceso de minería de datos.
2. Ahmad et al. (2013) proponen un sistema para la detección automática de estilos de aprendizaje. Esta detección se realiza a través de un método basado en literatura que utiliza el modelo de Felder y Silverman y es el proceso principal del módulo del aprendiz. Este método identifica los estilos de aprendizaje según el comportamiento del estudiante frente al sistema LMS utilizado, el cual contiene a los objetos de aprendizaje suministrados. De acuerdo con el comportamiento del estudiante frente al LMS, es como se liberan los objetos de aprendizaje logrando de esta manera la adaptabilidad.
3. Velandia (2015) propone un Recurso Educativo Digital Adaptativo (REDA) para la mejora de las HC utilizando como método el cuestionario de estilos de Felder y Silverman, y soportado en técnicas de IA y contenidos que resultan ser técnicas que interactúan con HC. El REDA se compone por una interfaz multimedia y una base de datos apoyada en técnicas de minería de datos y sistemas de decisiones, y tres módulos: el módulo del estudiante que contiene la base de datos con información que es

obtenida a partir del test de EA del estudiante; el módulo dominio, donde se encuentra toda la información de los contenidos que van a ser adaptados a partir del análisis de las falencias que tiene el usuario y de las mejoras alcanzadas; y el módulo pedagógico donde se define los procesos de adaptabilidad basados en técnicas de IA a partir de procesos estadísticos de minería de datos.

4. Qodad et al. (2016), proponen un sistema adaptativo basada en un modelo laboral, el cual se compone de cinco submodelos. Estos cinco componentes son: dominio, compuesto por los Objetos de Aprendizaje (OA), entendidos como materiales educativos digitales interactivos, accesibles, e incluso amenos (Carrillo et al., 2019); estudiante, basado en el conocimiento y estilo del estudiante; trabajo, representado por los elementos de la oferta laboral; instrucción que representa la pedagogía utilizada; y adaptabilidad para los OA específicos de acuerdo con cada EA. El modelo propuesto se basa en el concepto de instrucción diferenciada tomando sus elementos: el contenido, el estilo de aprendizaje, el producto (evaluación del estudiante) y el ambiente de aprendizaje; y utiliza el modelo de Felder y Silverman para la detección del estilo de aprendizaje. En el artículo no se especifica una técnica de inteligencia artificial para el submodelo de adaptación, sin embargo, de forma concluyente sugiere el uso de una Red Bayesiana en trabajos futuros que se basen en el sistema propuesto.
5. Salazar-Ospina et al. (2017) proponen un sistema de recomendación adaptativo que utiliza el filtrado colaborativo. Dentro de su estructura se contempla un módulo con el perfil del estudiante (módulo del aprendiz) y un módulo del objeto de aprendizaje. La adaptabilidad se logra a través de diferentes agentes (usuario, interfaz, evaluador, de recomendación, para los repositorios local y remoto y el agente central o coordinador).
6. Balasubramanian & Margret (2018) proponen un modelo para la detección de los EA, el cual se advierte como innovador y detallado, ya que no se basa en estilos de aprendizaje predefinidos como los estilos de Felder y Silverman, sino que detecta para cada sujeto de estudio su estilo de aprendizaje en particular. Otro diferenciador con respecto al resto de los trabajos estudiados es que consideran un enfoque netamente cuantitativo en cuanto a sus resultados los cuales se parcializan para cada uno de los submodelos que lo integran.
7. Firdausiah, Yusofa & Hoirul (2019) proponen un modelo de aprendizaje personalizado basado en un algoritmo de aprendizaje profundo que realiza una analítico del comportamiento del estudiante logrando de esta manera la adaptabilidad. El modelo propuesto se integra por cuatro componentes: componente basado en competencias en donde se encuentran los objetos de aprendizaje; individualizado, en el que se tiene el modelo del aprendiz; diferenciado; y adaptativo.
8. Hassan et al. (2019) proponen un modelo de aprendizaje adaptativo que utiliza la gamificación en un sistema e-learning detectando los estilos de aprendizaje de acuerdo con el modelo de Felder y Silverman. Se integra por cuatro submodelos: del estudiante, del estilo de aprendizaje (aprendiz), dimensión de interacción y aprendizaje (adaptación) y adaptación de comportamiento que contiene a los objetos de aprendizaje o juegos seleccionados para el estudiante.
9. Koch et al. (2020) proponen un sistema que utiliza herramientas adaptativas de inteligencia artificial para la enseñanza del idioma inglés. El sistema se basa en una herramienta desarrollada por la empresa Pearson denominada AIEd y se integra por tres submodelos (aprendiz, dominio y pedagógico) estos tres

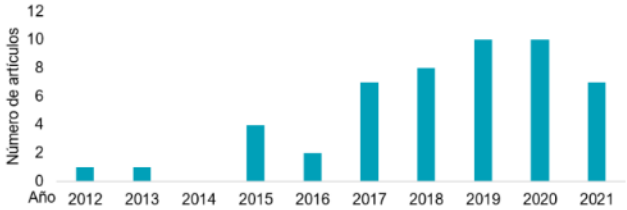
submodelos convergen e interactúan con los algoritmos de inteligencia artificial provistos por AIEd.

10. Troussasa, Krouskaa & Sgouropouloub (2020), proponen un sistema de colaboración y personalización mediante técnicas de lógica difusa para el aprendizaje ludificado en la educación superior. El sistema se integra por tres módulos: evaluación del conocimiento que contiene al modelo del dominio del conocimiento (objeto de aprendizaje); el módulo ejecutor de la recomendación para la colaboración en juego grupal el cual contiene la representación vectorial de los aprendices y el generador dinámico de consejos el cual se basa en el modelo de lógica difusa que se concibe como la técnica de adaptación.
11. Ríos-Rodríguez, Román-Cao & Pérez-Medinilla (2021) realizan una intervención educativa con un sistema de aprendizaje adaptativo previamente desarrollado denominado APA-Prolog. Es sistema APA-Prolog se compone de tres módulos (denominados modelos en la investigación): estudiante, adaptación y dominio. Para lograr la adaptación utilizan cinco agentes inteligentes: 1) Adis, el cual verifica el perfil del estudiante; 2) Tivo, el cual contiene los objetos de aprendizaje sugeridos y el seguimiento del progreso del estudiante; 3) Teo, realiza decisiones sobre los objetos de aprendizaje; 4) Tica, verifica el progreso del estudiante y emite recomendaciones y 5) Eva, se enfoca en la evaluación del estudiante.

De la muestra de estudios analizada, se observa una tendencia creciente en las investigaciones sobre el aprendizaje adaptativo. Más de la mitad de los trabajos revisados corresponden a los últimos tres años (Figura 11). Lo anterior se atribuye a la evolución y uso de las técnicas de inteligencia, lo cual se ha intensificado en los últimos años.

Figura 11

Distribución de los Trabajos Analizados por Año de Publicación



Nota. Distribución cronológica del número de estudios analizados, elaboración propia (2022).

De igual forma, los resultados muestran que el mayor número de estudios analizados optaron por el desarrollo de un sistema de aprendizaje adaptativo (31), permitiendo de esta forma aplicar una intervención en un ambiente escolar en 29 casos y en un ambiente laboral en 2 únicamente. Lo anterior, indica una amplia predominancia de sistemas de aprendizaje adaptativo en entornos académicos contra una casi nula investigación de los entonos laborales. Un caso similar se observa en las propuestas de modelos de aprendizaje adaptativo de las cuales solo 3 de 16 se enfocaron en contextos laborales.

Los resultados también develan una predominancia de estudios sobre aprendizaje adaptativo con respecto a aquellos que indagaron sobre el aprendizaje personalizado. En lo que respecta al origen la mayor parte de los estudios considerados para el estudio (21) corresponden a la región. La segunda región con mayor número de estudios considerados es Asia (11) mientras que la tercera fue Europa.

Discusión

A través del análisis de contenido y apoyado en la metodología PRISMA como directriz para la construcción de revisiones sistemáticas, es posible conocer el estado del arte de los modelos o sistemas tanto adaptativos como personalizados. Los resultados correspondientes a la dimensión "Objeto de Estudio" indican que: 1) Los estudios analizados se centraron en primera instancia en los objetos de aprendizaje, que es consecuencia del problema al que los profesionales de la educación se enfrentan hoy en día en elaborar diseños instruccionales adecuados; 2) El segundo enfoque más frecuentemente abordado está orientado a la

Quintanar-Casillas, R. & Hernández-López, Ma.S. (2022). Modelos Tecnológicos de Aprendizaje Adaptativo Aplicados a la Educación: Revisión Sistemática de la Literatura. *Revista Tecnológica-Educativa Docentes 2.0*, 15(1), 41-58. https://doi.org/10.37843/rted.v15i1.308

construcción de plataformas LMS adaptativas, en otras palabras, en ofrecer una solución integral de aprendizaje a los estudiantes.

Dado que el aprendizaje adaptativo se apoya en el uso de herramientas tecnológicas se justifica que el segundo mayor enfoque de los estudios analizados hayan sido los sistemas LMS; y 3) Se ha explorado en mayor medida el estudio de los estilos de aprendizaje en comparación con las habilidades cognitivas, aunque este último tipo de investigaciones tiene una tendencia al alza en años recientes.

En lo que refiere a la dimensión "Estructura", se identificaron tres componentes comúnmente empleados en los modelos o sistemas propuestos: el Módulo del Aprendiz (MAP), el Módulo del Dominio (MDOM) y el Módulo de Adaptabilidad (MAD). En el MAP se obtiene y almacena toda la información necesaria del perfil del estudiante, en el MDOM se integran los objetos de aprendizaje que se entregan al estudiante y en el MAD se utilizan las diferentes técnicas de adaptabilidad que permiten la correcta asignación de los objetos de aprendizaje a los estudiantes dadas sus características.

De acuerdo con los resultados obtenidos, se observa una mayor tendencia al empleo de técnicas de inteligencia artificial para lograr la adaptabilidad de los modelos o sistemas de aprendizaje adaptativo. De los trabajos analizados correspondientes a los últimos tres años, la mitad utilizan técnicas de inteligencia artificial; en los últimos dos años el 70% de los estudios revisados emplearon dichas técnicas, mientras que en el último año la totalidad de los trabajos realizado se basan en inteligencia artificial. Investigaciones similares a la desarrollada en el presente trabajo aparecen en los estudios de Afini et al (2018), Xie et al. (2019) y Zhai et al. (2021). Las discrepancias entre los trabajos mencionados y la presente investigación se exponen en las siguientes líneas.

En el trabajo de Afini et. al (2018), se identifican los perfiles individuales más comúnmente usados en el modelo del aprendiz, las técnicas aplicadas en la identificación de estos perfiles, y los problemas suscitados en el modelo de adaptación. Clasifican el modelo del aprendiz en tres tipos con base en la Taxonomía de Bloom: cognitivo, afectivo y mixto. De acuerdo con esta clasificación, los estilos de aprendizaje caen dentro del plano cognitivo, sin embargo, ubican a las habilidades cognitivas como un enfoque mixto, siendo que las investigaciones basadas en habilidades cognitivas no necesariamente involucran a las habilidades emocionales dentro de su objeto de estudio. Por lo anterior en la presente investigación se optó por tratar de forma independiente a las habilidades cognitivas de los estilos de aprendizaje.

Xie et al. (2019) definen las diferencias entre aprendizaje adaptativo y personalizado, aunque en sus resultados no muestran una distinción entre ambos tipos de estudios; en el presente trabajo se clasifican los estudios analizados con base en las diferencias previamente descritas. Por su parte, Zhai et al. (2021) identifican cuatro tendencias en la aplicación de las técnicas de inteligencia artificial pero no exponen cuáles fueron las técnicas empleadas en los trabajos analizados, como ocurre en el caso del presente estudio.

Dentro de los puntos coincidentes que la presente investigación mantiene con respecto a los resultados expuestos en los trabajos antes mencionados están los siguientes: a) en años recientes ha habido un incremento sostenido en la cantidad de estudios sobre aprendizaje adaptativo, b) uno de los abordajes más utilizado en el modelo del aprendiz es el de los estilos de aprendizaje, c) existe una escasa investigación sobre las habilidades cognitivas en su relación con el aprendizaje adaptativo, d) el uso de las técnicas de inteligencia artificial en los modelos y sistemas de aprendizaje adaptativo es muy frecuente y se observa en claro crecimiento, e) la distribución de la técnicas de inteligencia artificial empleadas es casi uniforme lo que hace difícil establecer un patrón de predominancia; no obstante, existe una tendencia emergente a un mayor uso de las redes neuronales y el aprendizaje por refuerzo.

Cabe señalar que, durante el proceso de búsqueda en los diferentes motores y bases de datos científicas, no se encontró ningún trabajo de revisión sobre modelos o sistemas de aprendizaje adaptativo generados en Hispanoamérica; los

trabajos de revisión encontrados se originaron en el sureste asiático. Lo anterior se debe a la escasa investigación que existe en la región hispana sobre esta temática. De los 50 trabajos analizados 16 se crearon en esta región, de los cuales 14 abordan problemáticas en torno al aprendizaje adaptativo y 2 sobre aprendizaje personalizado.

Dentro de las limitaciones que enfrentó la investigación están el acceder a trabajos no gratuitos y de alto costo, así como a estudios con acceso restringido. Por otro lado, la mayor parte de los estudios considerados para el análisis tiene su origen en Hispanoamérica, debido a que se priorizó la búsqueda y recuperación de trabajos originados en esta región. Sin embargo, cabe señalar que, dada su exigüidad, el ubicar e integrar trabajos elaborados en la región de Hispanoamérica que cumplieran con los criterios de inclusión se tradujo en una tarea laboriosa.

Otra de las dificultades que se tuvo durante el desarrollo de la investigación fue identificar el tipo de técnica de adaptabilidad, dada la amplia variedad de herramientas de IA disponibles en la actualidad. Los resultados alcanzados en la presente investigación abren la pauta para que trabajos futuros enfoquen en mayor medida al estudio de las habilidades cognitivas y competencias del estudiante que en los modelos clásicos de estilos de aprendizaje, en los que ya se define un conjunto delimitado dentro del cual se categoriza al sujeto de estudio.

Lo anterior permite establecer nuevos paradigmas que consideren: a) estilos de aprendizaje mixtos y difusos con el apoyo de herramientas de Inteligencia Artificial de agrupamiento difuso (*fuzzy clustering*) como el algoritmo Fuzzy C-Means; b) el seguimiento del progreso cognitivo del estudiante a través de estudio individualizado de los estilos de aprendizaje y el desarrollo de las habilidades cognitivas apoyado en tutores-agentes como los utilizados en el aprendizaje por refuerzo; y c) el uso de sistemas de recomendación para asignar los objetos de aprendizaje de mayor influencia en el aprendizaje del estudiante.

Conclusiones

La presente investigación contribuyó en la construcción del estado del arte de los Modelos y Sistemas de Aprendizaje Adaptativo y Aprendizaje Personalizado para un periodo de tiempo comprendido entre los años 2012 y 2022 tomando en consideración a los autores de origen hispanoamericano. Como producto del análisis desarrollado en la investigación, se observó que los modelos o sistemas de aprendizaje adaptativo para la solución de problemáticas educativas ha generado beneficios a las instituciones en donde se ha llevado a cabo su implementación, ya que los estudiantes tienen acceso a una modalidad de enseñanza que permite la entrega dinámica de los recursos de aprendizaje con los que mejor se desempeñan.

Además, la incorporación de técnicas de Inteligencia Artificial en el diseño de estos modelos ha permitido al aprendiz contar con cierto grado de autonomía y acceder a esquemas de autoaprendizaje, e incluso de autoevaluación. Por otra parte, el estudio devela una baja producción de propuestas de modelos y sistemas de aprendizaje adaptativo en la región hispanoamericana, dadas las razones expuestas en la discusión de los resultados, por lo que se infiere que este tipo de tecnologías educativas se aplican en menor medida a los estudiantes de la región.

Dentro de los campos poco explorados o nichos de investigación sobre aprendizaje adaptativo aplicado a la Educación en lo que refiere al objeto de estudio, se encuentra el campo laboral; en tanto que la aplicación de la lógica difusa para lograr la adaptabilidad de los Modelos de Aprendizaje Adaptativo también se advierte como una técnica poco empleada que puede ser ampliamente explotada en investigaciones que consideren que los individuos tienen un cierto grado de pertenencia a uno o varios estilos de aprendizaje.

Se propone entonces, que las problemáticas de aprendizaje que enfrentan los investigadores y académicos, sobre todo aquellas que tienen origen en la región hispanoamericana, sean abordadas en la medida de lo posible desde el enfoque del aprendizaje adaptativo. Además, se sugiere que estas futuras investigaciones: a) propongan mejoras a las técnicas de adaptabilidad que ya han sido empleadas en trabajos previos, o bien, introduzcan una nueva técnica que resulte innovadora; b) tomen en consideración para el diseño de la estructura del modelo propuesto, un esquema similar al del modelo MAP-MDOM-

MAD dados sus resultados positivos; y c) utilicen el enfoque orientado a las habilidades cognitivas y los estilos de aprendizaje, lo cual permite generar un modelo del perfil del estudiante más cercano a la realidad.

Referencias

Afini N., Shuib L., Md Nasir H.N., Bimba A., Idris N. & Balakrishnan V. (2018). Identification of personal traits in adaptive learning environment: Systematic literature review, *Computers & Education. 130.* 168-190. https://doi.org/10.1016/j.compedu.2018.11.005

Ahmad, N., Tasirb, Z., Kasimc, J., & Sahata, H. (2013). Automatic Detection of Learning Styles in Learning Management Systems. *Procedia - Social and Behavioral Science, 103*, 181 – 189. http://doi.org/10.1016/j.sbspro.2013.10.324

Alvarado, Y., Beltrán, M., Escobar, M., Espinosa, J., Gamboa, F., Gamboa, C., & Martínez, L. (2013). *Estado del arte sobre el concepto de trabajo cooperativo en el marco del desarrollo de habilidades de lenguaje con proyección social*. [Tesis de Especialidad, Universidad de San Buenaventura]. Repositorio de la Universidad de San Buenaventura, Colombia. http://bibliotecadigital.usb.edu.co/handle/10819/3224

Arias-Gómez, J., Villasís-Keever, M., & Miranda-Novales, M. (2016). El protocolo de investigación III: la población de estudio. *Alergia Mex. 63*(2), 201-206. https://www.redalyc.org/pdf/4867/486755023011.pdf

Balasubramanian, V., & Margret, S. (2018). Learning style detection based on cognitive skills to support adaptive learning environment – A reinforcement approach. *Ain Shams Engineering Journal*, *9*, 895-907. https://doi.org/10.1016/j.asej.2016.04.012

Battou, A. (2017). Designing an Adaptive Learning System Based on a Balanced Combination of Agile Learner Design and Learner Centered Approach. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS) 37*(1), 178-186. https://www.asrjetsjournal.org/index.php/American_Scientific_Journal/article/view/3519/1292

Brateanu, A. Strang, T., Garber, A., Mani, S., Spencer, A., Spevak, B., Thomascik, J., Mehta, N., & Colbert, C. (2019). Using an Adaptive, Self-Directed Web-Based Learning Module to Enhance Residents' Medical Knowledge Prior to a New Clinical Rotation. *Medical Science Educator, 29,779*–786. https://doi.org/10.1007/s40670-019-00772-8

Chávez, A. (2019). Uso de la tecnología en el aprendizaje adaptativo: propuesta para favorecer la resolución de problemas matemáticos en primaria. *Educando para educar. 20*(37) 71-89. https://beceneslp.edu.mx/ojs2/index.php/epe/article/view/50

Duque-Méndez, N., Ovalle-Cerranza, D. & Carrillo-Ramos, A. (2020). Sistema basado en reglas para la generación personalizada de curso virtual. TecnoLógicas. 23(47), 229-242. DOI: https://doi.org/10.22430/22565337.1494

El Asame, M., Wakrim, M., & Battou, A. (2018). An Improved Competency Meta-model for Adaptive Learning Systems. En M. Ben Ahmed and A. A. Boudhir (Eds.). *Springer International Publishing AG.* 37, 411–419. https://doi.org/10.1007/978-3-319-74500-8_38

Fernández, L., Ramírez, J. & Blanco, A. (2017). Uso de la Adaptabilidad en el aprendizaje de la Expresión Gráfica. *IV Congreso Internacional sobre Aprendizaje, Innovación y Competitividad (CINAIC 2017).* 515-519. http://doi.org/10.26754/CINAIC.2017.000001_108

Firdausiah, A., Yusofa, N., & Hoirul, A. (2019). Personalized Learning Model based on Deep Learning Algorithm for Student Behaviour Analytic. *Procedia Computer Science, 163*, 125–133. http://doi.org/10.1016/j.procs.2019.12.094

Gómez, M., Galeano, C., & Jaramillo, D. (2015). El estado del arte: una metodología de investigación. *Revista Colombiana de Ciencias Sociales*, *6*(2), 423-442. https://www.redalyc.org/pdf/4978/497856275012.pdf

Gonçalves, H., & García, F. (2012). *Adaptive Hypermedia Knowledge Management eLearning System (AHKME) – Management and Adaptation of Learning Objects and Learning Design in a Web-Based Information System Towards the Third Generation of Web* [Tesis de Doctorado, Universidad de Salamanca]. Repositorio de GRIAL. https://repositorio.grial.eu/handle/grial/186

González, J. L. (2020). Técnicas e instrumentos de investigación científica. CONCYTEC. http://repositorio.concytec.gob.pe/handle/20.500.12390/2238

González, M., Benchoff, D., Huapaya, C., & Remon, C. (2017). Aprendizaje Adaptativo: Un Caso de Evaluación Personalizada. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología.* 65-72. https://dialnet.unirioja.es/servlet/articulo?codigo=6068136

González, M., Becerra, J., & Olmos, J. (2018). Promoción de la autogestión a través de objetos de aprendizaje adaptativos en alumnos de educación superior. *Edutec. Revista Electrónica De Tecnología Educativa,* (63), 15-28. https://doi.org/10.21556/edutec.2018.63.1037

Guevara, C., & Aguilar, J. (2019). Caracterización Semántica de Objetos de Aprendizaje Adaptativos mediante una Ontología de Tareas. *Revista Ciencia e Ingeniería. 40* (2), 195-208. https://www.redalyc.org/journal/5075/507567825010/html/

Guevara, C., & Aguilar, J. (2019). Modelo Ontológico del Estándar LOM Extendido para la Gestión de Objetos de Aprendizaje Adaptativos En Avances y retos de la ciencia e ingeniería. https://www.researchgate.net/publication/336019607.

Hassan, M., Habiba, U., Majeed, F., & Shoaib, M. (2019). Adaptive gamification in e-learning based on students' learning styles. *Interactive Learning Environments*.29(4), 545-565. https://doi.org/10.1080/10494820.2019.1588745

Hernández-Sampieri, R., & Mendoza, C. (2018). *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta*. Editorial Mc Graw Hill Education, ISBN: 978-1-4562-6096.

Hünicken, L., González, A., Haag, M., & Ruppel, D. (2020). Gamificación y aprendizaje adaptativo en la universidad. En M. R. De Gusti. (Ed.). *Actas I Congreso Internacional de Ingeniería Aplicada a la Innovación y Educación y Asamblea General de ISTEC 2019*. Repositorio institucional de la UNLP.

Jiménez, I., Fernández, O. E., & Almenárez, F. T. (2020). Diseño pedagógico adaptativo para el desarrollo de MOOC: una estrategia para el desarrollo de competencias en contextos corporativos. *Revista Electrónica de Investigación Educativa*, *22* (16), 1- 18. https://doi.org/10.24320/redie.2020.22.e16.2192

Kellman, P., & Krasne, S. (2018). Accelerating expertise: Perceptual and adaptive learning technology in medical learning. *Medical Teacher. 40* (8), 797-802. https://doi.org/10.1080/0142159X.2018.1484897

Koch, H., De Azevedo, A., Sebastiany, M., & Cortina, A. (2020). Artificial intelligence adaptive learning tools: the teaching of English in focus. *Brazilian English Language Teaching Journal. 11*(2). 1-19. http://dx.doi.org/10.15448/2178-3640.2020.2.38749

Koshova, S. (2020). Specifics of adaptive learning in the field of public governance. *Public Administration Aspects, 8*(4), 107-117. https://doi.org/10.15421/152085

Krechetov, I., & Romanenko, V. (2020). Implementing the Adaptive Learning Techniques. *Voprosy obrazovaniya / Educational Studies Moscow, 2*. 252–277. https://vo.hse.ru/data/2020/06/17/1607977770/Krechetov.pdf

Kuniyoshi, K., & Kurahashi, S. (2020). Simulation of learning effects of adaptive learning. *Procedia Computer Science*, *176,* 2164-2172. https://doi.org/10.1016/j.procs.2020.09.253

Leris, D., Vea, F., & Velamazán, A. (2015). Aprendizaje adaptativo en Moodle: tres casos prácticos. *EKS. 16*(4) 138-157. http://dx.doi.org/10.1109/FINTDI.2011.5945972

Lopera, J., Ramírez, C., Zuluaga, M., & Ortiz, J. (2010). El método analítico como método natural. *Nómadas. Revista Crítica de Ciencias Sociales y Jurídicas. 25* (1). https://revistas.ucm.es/index.php/NOMA/article/view/NOMA1010140327A/25986

Liu, M., McKelroy, E., Corliss, S., & Carrigan, J. (2017). Investigating the effect of an adaptive learning intervention on students' learning. *Education Tech Research Dev. 65,* 1605–1625. http://doi.org/10.1007/s11423-017-9542-1

Lwande, C., Muchemi, L., & Oboko, R. (2021). Identifying learning styles and cognitive traits in a learning management system. *Heliyon. 7*. https://doi.org/10.1016/j.heliyon.2021.e07701

May, H. (2016). Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities. *Computers in Human Behavior. 55*. 1185-1193. http://dx.doi.org/10.1016/j.chb.2015.02.014

Mira-Giménez, M. (2017). COMALAT: Aprendizaje de idiomas adaptativo y móvil. *Cultura y Educación. 29* (4), 863–870. https://doi.org/10.1080/11356405.2017.1370820

Otero, A., Rivera, W., Pedraza, C., & Canay, J. (2019). TIC para la educación: sistema adaptativo basado en mecanismos de aprendizaje automático para la apropiación de tecnologías en estudiantes de educación media. *Telos, 21*(3), 526-543. https://www.redalyc.org/articulo.oa?id=99360575003

Otzen, T., & Manterola, C. (2017). Técnicas de muestreo sobre una población a estudio. *Int. J. Morphol., 35*(1), 227-232. http://www.intjmorphol.com/wp-content/uploads/2017/04/art_37_351.pdf

Page, M., McKenziea, J., Bossuyt, P., Boutron, I., Hoffmann, T., Mulrow, C., Shamseer, L., Tetzlaff, J., Akl, E., Brennan, S., Chou, R., Glanville, J., Grimshaw, J., Hróbjartsson, A., Lalu, M., Li, T., Lodero, E., Mayo-Wilson, E., McDonald, S., …, & Moherv. D. (2021). Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas. *Rev Esp Cardiol. 74* (2), 790-799. https://www.revespcardiol.org/es-declaracion-prisma-2020-una-guia-articulo-S0300893221002748

Peña, N., González, C., & Mendoza, M. (2015). Diagnóstico de estilos de aprendizaje para favorecer la personalización de materiales educativos mediante redes bayesianas: servicios web para Moodle. *Revista Virtual Universidad Católica del Norte*, *46*, 4-20. http://www.redalyc.org/articulo.oa?id=194242285002

Qodad, A., Seghroucheni, Y., Al Achhab, M., El Yadari, M., El Kenz, A., & Benyoussef, A. (2016). An Adaptive Learning System based on a Job Model, the Differentiated Instruction and Felder and Silverman's Learning Styles Model. *4th IEEE International Colloquium on Information Science and Technology.* IEEE. https://doi.org/10.1109/CIST.2016.7805100

Ramos, C. (2015). Los paradigmas de la Investigación Científica. *Av. psicol. 23*(1), 9-17 http://www.unife.edu.pe/publicaciones/revistas/psicologia/2015_1/Carlos_Ramos.pdf

Rasheed, F., & Wahid, A. (2021). Learning style detection in E-learning systems using machine learning techniques. *Expert Systems With Applications. 174*. https://doi.org/10.1016/j.eswa.2021.114774

Real-Fernández, A., Molina-Carmona, R., & Llorens-Largo, F. (2017). Aprendizaje adaptativo basado en competencias y actividades. *IV Congreso Internacional sobre Aprendizaje, Innovación y Competitividad*. Universidad de Zaragoza. https://doi.org/10.26754/CINAIC.2017.000001_017

Renn, B., Arean, P., Raue, P., Aisenberg, E., Friedman1, E., & Popovic, Z. (2020). Modernizing Training in Psychotherapy Competencies With Adaptive Learning Systems: Proof of Concept. *Research on Social Work Practice*. https://doi.org/10.1177%2F1049731520964854

Ríos-Rodríguez, L., Román-Cao, E., & Pérez-Medinilla, Y. (2021). La dirección del trabajo independiente mediante el ambiente de enseñanza-aprendizaje adaptativo APA-Prolog. *Revista Electrónica Educare*. *25*(1), 1-22. http://doi.org/10.15359/ree.25-1.11

Roque-Alayón, Y., Sánchez-Díaz, A., & López-Padrón, A. (2015). Entorno de Aprendizaje Personalizado (PLE) para la asignatura de Investigación de Operaciones en Ingeniería Agrícola. *Revista Ciencias Técnicas Agropecuarias*, *25*(1). https://www.redalyc.org/journal/932/93242698010

Rothman, D. (2018). *Artificial Intelligence by example*. Packt Publishing.

Sabita, R. (2021). Everything you need to know about Machine Learning. *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2021/03/everything-you-need-to-know-about-machine-learning/

Salazar-Ospina, O., Rodríguez-Marín, P., Ovalle-Carranza, D., & Duque-Méndez, N. (2017). Interfaces adaptativas personalizadas para brindar recomendaciones en repositorios de objetos de aprendizaje. *Tecnura*, *21* (53), 107-118. https://doi.org/10.14483/udistrital.jour.tecnura.2017.3.a07

Serrao-Neumann, S., Cox. M., & Low, D. (2018). Bridging Adaptive Learning and Desired Natural Resource Management Outcomes: Insights from Australian Planners. *Planning Practice & Research. 34*(2), 149–167. https://doi.org/10.1080/02697459.2018.1549188

Sfenrianto, S., Hartarto, Y., & Akbar. H. (2018). An Adaptive Learning System based on Knowledge Level for English Learning. *iJET*. *13*(12). https://doi.org/10.3991/ijet.v13i12.8004

Shawky, D., & Badawi, A. (2018). A Reinforcement Learning-Based Adaptive Learning System. En A. E. Hassanien et al. (Eds.): International Conference on Advanced Machine Learning Technologies and Applications. *Springer*. https://doi.org/10.1007/978-3-319-74690-6_22

Sheeba, T., & Krishnan. R. (2019). Automatic Detection of Students Learning Style in Learning Management System. *Springer*. 45-53. https://doi.org/10.1007/978-3-030-01659-3_7

Sokol, V., Bronin, S., Karnaukh, V., & Bilov, M. (2020). Developing Adaptive Learning Management Application for Project Team in IT-Industry. *Bulletin of National Technical University "KHPI". Series: System Analysis, Control and Information Technologies*. *1*(3), 97-105. https://doi.org/10.20998/2079-0023.2020.01.17

Tabares-Morales, V., Duque-Méndez, N., & Fabregat-Gesa, R. (2020). Raim: framework para la inclusión adaptativa en entornos educativos para todos. TecnoLógicas. 23(47), 179-196. https://doi.org/10.22430/22565337.1495

Troussasa, C., Krouskaa, A., & Sgouropouloub, C. (2020). Collaboration and fuzzy-modeled personalization for mobile game-based learning in higher education. *Computers & Education. 144*. https://doi.org/10.1016/j.compedu.2019.103698

Velandia, J. (2015). *Mejoras en habilidades cognitivas con el apoyo de un recurso educativo digital adaptativa.* [Tesis de Maestría, Universidad de la Sabana]. Repositorio de la Universidad de Salamanca. https://intellectum.unisabana.edu.co/handle/10818/22992

Véliz, A., Madrigal, O., & Correa, V. (2021). Aprendizaje adaptativo basado en Simuladores de Realidad Virtual. *Revista Cubana de Ciencias Informáticas, 15*(2) 138-157. https://www.redalyc.org/articulo.oa?id=378367420008

Wang, S., Christensen, C., & Cui, W. (2020). When adaptive learning is effective learning: comparison of an adaptive learning system to teacher-led instruction. *Interactive Learning Environments*. 1–11, https://doi.org/10.1080/10494820.2020.1808794

Xie, H., Chu, H., Hwang, G., & Wang, Ch. (2019). Trends and development in technology-enhanced adaptive/ personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers and Education. 140*. https://doi.org/10.1016/j.compedu.2019.103599

Yaplea, Z., & Yu, R. (2019). Fractionating adaptive learning: A meta-analysis of the reversal learning paradigm. *Neuroscience and Biobehavioral Reviews, 102*. 85-94. https://doi.org/10.1016/j.neubiorev.2019.04.006

Yu, J., Wang, Y., Zhong, Q., Luo, G., Mao, Y., Sun, K., Feng, W., Xu, W., Cao, S., Zeng, K., Yao, Z., Hou, L., Lin, Y., Li, P., Zhou, J., Xu., B., Li, J., Tang, J., & Sun., M. (2021). MOOCCubeX: A Large Knowledge-centered Repository for Adaptive Learning in MOOCs. *CIKM '21* 4643-4652. https://dl.acm.org/doi/pdf/10.1145/3459637.3482010

Zhai, X., Chu. X., Sing, C., Yung, M., Istenic, A., Spector, M., Liu, J., Yuan, J., & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*. https://doi.org/10.1155/2021/8812542

Zhikharev, A., Deeney, A., Igrunova, S., Klyuchnikov. D., & Frolova, A. (2021). To the development of intelligent adaptive learning systems. *Journal of Physics: Conference Series 2060*. http://doi.org/10.1088/1742-6596/2060/1/012012

Zulfiani, Z., Suwarna, I., & Miranto. S. (2018). Science education adaptive learning system as a computer-based science learning with learning style variations. *Journal of Baltic Science Education. 17*(4), 711-727. http://www.scientiasocialis.lt/jbse/?q=node/695

Array 19 (2023) 100301

Contents lists available at ScienceDirect

Array

journal homepage: www.sciencedirect.com/journal/array

Optimizing student engagement in edge-based online learning with advanced analytics

Rasheed Abdulkader [a], Firas Tayseer Mohammad Ayasrah [b], Venkata Ramana Gupta Nallagattla [c], Kamal Kant Hiran [d], Pankaj Dadheech [e], Vivekanandam Balasubramaniam [f], Sudhakar Sengan [g,*]

[a] Electrical Engineering Department, Imam Mohammad Ibn Saud Islamic University (IMSIU), Saudi Arabia
[b] College of Education, Humanities and Science, Al Ain University Al Ain, United Arab Emirates
[c] Department of Computer Science and Engineering, Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, 520007, India
[d] Department of Computer Science and Engineering, Lincoln University College, Malaysia
[e] Department of Computer Science & Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan (SKIT), Jaipur, Rajasthan, 302017, India
[f] Faculty of Computer Science and Multimedia, Lincoln University College, Wisma Lincoln, No. 12-18, Jalan SS 6/12, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia
[g] Department of Computer Science and Engineering, PSN College of Engineering and Technology, Tirunelveli, 627152, Tamil Nadu, India

ARTICLE INFO

Keywords:
Deep learning
VGG-19
Online learning
Edge computing
Face detection
DLIP

ABSTRACT

Edge-Based Online Learning (EBOL), a technique that combines the practical, hands-on approach of EBOL with the convenience of Online Learning (OL), is growing in popularity. But accurately monitoring student engagement to enhance teaching methodologies and learning outcomes is one of the difficulties of OL. To determine this challenge, this paper has put forth an Edge-Based Student Attentiveness Analysis System (EBSAAS) method, which uses a Face Detection (FD) algorithm and a Deep Learning (DL) model known as DLIP to extract eye and mouth landmark features. Images of the eye and mouth are used to extract landmarks using DLIP or Deep Learning Image Processing. Landmark Localization pre-trained models for Facial Landmark Localization (FLL) are one well-liked DL model for facial landmark recognition. The Visual Geometry Group-19 (VGG-19) learning model then uses these features to classify the student's level of attentiveness as fatigued or focused. Compared to a server-based model, the proposed model is developed to execute on an Edge Device (ED), enabling a swift and more effective analysis. The EBOL achieves 95.29% accuracy and attains 2.11% higher than existing model 1 and 4.41% higher than existing model 2. The study's findings have shown how successful the proposed method is at assisting teachers in changing their teaching methodologies to engage students better and enhance learning outcomes.

1. Introduction

Numerous facets of daily life are transforming due to the Internet of Things (IoT). The ubiquitous nature of IoT technologies sets them apart from earlier advancements and encourages the development of intelligent and autonomous solutions [1]. A significant strategic technology trend is the development of the IoT [2]. The conceptual model in the novel approach to learning is predicated on having multiple sensors that connect the physical and virtual machines. Sensors are embedded into any object, which supports this change in basic assumptions. Next, embedded sensors use Machine-to-Machine (M2M) communication to connect billions of devices to the Internet [3]. Face Detection (FD) using Deep Learning is a popular Computer Vision (CV) task that involves identifying and locating human faces in images or video streams. DL techniques, specifically Convolutional Neural Networks (CNNs), have proven highly effective for FD tasks.

On the whole, the physical world is going online quickly. The IoT is causing worldwide interest and apprehension as it develops swiftly and

* Corresponding author. .
E-mail addresses: rmabdulkader@imamu.edu.sa (R. Abdulkader), firas.ayasrah@aau.ac.ae (F. Tayseer Mohammad Ayasrah), nallagattla@gmail.com (V.R.G. Nallagattla), pdf.drkkhiran@lincoln.edu.my (K. Kant Hiran), pankajdadheech777@gmail.com (P. Dadheech), vivekanandam@lincoln.edu.my (V. Balasubramaniam), sudhasengan@gmail.com (S. Sengan).

https://doi.org/10.1016/j.array.2023.100301
Received 21 April 2023; Received in revised form 19 June 2023; Accepted 24 June 2023
Available online 28 June 2023
2590-0056/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

has become an ever-growing topic [4]. Several signs show that the IoT's potential can transform many industries, including Higher Education (HE) institutions, particularly universities [5].

As a result of the internet's rootedness in our schools, e-Learning is becoming a widely used practice in many school education systems. But the IoT has many educational uses, and this disruption has enormous ramifications. Schools may now increase the safety of their campuses, monitor essential resources, and improve information access in the learning platform due to the development of smartphone technology and the IoT [6]. Instead of the Traditional Lesson Plans (TLP) of the past, this technology for developing "Smart Lesson Plans (SLP)" by teachers [7]. From paper books, college students significantly progressively switch to handheld devices. Now, the learning students are done at their own pace and have a comparable learning experience in their homes and classrooms because they have access to all the required material at their fingertips.

Additionally, when this tendency increases students' accessibility, it helps professors teach more effectively. The proliferation of linked technology has eliminated the need for tutors to grade tests on paper manually or conduct other daily activities. Instead, tutors concentrate on customized learning to help their students the best. Professors can collect data on their students using cloud-connected devices and then decide the one that requires the most outstanding individualized care and attention. These data support tutors in enhancing student engagement and modifying their lecture plans for upcoming classes.

On the one hand, due to the limitations of various hardware facilities, today's online instructional System (OTS) cannot apply and manage the teaching resources. On the other hand, while Cloud Computing (CC) is becoming increasingly widespread, it typically uses centralized management, making it impossible to guarantee online classes' low delay and reliability and even posing some data security challenges [8]—Edge Computing (EC) technology to enhance the existing OTS. EC technology integrates a network, a computer system, storage spaces, and application capabilities on free-standing platforms. It starts on edge to provide fast network service responses and fulfill industry standards for real-time business, application intelligence, security, and privacy. Between the physical subject and industrial connection, or even above the physical subject, is where the edge computer is situated. Data processing and analysis with EC is faster than with CC [9].

Online Learning (OL) is being used more in schools for readings, articles, video lectures, virtual learning environments, and timed tasks [10], although there are still problems in the online setting in traditional classrooms. Some learning resources given to the students include timed tasks with long reading materials attached or 1-h video lectures; these harm some students' attention spans [11]. OL and their attention span impact the motivation of students. Because online classes are performed over the Internet, students cannot communicate in person with their classmates, and procrastination is inevitable due to the deadlines' flexibility [12]. Higher motivation levels may cause the student to persevere and seek out more challenging tasks even when challenges are performed, but lower motivation levels may cause the student to disengage from an activity [13]. Their research aims to determine how OL affects motivation and attention span. Earlier researchers suggested that several factors can affect a student's ability to pay attention in an online course. Eye tracking, reduced class feedback, and facial direction, when seen through a camera, can do it [14].

Researchers have been drawn to the study of OL attention, which has produced fruitful results. In real life, it might be challenging to gauge how attentive kids are in class. Accurately detecting each student's status of learning is a challenging task. Researchers have conducted studies on recognizing Learning Attention (LA) [15,16], producing potential outcomes [17]. examined the level of LA at home and abroad, split the emphasis of the research into two classes—Attention Recognition (AR) based on behavior and AR based on facial expression—and then discussed the AR's development phase. OL platforms and offline classroom settings are the main focus of research scenarios. Researchers have been working hard to understand AR in online classes for the convenience of voice and image acquisition. Using 636 students from management and economics disciplines at national universities as samples, a technique of developing a Triadic Theory of Learning based structural equation model that analyzed the effectiveness of OL was proposed by Ref. [18]. Deep Learning (DL) and Shallow Learning (SL) are the two types of teaching quality. This method contends that learning vigor, teaching methods, and teaching contents significantly impacted students' LA, which led to further optimization of the OTS design to increase the effectiveness of OL [19]. proposed a technique with three sub-modules. For head pose detection, the first module primarily identifies the angle of deviation of each student's head. By closing the mouth and eyes, the second module scores fatigue. The third module scores emotion by finding facial expressions. After combining the abovementioned data, a fuzzy comprehensive evaluation technique objectively measures their LA.

To check students' fatigue levels in real-time and with accuracy [20], developed a Fatigue Monitoring System (FMS) that assessed a student's fatigue state based on changes in the contour of their eyes. In order to realize the checking of quality of the class [21], developed a method to employ Facial Recognition (FR) technology for calculating the classroom learners' head-up rate by checking the rotation Angle of the head [22]. proposed a method to recognize learners' 3D head rotation angle, degree of eye closure, and facial expression to assess their facial concentration degree [23]. proposed an FR-based fuzzy comprehensive evaluation technique. The four detection parts—left (right) head-turning angle, head-lifting (low) head-turning angle, eye closure, mouth closure, and facial expression—were found by analyzing face images. The learning focus was processed using the quantitative scores of the four detection parts.

DL technology-based image feature classification has appeared as a renowned research direction while computer software and hardware development, Artificial Intelligence (AI) technologies, and digital image processing have continuously developed. As known earlier, DL research has set off in recent years, thus covering a wide range of topics, including text, pronunciation, and visual elements. DL has been successfully used in recognition and image classification and may efficiently circumvent the issues associated with Feature Extraction's (FE) artificial selection. Though the CNN-based student fatigue state detection has superior accuracy, it is challenging for real-time application on the terminal side. These are the two other key issues with the current research: (a) the eye's positioning is subject to the peripheral situation, and (b) despite the incredible accuracy of the CNN-based student fatigue state detection, it is challenging for real-time implementation of the terminal side.

Using a robust Face Detection (FD) algorithm and a sophisticated FE model, DLIP, which extracts eye and mouth landmark features, provides an edge-based student attentiveness analysis system in this study. The FE is fed into the Visual Geometry Group-19 (VGG-19) learning model to determine whether the student is focusing or fatigued on paying attention. The model's performance, as applied in Edge Devices (ED), is linked to server-based models. The findings that prove the viability of the proposed model are summarized.

2. Research contribution

(a) **Integration of FD and AR:** The paper proposes the integration of Face Detection (FD) and Attentiveness Recognition (AR) using the MTCNN algorithm for FD and DLIB for extracting Eye and Mouth Aspect Ratios (EAR and MAR). This combination enables the system to identify and analyze the attentiveness of students.

(b) **Deployment of VGG-19 Network:** This work proposes using the VGG-19 learning network for FE and analysis. By leveraging this DL model, the EBSAAS can effectively analyze the facial FE to determine student attentiveness.

(c) **Implementation of Face Recognition (FR) System: The** paper describes implementing an FR system within the EBSAAS framework. The FR system utilizes the ED (a smartphone) and the Edge Server (ES) for processing and recognition tasks, providing the ability to recognize individuals and assign names to their faces.

(d) **Training and Processing Pipeline:** The paper presents a training and processing pipeline where the EBSAAS model is initially trained on a face dataset within the ED. Once trained, the system can capture images of individuals or groups, perform FS, recognize faces, and analyze attentiveness by sending images to the ES for processing. This pipeline enables real-time analysis of student attentiveness.

(e) **Integration of Edge Server:** The paper highlights the inclusion of an ES in the EBSAAS architecture. The ES is responsible for processing the attentiveness analysis and training the recognition classifier. By offloading these tasks to the ES, the EBSAAS can leverage its computational capabilities and potentially enhance processing speed.

The article is systematized as follows: The overview and the research works are discoursed in Section 1, Section 2 has the proposed EBSAAS model, Section 3 has the investigational performance analysis with comparative illustration, and Section 4 is the paper's conclusion with recommendations for future work.

3. Proposed edge-based student attentiveness analysis system (EBSAAS)

FD and AR are the two significant aspects of EBSAAS. Using the Multi-task Cascaded Convolutional Networks (MTCNN) algorithm, a face's potential location is found inside an image during the FD phase. In order to find students who are LA, the attentiveness recognition phase uses DLIB for extracting the student's Eye and Mouth Aspect Ratios. The FE is then put into the VGG-19 learning network, which analyzes the student attentiveness. The EBSAAS is connected to the ED (a smartphone) and the ES to find out the FR system's processing time.

The learning model is first trained in the ED using a face dataset. The EBSAAS then uses its camera to snap an image of a person or group and starts the FD process. Once the Face is recognized, it is marked on the photo, and the person's name is assigned. In order to recognize, the photo is sent to the ES by the ED using the Internet rather than internally processing the attentiveness analysis. In the ES, the training of the recognition classifier takes place, which is very much like the ED. Following the FD and attentiveness recognition procedures, it passes the tagged and labelled photo back to the ED after obtaining the image from the same. The depiction of the ES-FR system through a block diagram is shown in Fig. 5. In Fig. 1, the framework outlined in this paper is depicted. The specifics of each part's implementation will be covered in detail.

3.1. Face Detection

In order to perform FD and face alignment rapidly and effectively for this work, this work employs MTCNN (Fig. 2), a DL-based FD and face alignment algorithm [24]. The left and right corners of the mouth, the nose, and the left and right eyes are the five facial features that MTCNN can recognize. The five significant points, however, are not enough to extract information about facial fatigue; hence this work employs MTCNN for FD. Proposal Network (*P*-Net), Refine Network (R-Net), and Output Network (ONet) are the three subnets that face MTCNN. They are arranged in cascades to form MTCNN [25].

In FD, the goal is to identify and locate the presence of faces within an image or video frame. The primary focus is detecting the overall face region rather than specific facial features. However, during the FD process, certain facial features can be used as cues to locate and validate the presence of a face. Some common facial features that can be applied are eye, nose, mouth and face contour. The facial features are retrieved using DLIB.

P-Net. This network's primary task is to decide the candidate window's regression box and BB. Highly overlapping windows are removed after calibrating the candidate window using non-maximum suppression. *P-Net* is a face region-specific proposal network. The network employs border regression and a locator of essential points in a face to generate a face area's primary proposal after using a face classifier to identify if a face is present. This section will produce many candidate windows, which R-Net will then use as input.

R-Net: This network's primary job is to weed out bogus samples while continuing to gather Bounding Boxes (BB) and Regression Vectors (RV). *R-Net* has a connecting layer that is more conclusive than the preceding network. This network's primary job is to weed out false samples while continuing to gather BB and RV. *R-Net* has a connecting layer that is further complete than the preceding network. Many candidate windows are obtained once the test sample passes on the *P-Net* layer. The network filters out a high number of candidate windows. The prediction outcomes are further improved on the selected candidate boxes using Non-Maximum Suppression (NMS) and BB regression.

O-Net. The third network is more complex than the first two. There are 256 Fully Connected (FC) layers in *O-Net*. Also, this network layer will calculate the facial feature points' position after further filtering the *R-Net* candidate window. Additionally, this procedure can eliminate various impediments like hats, sunglasses, and ordinary glasses.

3.2. Detection method based on eye and mouth feature

3.2.1. Face localization

The system localizes the FD after successfully detecting it. Face localization includes Facial Landmarks Detection (FLD). The purpose of FLL is to precisely identify and locate specific key points or landmarks on a human face. These landmarks represent important facial features such as the eyes, eyebrows, nose, mouth, and jawline. By accurately localizing these landmarks, performing several tasks related to facial analysis, FR, tracking, and FD becomes possible. FLD is the image shape predictors' subset. The following is a list of the facial landmarks found in

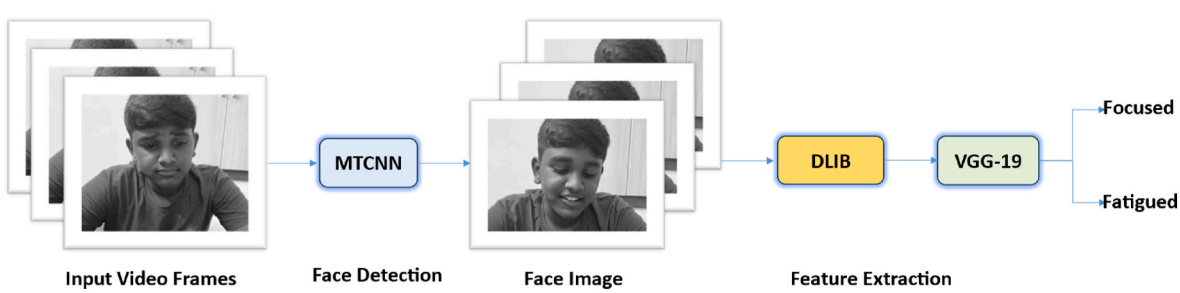


Fig. 1. Framework of fatigue detection.

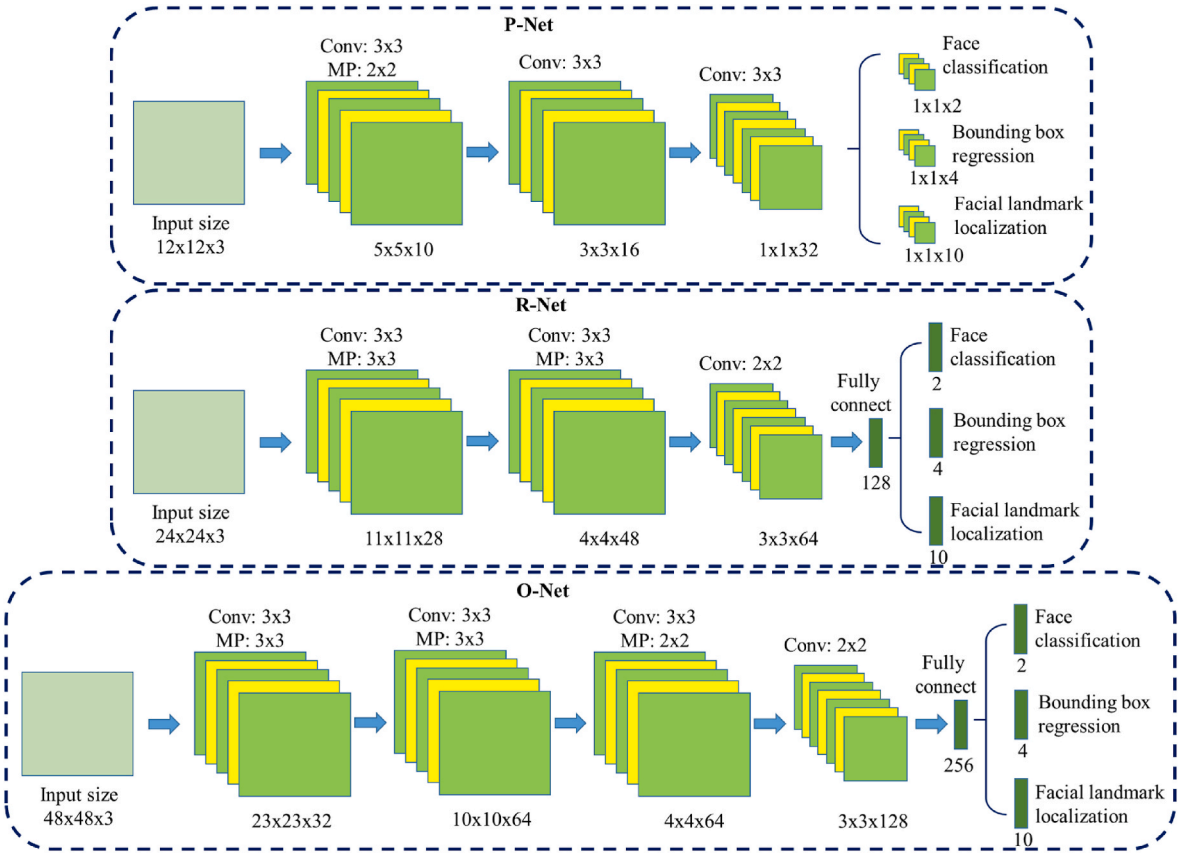


Fig. 2. MTCNN architecture.

the face region.

Eyes
Eyebrows
Lips
Nose
Jawline

Using the shape predictor method, one may localize these landmarks on the Face by quickly extracting the eye region. These landmarks are found around the eyes. The DLIB library used $C++$ programming to help the Machine Learning (ML) algorithm. Using the function named *"dlib. shape predictor"* ("shape predictor 68 landmarks. dat"), 68 unique feature points are returned by the algorithm in the given frames for each person [26]. This article uses the *"dlib.get_frontal_face_detector"* function to obtain the frames. A semi-automatic method for the landmark points was introduced by Ref. [27], and the outcomes of 300 different faces in "The Wild Challenge (300-W)" was reported in that article. It was the prime FLD test where earlier techniques were compared. The 68 individual landmark points that were detected are listed with precise numerals in Fig. 3. This study determined the lip distance and EAR using these landmarks.

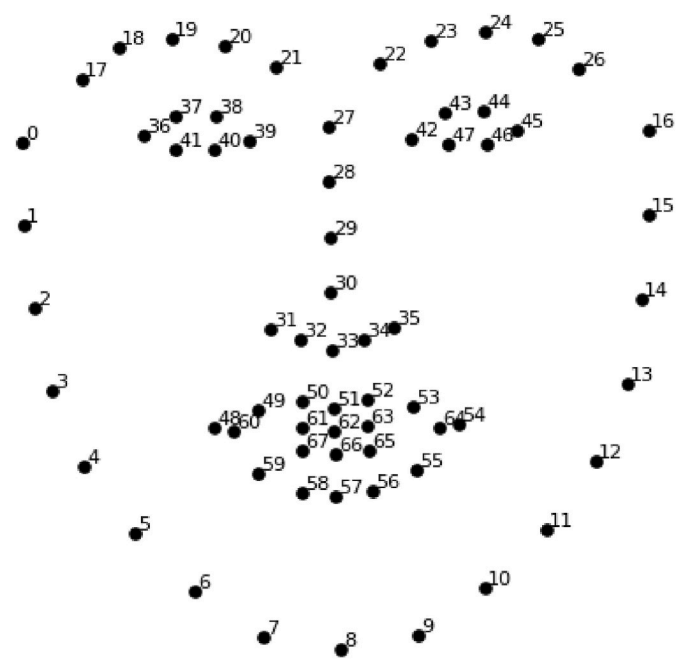


Fig. 3. The 68 FLD coordinates from the iBUG-300 W dataset.

3.2.2. Eye aspect ratio (EAR)

The EAR, a crucial part of this algorithm, can be used to assess whether or not someone is closing their eyes in the provided video frame.

Fig. 4 shows the eye that a set of six facial points with labels and specific coordinates. The distance between points p_1 and p_4 (the width of an eye) is along a horizontal line, whereas the distance along a vertical line is between the middle of p_2 and p_3 and p_6 and p_5 (the height of an eye). While the vertical line's length will vary when the eye opens and closes, the horizontal line's length will always remain constant. This work detects blinking by assessing the ratio between the lengths of these two lines. While the eye is open, this ratio will remain constant, but it will swiftly go to '0' when the eye blinks.

The aspect ratio will increase and remain constant over time, as seen

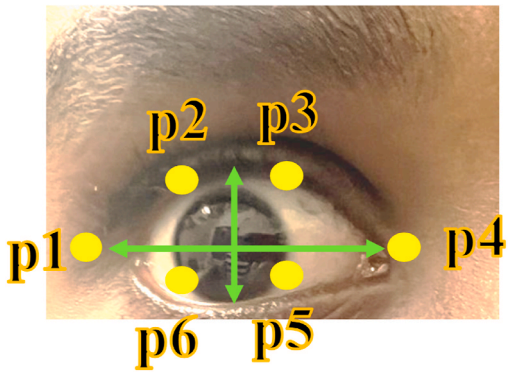


Fig. 4. Eye key points.

in Fig. 5(a). Fig. 5 (b), on the other hand, shows that the EAR will be equal to '0', showing that the person is blinking at that moment. Equ (1) below can be used to calculate the EAR.

For the active, fatigued, and sleep stages, the EAR values range from 0.38 to 0.30, 0.255 to 0.18, and 0.155 to 0.03. The experiment created a non-overlapping graph that helps calculate the classification boundary for the three states together. The threshold range (THrange) was measured using this graph. The range is,

$$TH_{\text{range}} = \begin{cases} EAR \geq 0.28\ ;\ \text{Active} \\ 0.17 < EAR \leq 0.27\ ;\ \text{Fatigue} \\ EAR \leq 0.17\ ;\ \text{Sleep} \end{cases} \tag{2}$$

In this study, the average eye blinks were prevented from interfering with fatigue detection based on EAR by choosing a time rate of 1000 ms, or 1 *Sec*. Drowsiness was detected by the proposed system when the EAR value, as EQU (2), goes beyond the threshold for a continuous time of 1000 ms.

3.2.3. Mouth aspect ratio (MAR)

Another symptom of fatigue, the yawn, is classified by lip distance. "Lip landmark points," or exciting points for calculating lip distance. The distance is calculated by deducting the upper lip weight's meaning from the lower lip weight's meaning. Fig. 6 depicts the distinctive top and lower lip points. Notably, each of the top and lower lips have eight distinct points. The weights of the points are added to determine the meaning of each lip.

The exciting point weights of the top lip are added together in EQU (3).

$$Top_{lipW} = \sum_{i=50}^{53} W_{P_i} + \sum_{i=61}^{64} W_{P_i} \tag{3}$$

To add up every lower lip's point weights, EQU (4) is used.

$$\text{Lower}_{lipW} = \sum_{i=56}^{59} W_{P_i} + \sum_{i=66}^{68} W_{P_i} \tag{4}$$

Then, as in the following, the mean of the top and lower lips is determined: The lip distance is determined as EQU (5) and EQU (6)

$$Top_{lip_mean} = \frac{\text{Top}_{lipW}}{8} \tag{5}$$

$$Lower_{lip_mean} = \frac{\text{Lower}_{lipW}}{8} \tag{6}$$

Top_{lip_mean} and $Lower_{lip_mean}$ are derived in EQU (7).

$$\text{Lip}_{distance} = \left\| Top_{lip_mean} - Lower_{lip_mean} \right\| \tag{7}$$

3.3. VGGNet architecture

Authors [28] from the University of Oxford devised the CNN model as VGGNet. This study's primary aim is to examine the impact of CNN depth on accuracy. A 224 × 224 RGB image serves as the input to the VGG-based convNet. The preprocessing layer subtracts RGB images with pixel values between 0 and 255 from the ImageNet training set's mean image values.

Weight and CL transform input images into training imagesVGG16 using 13 CL and 3 FC layers, with smaller (3 × 3) large filters instead of huge ones. With one 7 × 7 Convolutional Layer (CL), it now has the same effective receptive field. VGGNet is the version used in this study, which includes 19 wt, 16 CL, 3 FC, and 5 Pooling Layers (PL) (Fig. 7). The VGGNet uses two FC layers with 4096 channels each, followed by a further FC layer with 1000 channels and a SoftMax layer to classify data.

3.3.1. Details of architecture

- The CL holds the first two layers with 3 × 3 filters, and since 64 filters are used in the first two layers, the volume is 224 × 224 × 64 due to the deployment of the same convolutions. The filters always have a 3 × 3 stride of 1.
- The volume was reduced from 224 × 224 × 64 to 112 × 112 × 64 using a PL with a 2 × 2 size and stride 2.
- Two CLs have 128 filters, resulting in a new dimension of 112 × 112 × 128.
- The PL reduces the volume to 56 × 56 × 128.
- The size of the Down-sampling is 28 × 28 × 256, followed by two CLs with 256 filters each.
- A max-PL separates two stacks with three CLs.
- After the last PL, volume 7 × 7 × 512 is flattened into an FC layer with 4096 channels and 1000 SoftMax classes.

4. Experiment

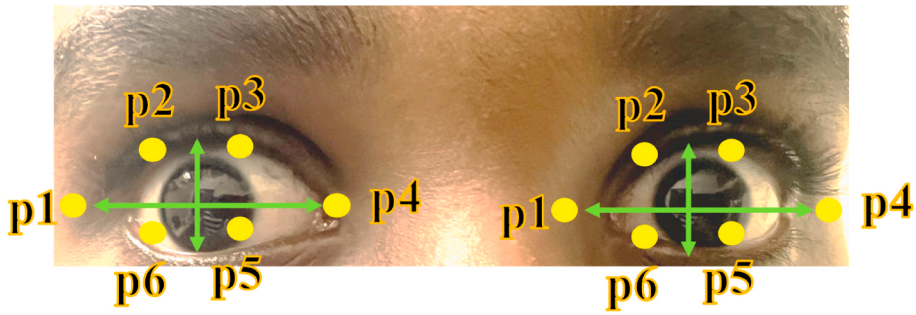
4.1. Real-life drowsiness dataset (RLDD)

RLDD includes 30 h of RGB videos of 60 people in good health. 180 videos—one for alertness, low vigilance, and drowsiness—were given to each tester. Staff and students who volunteered or received extra credit participated. Everyone who took part was older than 18.51 men and 9 women aged 20–59, with a mean age of 25 and standard deviation of 6. From 180 videos, frames were extracted for training and testing. The training uses 40% of the frames, and 60% is achieved using 40% of testing.

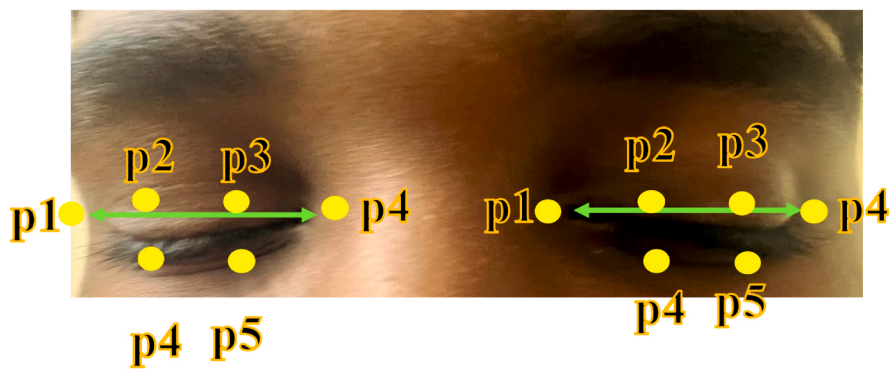
10 Caucasians, 30 Indo-Aryans and Dravidians, 8 Middle Easterners, and 7 East Asians. In 72 of the 180 videos, the subjects had much facial hair, and 21 of the 180 videos had the subjects' wearing glasses. Videos were shot at different angles against unusual circumstances from real life. The students input each video themselves using a webcam (or) smartphone. The video frame rate was consistently below 30 *fps*, typical of client cameras. Fig. 8 indicates RLDD images.

4.2. Experimental results and analysis

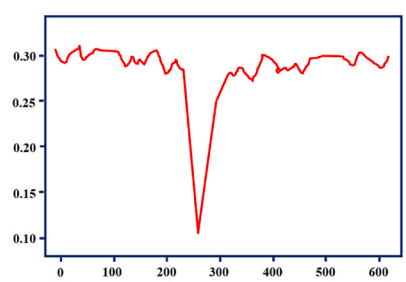
The Operating System (OS) environment used for this experiment is Linux (Ubuntu14.04). The computer's configuration details are Windows 10 OS, Intel (R) Xeon (R) CPU E5-2630 v3, 8 GB of RAM, and GTX1080 graphics card. Application based on smartphone technology for fatigue detection is created using Android Studio, and the ES program is created using Java and Eclipse IDE. Using MTCNN for detecting and cropping the images of faces, this work divided the video dataset into images for the experiment. The DLIB library labels the basic facial features after cropping the face image to determine the state value of the eye and mouth. This paper detects close fatigue by deciding the aspect ratio of the eyes and mouth. Ten videos from the RLDD test set were



(a) EAR for an Opened eye will be Grater and Constant



(b) EAR for a closed eye will be equal (or) Closer to '0' and Constant



(c)

Fig. 5. **(A)** Open eyes; **(b)** Closed eyes; **(c)** with automatically recognized landmarks with EAR drop graph.

$$E_A_R = \frac{\| X_2 - X_6 \| + \| X_3 - X_5 \|}{2 \| X_1 - X_4 \|} \tag{1}$$

chosen randomly for this study, and the findings are displayed in Table 1. In the 10 selected videos at random, the detection accuracy of fatigued driving is 95%. The results of sample detection are shown in Fig. 9.

Our algorithm proves that the device can properly detect the driver's fatigue state and function at high speeds while working in many conditions. The class labels assigned in the research work are fatigue and focused, where the outcome is depicted in Fig. 9. This proposed approach increases the fatigue driving detection algorithm's accuracy compared to FaceNet + KNN [29] and MTCNN + LSTM [30] algorithms. Additionally, it performs better in real-time, which satisfies the needs of the attentiveness analysis system [31,32]. In Table 1, the comparative result is displayed.

The detection of student reaction has to be quick to predict the nature of listening, and it is necessary to analyze the performance speed for the existing model with the proposed model. A comparatively proposed model identifies the outcome quickly, which is minimal than existing models.

4.3. Edge model analysis

Combining 10 random images into 1 image, this work adjusted the dataset to replicate a classroom environment to evaluate the ES and ED performance. It allowed us to analyze the performance associated with the proposed model. The processing time for FD by the ES and ED is shown in Fig. 10. As the number of faces rises in the ED, there is a drastic increase in the detection time. On the other hand, the ES experiences a modest increase in FD time for a similar instance.

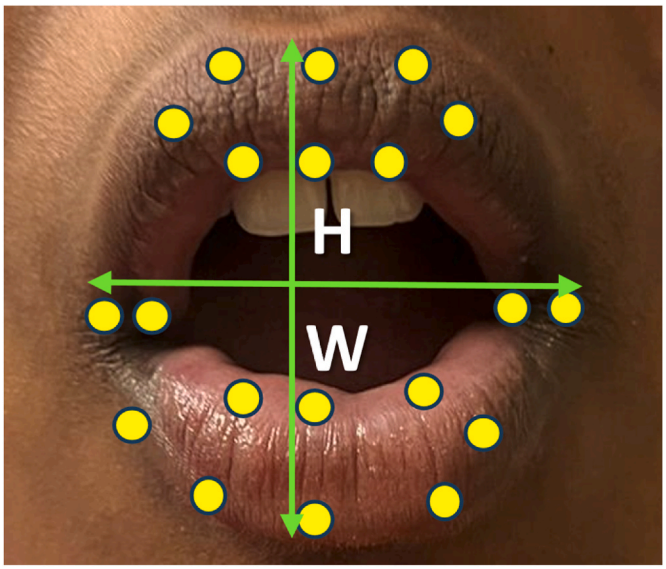


Fig. 6. Top and lower lips landmark points.

The image resolution influences the FD time rather than the image file size. For different image file sizes (image resolution 446 × 614 pixels), the FD time difference is displayed in Fig. 11 for the ES and ED.

The processing time for attentiveness recognition in the ED and ES is displayed in Fig. 12. Similar patterns are traced in the attentiveness recognition time measurement. The more faces in an image, the longer it takes for the ED to recognize them attentively. The ES, however, causes a moderate increase in recognition time. In the image, according to the number of faces, Fig. 12 compares the overall processing time between the ES and ED. Regarding ED, the FD and fatigue recognition times internal to the ED are added to determine the total processing time.

The RLDD dataset is used in the research, and the DLIB library features are considered. The performance measures, like accuracy, speed, and time for ES and ED, are analyzed. The processing time for attentiveness recognition in the ED and ES is illustrated in Figs. 10–11. The ED takes longer to FR carefully when more people are in an image. However, the ES brings a slight increase in recognition time. To forecast the nature of listening, it is vital to identify student reactions quickly, and it is essential to compare the performance speeds of the proposed and current models. In comparison to existing models, the proposed model predicts the result in a short amount of time. The accuracy of classification is compared, and EBOL outperforms the *state-of-the-art* techniques.

Table 1
Accuracy analysis.

Algorithms	Accuracy (%)	Speed (ms $\cdot f^{-1}$)
Model 1	93.18	53.61
Model 2	90.88	61.54
Proposed Model	95.29	47.23

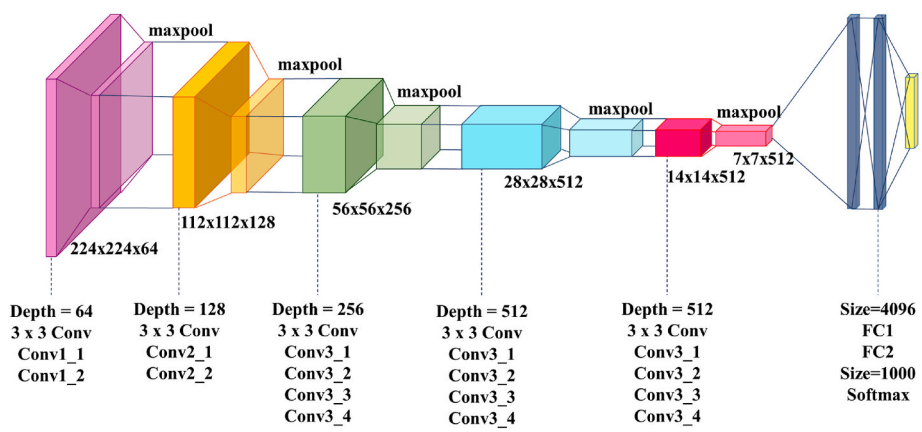


Fig. 7. Diagram of the network model of VGG-19.

Fig. 8. The sample of the RLDD dataset.

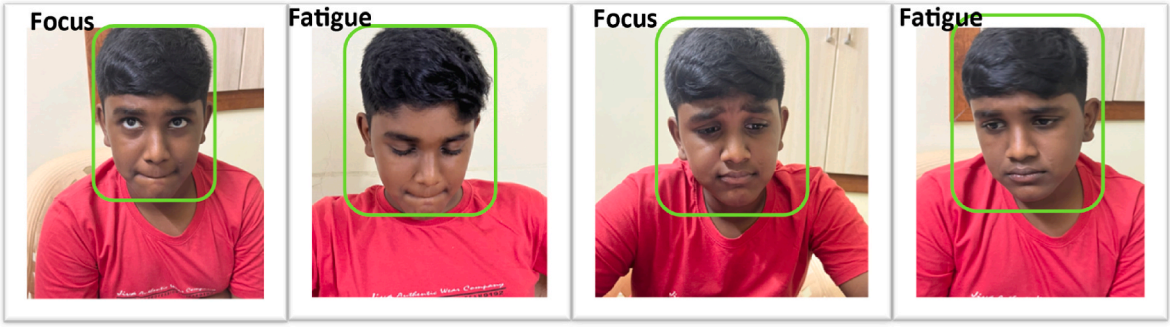


Fig. 9. Example of detection results.

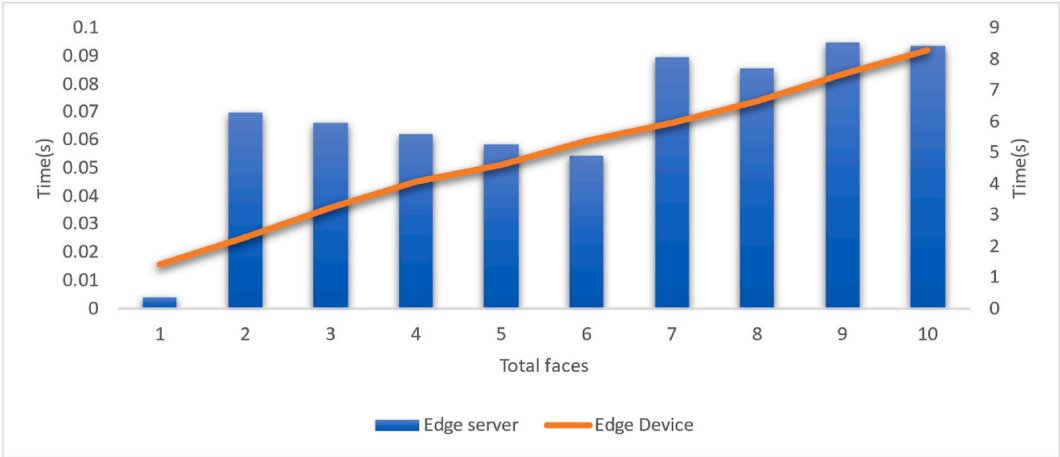


Fig. 10. The ES's and ED's-FD time.

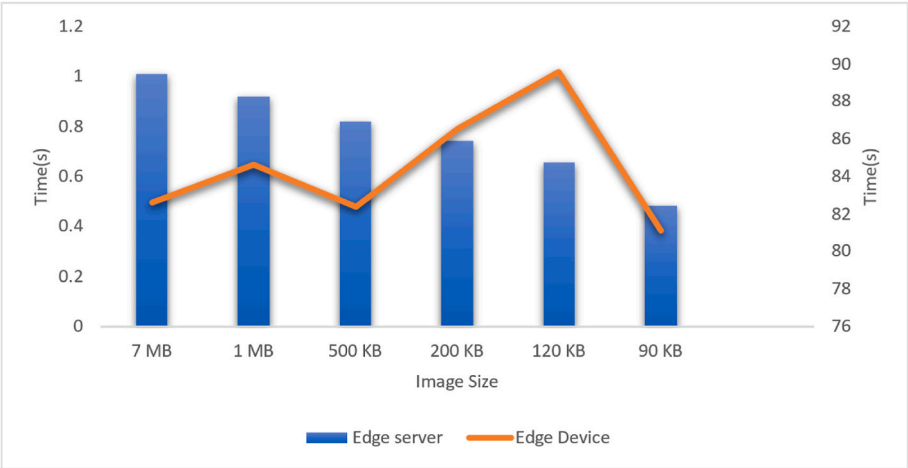


Fig. 11. Variation in FD time, including image file size.

5. Conclusion and future work

In conclusion, the development of Edge-based Online Learning (EOL) has given students more ease and flexibility in their academic endeavors. While taking online classes, measuring, and assessing students' involvement has been particularly challenging. This problem can be effectively solved by the proposed Edge-Based Student Attentiveness Analysis System (EBSAAS), which enables teachers to precisely measure and assess pupils' levels of attention using innovative Deep Learning (DP) models. The system uses innovative technology to analyze student attentiveness more quickly and effectively, enhancing learning outcomes and the efficiency of Online Learning (OL). The proposed model states whether the student is fatigued or focused, and the detection is achieved with a speed of 47.23, which is minimal than other models. The proposed model stands for a significant improvement in the evolution of contemporary education by giving teachers a potent tool to enhance their educational approaches and guarantee that students reach their full potential.

Future research can focus on developing models that can generalize well across different domains, including different ethnicities, age

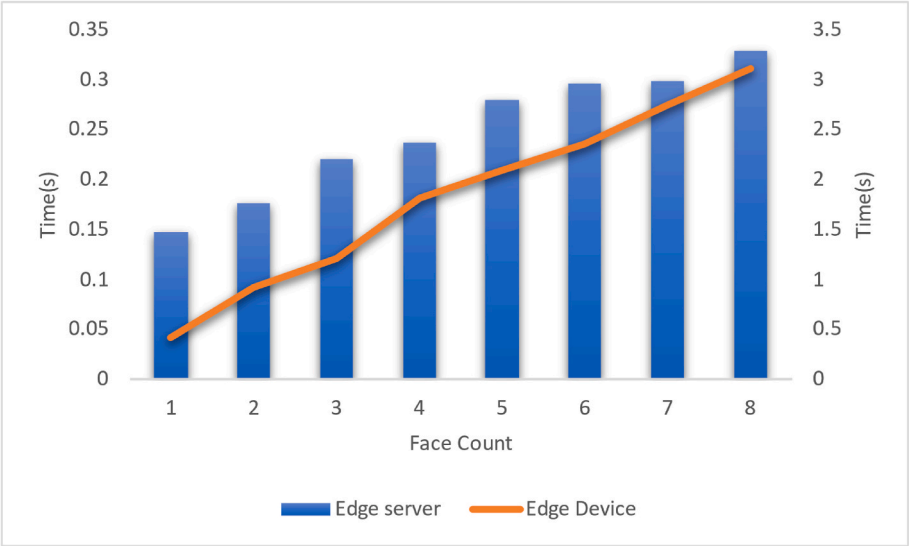


Fig. 12. Performance comparison of the complete processing time between the ES and ED.

groups, and cultural contexts. While current Face Detection (FD) algorithms have achieved high accuracy, there is always research direction for improvement. Research can focus on developing more accurate models that can handle challenging scenarios, such as occlusions, extreme poses, variations in lighting conditions, and diverse demographics.

Credit author statement

Rasheed Abdulkader: Software, Firas Tayseer Mohammad Ayasrah: Resources, Venkata Ramana Gupta Nallagattla: Supervision, Kamal Kant Hiran: Data Curation, Pankaj Dadheech: Visualization, Project administration, Vivekanandam Balasubramaniam: Validation, Sudhakar Sengan: Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Writing - Original Draft, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] Pallavi S, Smruti SR. Internet of Things: architectures, protocols, and applications. J Electrical Comput Eng 2017;2017:1–25. https://doi.org/10.1155/2017/9324035.
[2] Kumar S, Tiwari P, Zymbler M. Internet of Things is a revolutionary approach for future technology enhancement: a review. J Big Data 2019;6:111.
[3] Kim J, Maher ML. Conceptual metaphors for designing smart environments: device, robot, and friend. Front Psychol 2020;11:198.
[4] Greengard S. The internet of things. MIT Press; 2021.
[5] Oke A, Fernandes FAP. Innovations in teaching and learning: exploring the perceptions of the education sector on the 4th industrial revolution (4IR). J Open Innovat: Technol Market Complex 2020;6(2):31.
[6] Bajracharya B, Blackford C, Chelladurai J. J. Prospects of Internet of Things in education system. Prospects 2018;6(1):1–7.
[7] Mah PM, Skalna I, Munyeshuri E, Akoko J. Influence of Internet of Things on human psychology (internet of Thoughts) for education, healthcare, and businesses. EAI Endorsed Trans Mobile Commun Appl 2022;7(2).
[8] Fanian F, Khatibi V, Shokouhifar M. A new task scheduling algorithm using firefly and simulated annealing algorithms in cloud computing. Int J Adv Comput Sci Appl 2018;9(2):228–31.
[9] Dias de M, Assunção M, da Silva A, Veith A, Buyya R. Distributed data stream processing and edge computing: a survey on resource elasticity and future directions. J Netw Comput Appl 2018;103(8):1–17.
[10] Simamora RM. The Challenges of online learning during the COVID-19 Pandemic: an Essay Analysis of performing arts education students. Stud Learn Teach 2020;1(2):86–103.
[11] May SG. The excess of content and the decreasing attention span in students Get our Weekly. 2020. p. 1–5.
[12] Balan AK, Jacintos AR, Montemayor T. The influence of online learning towards the attention span and motivation of college students. Mapua University; 2020 (Unpublished undergraduate research).
[13] Hartnett M, Hartnett M. The importance of motivation in online learning. In: Motivation in online education. Massey Univeristy; 2016. p. 5–32. https://doi.org/10.1007/978- 981-10-0700-2_2.
[14] Deng Q, Wu Z, Students' Z. Attention assessment in eLearning based on machine learning. Earth Environ Sci 2018;26(4):741–52.
[15] Huang L, Su J, Pao T. A context-aware smart classroom architecture for smart campuses. Appl Sci 2019;9:1837.
[16] Francisti J, Balogh Z, Reichel J, Magdin M, Koprda S, Molnár G. Application experiences using IoT devices in education. Appl Sci 2020;10:7286.
[17] Sun S, Liu R. A review of research on attentiveness recognition. Sci Tech Inf 2021;4:6–8.
[18] Wang P, Wang D. Research on the application of focus recognition technology in teaching monitoring. Comput Knowl Technol 2020;16:38–42.
[19] Chang Z. Analysis and reflection on concentration of online teaching in vocational colleges. J. Wuhan Eng. Inst. 2020;32:91–4.
[20] Yang SZW, Hu. Design of fatigue monitoring system based on face recognition technology. Instrum Technol 2020;8.
[21] He Z. Research on application of face recognition technology in quality monitoring of college students' lectures. China CIO News 2018;3:102.
[22] Chen H. Research on application of face recognition in learning attentiveness monitoring. J Ind 2020;10(9):70–1.
[23] Zhong ZL. He. Study on online education focus degree based on face detection and fuzzy comprehensive evaluation. Computer Science 2020;47(S2):196–203.
[24] Zhang K, Zhang Z, Li Z, Qiao Y. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process Lett 2016;23(10):1499–503.
[25] Ji Y, Wang S, Zhao Y, Wei J, Lu Y. Fatigue state detection based on multi-index fusion and state recognition network. IEEE Access 2019;7:64136–47.
[26] Bezerra G, Gomes R. Recognition of occluded and lateral faces using MTCNN. dlib and homographies; 2018. p. 1–4.
[27] Sagonas C, Tzimiropoulos G, Zafeiriou S, Pantic M. 300 faces in-the-wild challenge: the first facial landmark localization challenge. In: International conference on computer vision workshops; 2013. p. 397–403.
[28] Karen S, Andrew Z. Very Deep convolutional networks for large-scale image recognition. International Conference on Learning Representations; 2015.
[29] Adhinata FD, Rakhmadani DP, Wijayanto D. Fatigue detection on face image using FaceNet algorithm and K-nearest neighbor classifier. J Inf Syst Eng Bus Intell 2021;7(1):22–30.
[30] Chen L, Xin G, Liu Y, Huang J. Driver fatigue detection based on facial key points and LSTM. Security and Communication Networks; 2021. p. 1–9.
[31] Abhishek G, Alagan A, Ling G, Ahmed SK. Deep learning for object detection and scene perception in self-driving cars: survey, challenges, and open issues. Array 2021;10:100057. https://doi.org/10.1016/j.array.2021.100057.
[32] Charalambos T, Vladan V, Vladimir D, Fredi N. Visual SLAM algorithms and their application for AR, mapping, localization and wayfinding. Array 2022;15:100222. https://doi.org/10.1016/j.array.2022.100222.

Computers and Education: Artificial Intelligence 5 (2023) 100191

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

Attention monitoring of students during online classes using XGBoost classifier

Muhammad Kamal Hossen [a,b,*], Mohammad Shorif Uddin [b]

[a] Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chattogram 4349, Bangladesh
[b] Department of Computer Science and Engineering, Jahangirnagar University, Dhaka 1342, Bangladesh

ARTICLE INFO

Keywords:
Online class
Student engagement
Attention monitoring
Artificial intelligence
Machine learning
XGBoost

ABSTRACT

With the ever-growing influence of globalization and digitization on education, online learning has gained immense popularity, driven in large part by the global COVID-19 pandemic. In response, educational institutions have rapidly expanded their virtual course offerings and assessment methods. One significant challenge in the realm of online education is the effective assessment of student involvement, engagement, and attentiveness during virtual classes. To address this challenge, we introduce a specialized system tailored to identifying and comprehending the complex range of student behaviors exhibited during online classes. Our system takes into account the diverse contexts and individual considerations that shape students' responses to the online learning environment. It employs user authentication through facial recognition technology and seamlessly integrates vital components, including facial detection, hand tracking, mobile phone detection, and pose estimation modules. Leveraging these components, we extract high-level features to construct a comprehensive dataset for training machine-learning models designed to detect students' attention levels during online classes. After evaluating six prominent machine learning algorithms, we have selected the extreme gradient boosting (XGBoost) algorithm, which demonstrated an impressive 99.75% accuracy on the test data. Finally, the system generates a comprehensive anonymous report on the level of attentiveness of students, accessible through a dedicated webpage that includes a summary of the behavior of students throughout online classes, providing valuable insights for personalized interventions and enhanced online learning experiences.

1. Introduction

The virtualization of education has turned conventional classrooms into virtual platforms, presenting both potential and difficulties. As a result of the COVID-19 epidemic, which has forced the majority of schools, colleges, and universities to move to online learning, extend their online course offerings, and hold online exams, the transition to remote education has become more prevalent (Bashir et al., 2021; Manca & Meluzzi, 2020). In virtual classrooms, the lack of being present in person and face-to-face interactions presents special challenges, notably in terms of student involvement and attention in real-time (Hu, 2022). A viable attention monitoring system that can give insightful information about the levels of involvement of students during online classes is urgently needed to close this gap and guarantee the best learning experiences (Ong & Quek, 2023; Dhawan, 2020).

Most schools, colleges and universities offer online classes and monitor the attentiveness of the students using Zoom, Google Meet, Microsoft Teams, Cisco Webex, Adobe Connect, and similar tools (Kansal et al., 2021). To maximize learning outcomes, it is indispensable to keep the attention of the students in online classrooms (Bedi, 2023; Khan et al., 2017; Moşteanu, 2021). Through conversations, body language, and facial expressions in typical settings in the classrooms, teachers may quickly assess how attentive their students are. These indications are frequently minimal or not present in online spaces, though. The problem is exacerbated by the scarcity of high-speed Internet and high-quality devices. Due to reentering the virtual classroom, students miss crucial lecture sections. Additionally, students find it difficult to focus during online lessons because the pandemic impairs their focus at home (Muthuprasad et al., 2021). For this reason, it is essential to have a robust system in place so that educators can see students who could be unmotivated, inattentive, or having trouble and offer prompt interventions and assistance (Villegas-Ch et al., 2023).

Numerous attention-monitoring methods have been created to assess the involvement, engagement, and attentiveness of the students during

* Corresponding author. Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chattogram 4349, Bangladesh.
E-mail addresses: mkhossen@cuet.ac.bd (M.K. Hossen), shorifuddin@juniv.edu (M.S. Uddin).

https://doi.org/10.1016/j.caeai.2023.100191
Received 20 August 2023; Received in revised form 4 November 2023; Accepted 25 November 2023
Available online 1 December 2023
2666-920X/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

online classes in order to solve these issues (Alhothali et al., 2022; Parvathy & Smitha, 2022). Modern tools like video conferencing, computer vision, artificial intelligence (AI), machine learning, and data analytics are used by these systems. Most of them utilized two or more factors among facial features and expressions, head movements, eye gaze, speech and voice, and data from student interactions and activities within the online learning platform to assess the level of attention of students (Alhothali et al., 2022; Parvathy & Smitha, 2022).

While current attention monitoring systems show promise, they also exhibit significant limitations, warranting further research and improvement. Primarily, these systems rely on facial expressions (Anh et al., 2019), eye movements (Guo & Xu, 2022), and head orientation (Reddy et al., 2022) to assess student attention, raising concerns about precision. Several factors contribute to the need for enhancing precision. Technical challenges, including internet connectivity, hardware limitations, and software compatibility, along with environmental distractions like noisy environments, physical comfort, and household responsibilities, as well as the diverse learning preferences of students such as learning style, time of day, and engagement with materials, can introduce variability in attention estimation. Moreover, the constraints of eye-tracking technology (Sharma, Joshi et al., 2023), while valuable, can limit system accuracy. Therefore, addressing these limitations and exploring advanced attention assessment techniques is imperative. These endeavors are crucial for refining the reliability of attention monitoring in online learning environments, ultimately benefiting both educators and students.

In addition, existing attention-monitoring systems may have a limited focus, potentially overlooking the entirety of student involvement. This encompasses a comprehensive understanding of how students engage in online learning, extending beyond tracking visual cues (Kainat et al., 2022) like eye movements (Sharma, Amin et al., 2023), cursor activity, and screen time. It encompasses a broader range of factors, including audio analysis and active participation, which influence students' cognitive and affective states during online classes. These overlooked dimensions are essential for achieving a comprehensive understanding of student attention (Novita et al., 2022). The significance of capturing the entirety of student involvement stems from the multifaceted nature of attention, which extends beyond what visual cues alone can reveal. Audio analysis can unveil nuances in speech patterns, ambient sounds, and verbal responses, providing deeper insights into a student's engagement (Elias, 2020). Furthermore, active participation, including responding to questions and engaging in discussions, reflects cognitive involvement and active learning (Hodges et al., 2020). Integrating these dimensions into attention monitoring offers a more accurate and comprehensive assessment, ultimately enhancing the effectiveness of online teaching and learning.

Lastly, the continuous surveillance of students' facial and visual data raises legitimate privacy concerns. Protecting student privacy and ensuring compliance with stringent data protection laws are paramount. This involves implementing robust measures for secure data transfer, storage, and usage (May and George, 2011). By addressing these privacy concerns, educational institutions can strike a balance between leveraging attention-monitoring technologies for improved learning experiences and safeguarding sensitive information of students (Martinovic & Ralevich, 2007). This commitment to data security not only upholds legal and ethical standards but also fosters trust among students, educators, and stakeholders in online education.

Taking all these things into consideration, our objective is to surpass the current limitations by creating a reliable and efficient online class attention monitoring system. In an effort to promote trust in online education, the objective is to emulate conventional face-to-face classes. This entails ensuring seamless and dependable execution while deterring dubious activities. Our proposed system identifies any disruptive conduct on the part of the students (Das et al., 2021; Lugrin et al., 2016; Reddy et al., 2022) and gauges their level of engagement during virtual lessons. Additionally, we have designed a simple and intuitive web interface to facilitate monitoring and generate comprehensive end-of-class reports anonymously that accurately capture the conduct and attentiveness of students during the course of the online classes. Furthermore, video processing takes place directly on individual computers of students. This approach offers advantages in terms of privacy and security, as it prevents the exposure of sensitive personal data and optimizes Internet bandwidth usage (Abate et al., 2021).

The proposed system utilizes various techniques to detect a diverse range of behaviors exhibited by students during online classes. These techniques include continuous face detection through the use of Deepface (Serengil & Ozpinar, 2020, October), hand tracking facilitated by MediaPipe (Lugaresi et al., 2019), and pose estimation using MediaPipe and solvePNP. Additionally, the system incorporates mobile phone detection via YOLOv7 (Wang et al., 2023). Each of the fundamental modules generates high-level features that are subsequently amalgamated to produce a dataset (Hossen & Uddin, 2023). These features are then utilized to train an XGBoost machine learning classifier (Chen & Guestrin, 2016, August), which is proficient in determining whether students are attentive during virtual lectures. All of the modules are seamlessly integrated into a user-friendly web interface designed for both teachers and students, created using Django (Django, 2023). Finally, an anonymous report detailing disruptive behaviors and attentiveness of the students during online classes is generated and displayed on a webpage in the section of teachers.

The major contributions of this paper are as follows:

- An attention monitoring system of the students containing a comprehensive anonymous report on the level of attentiveness is developed through machine learning techniques for conducting effective online classes.
- The most common disruptive behaviors of the students in online classes are identified by incorporating user authentication through face recognition and integrating basic components such as face detection, hand tracking, cell phone detection, and pose estimation modules.
- A dataset is also developed for detailed experimentations of the system.

The following sections of this manuscript are organized as follows. In Section 2, an overview of existing literature is provided, with a focus on the analysis of the characteristics previously mentioned to evaluate the attention levels of the students. Section 3 presents a comprehensive description of the proposed system, encompassing the machine learning techniques employed and the dataset developed for detecting the attentiveness of the students. We provide the analysis of the dataset, evaluation of the performance of the machine learning classifiers on the dataset and the findings of our proposed attention monitoring system, in terms of the percentage of attentiveness, in Section 4, along with limitations and potential directions for future research. Finally, Section 5 provides a succinct conclusion.

2. Literature review

2.1. Learner attention

Learning theories offer invaluable insights into the cognitive, behavioral, and social aspects of learning, serving as guiding frameworks for educators and researchers to comprehend the learning process and enhance teaching techniques. These theories shape teaching strategies, instructional materials, and educational practices, optimizing learning experiences (Stewart, 2021).

In this context, learner attention is a pivotal cognitive and emotional state during learning. It entails focused engagement, concentration on instructional content, active participation, and sustained mental effort (Durães et al., 2018; Hassed, 2016; Mason, 2010). Research demonstrates that attention is a linchpin in the learning process, impacting

information selection, processing, and encoding. Learners who maintain their attention comprehend and retain information more effectively (Pimenta et al., 2015). Consequently, learner attention stands as a critical determinant of the quality of teaching and learning outcomes in educational contexts (Hwang & Yang, 2009).

Numerous factors exert influence over learner attention. Notably, the relevance of educational content, which reflects its alignment with learners' goals and interests, holds a substantial sway on attention (Hidi & Renninger, 2006). Effective teaching methodologies, including active learning, multimedia presentations and real-world applications, assume pivotal roles in capturing and maintaining learner attention (Prince, 2004). The dynamics of intrinsic motivation, stemming from genuine interest, and extrinsic motivation, linked to external rewards, both significantly contribute to sustaining attention (Deci et al., 1991). Furthermore, individual differences, encompassing cognitive abilities and learning styles, shape how learners allocate their attentional resources (Riding and Sadler-Smith, 1997).

2.2. Online learning and learner attention therein

Online learning, also known as e-learning or distance education, is an educational approach that leverages digital technology and the internet to deliver instruction and facilitate learning outside of traditional physical classrooms. It provides learners with flexibility in terms of when and where they engage with educational content, offering diverse resources, interactive elements, and self-paced learning opportunities (Muthuprasad et al., 2021). Online learning has become increasingly prevalent, offering accessible and innovative educational experiences for a wide range of learners, from working professionals seeking skill upgrades to students pursuing formal degrees (Means et al., 2013). It encompasses a variety of formats, from fully online courses to blended learning environments that combine online and in-person instruction. This educational paradigm has reshaped the way knowledge is accessed and acquired, making learning more accessible and adaptable to individual needs (Hodges et al., 2020).

Though online learning provides learners with flexibility but also introduces distractions and new considerations. Learners engage with course materials, peers, and instructors through digital platforms (Anderson & Dron, 2011). The digital realm offers numerous distractions, such as messages from chats, social media, email notifications and other applications like music and news sites, which can be more attractive to students. Moreover, these applications can constantly run in the background which can divert learner attention (Pimenta, Carneiro, Neves, & Novais, 2016). Prolonged online learning sessions may lead to cognitive fatigue, affecting sustained attention. In anonymous classroom observations, around 80% of 350 college students reported increased difficulty in focusing and staying engaged during online classes. Technical issues, including slow internet connections and platform glitches, can disrupt learner attention (Peper et al., 2021).

Educators and instructional designers must address these challenges and implement strategies to optimize learner attention. These strategies may include interactive and multimedia content, timely feedback, and cultivating self-regulation skills among learners (Narciss, 2008; Zimmerman, 2002). However, innovative pedagogical strategies, multimedia resources, and interactive discussions can captivate and maintain learner attention in the digital environment (Prince, 2004). As the field of online education continues to evolve, ongoing research into learner attention will provide valuable insights into the intricacies of digital learning environments.

2.3. Previous assessments of learner attention

Assessing learner attention has been a subject of interest and research in education, psychology, and cognitive science. Understanding how learners engage with educational content is crucial for improving teaching methods and optimizing learning outcomes. Various assessment methods have been employed by the researchers to understand how learners engage with educational content (Kainat et al., 2022).

The authors (Anh et al., 2019) presented a system that automatically records and analyzes student behavior in the classroom using computer vision techniques, outperforming current methods. It has seven parts and uses face recognition, facial landmark recognition, face embedding, face classification, and gaze estimation to gather behavioral data. Promising outcomes in facial recognition, behavior detection, and gaze estimation are shown by the experiments. The YOLOv3 algorithm and facial recognition are used in this paper (Mindoro et al., 2020) to introduce a novel method for forecasting student behavior in the classroom. The model demonstrated effectiveness in gauging student interest based on visible behaviors by achieving an accuracy of 88.606%. For accurate predictions, the study emphasized the significance of dataset development and evaluation.

In this project, computer vision was used to track the head position, drowsiness, and yawning of students taking online classes. The system showed a high degree of accuracy and can be added as a feature for predictive analysis on platforms for live streaming. The need for resources and student participation, however, presented difficulties for ongoing observation. The analysis of emotions may be taken into account for future improvements (Reddy et al., 2022). This study suggested a software component for tracking student attention during online lectures using machine learning and computer vision (Abate et al., 2021). The module offered insightful information for enhancing lecture delivery, and experimental results showed favorable feedback. The effects of tools on learning outcomes and its ability to catch exam cheating require more investigation, which is currently underway.

Parambil et al. (2022) presented a real-time vision-based intelligent classroom system that made use of AI to track emotions and attention of the students. The system gave instructors real-time graphical feedback, allowing them to more accurately assess attention of the students and support efficient session management, potentially enhancing academic performance. It was capable of assigning numerical scores to the students based on their concentration levels in class. The article (Kavitha et al., 2023) suggested a methodology to evaluate the effectiveness of online learners by assessing attention using EEG data and emotional state using recorded lecture videos. Test results and attention levels were predicted by machine learning models, demonstrating a positive correlation. This study investigated the emotional reactions of students and made recommendations for future research on course effectiveness and recommendation engines.

This article presented Local-SVM, a machine learning-based technique for identifying student participation in online learning (Aljabarti & Othman, 2023). Utilizing local features taken from the MoveNet model, it performed better than already-in-use methods like KNN. By addressing the shortcomings of existing computer vision systems, the suggested system produced cutting-edge results. The study (Ouyang et al., 2023) combined learning analytics feedback with AI performance prediction models to improve group learning. As a result, students appeared to be more engaged, performing better, and feeling more satisfied. In addition to discussing the revolutionary potential of AI-driven learning analytics, the study suggested three AIEd paradigms. For the continued advancement and application of AI in education, future work should broaden its focus and sample size.

Sharma, Joshi et al. (2023) described a system that detected and tracked the levels of participation of the students in online learning using a webcam. The system had modules for engagement classification, distraction detection, and facial emotion recognition. It could potentially boost student performance and accurately categorize various levels of engagement. In this paper (Mershad & Said, 2022), DIAMOND, a voice diarization system for assessing student participation in online lectures, is introduced. The number of failing students decreased by 43%, according to the results, and the tool successfully raised student engagement and improved performance. Successful participation

strategies were used by instructors.

A new system was proposed (Arciniegas et al., 2021) that used AI and computer vision algorithms to track student attention during online learning. It could help improve the dynamics of the classroom by precisely identifying when students are paying attention. Different recurrent neural network architectures were used to test the efficacy of the system; LSTM performed the best. It is advised that more research be conducted to examine different student attitudes and applications for in-person instruction. Gupta et al. (2023) introduced a real-time engagement detection method for online learners based on face emotion recognition. The system outperformed previous works and predicted student engagement levels with high accuracy. By giving instructors real-time feedback, it sought to improve the online learning experience.

In this paper (Sharma, 2021), a smart online class monitoring system was introduced and its goal was to address non-responsive behavior in online classes. In order to increase attention, the system employed techniques like switching up the student list and promoting participation. Undergraduate test subjects for prototypes showed reliable information and a focus on preventing platform abuse. Saha et al. (2021) examined the use of technology in education during the COVID-19 pandemic, focusing on issues with online class attendance and attention spans. It suggested utilizing facial recognition, AI, AR, VR, and MR technologies to raise student engagement. The study emphasized how crucial it was to address these issues if virtual learning is to be successful.

The article in (Parvathy & Smitha, 2022) deliberated upon diverse automated systems for monitoring and tracing the levels of attention of students in the classroom. These systems were equipped with EEG data, facial recognition algorithms, and deep learning methodologies. In order to compute the levels of attention, a range of techniques were employed, comprising SVM, DWT, HOG descriptor, Fisherface Algorithm, and FAN. The paper (Sartika et al., 2022) examined the difficulties of distance learning during the pandemic and proposed an advanced online monitoring platform. The platform used AI to track student attendance and expressions in real-time. The system was created using design research methodology and system engineering, resulting in a comprehensive service system with desktop and website applications for students and instructors.

Al Yousufi et al. (2023) conducted a study on big data analytics (BDA) in education. The study highlighted advantages of BDA in monitoring student progress, personalized learning, and learning analytics. BDA had goals ranging from dropout risk detection to individualized learning strategies. However, integration faced security and privacy challenges. Despite these challenges, BDA can transform academic results and enhance productivity. The article (Rahul et al., 2020) outlined a framework that leveraged computer vision and artificial intelligence in monitoring attentiveness of the students in online classes. By utilizing various non-verbal cues, the model computed an attention score and provided timely feedback to both academic institutions and learners. Its implementation holds the potential to elevate the quality of online pedagogy whilst concurrently assessing student progress.

An easy-to-use convolutional neural network (CNN) model is presented in this study by Guo and Xu (2022) for identifying the levels of concentration and learning fatigue of students during online classes. When analyzing video footage, it used face recognition and machine learning to determine the states and positions of the eyes. The model successfully validated the usage of the PERCLOS value for tiredness detection and obtained good recognition rates. In order to improve the online learning experiences of middle school pupils, the project would track how well they were learning. The suggested techniques helped to increase the efficiency of digital teaching management and online learning models. In order to improve the instruction quality of online courses at vocational colleges, this research paper (Wang, 2022) emphasized the significance of putting in place a digital monitoring system. In order to overcome the shortcomings of conventional monitoring systems, it suggested a three-level digital system based on big data analysis. A case study on advertising design courses, which showed improved teaching quality, illustrated the viability of the system.

In order to track the attention of the students during computer-based instruction, the study of Wang and Xia (2021, April) suggested a Raspberry Pi-based system. It estimated head angles and determined the focus of the pupils inside the viewing area using edge computing and head posture assessment. The technology displayed attention levels visually while wirelessly transmitting data to a central server. Results from experiments confirmed its efficacy, highlighting the significance of attention in learning. With the possibility for improvements like lowering errors and including real-time video monitoring, the system provided a practical option for real-time attention monitoring. This paper (Useche & El-Sheikh, 2016) described the development of an intelligent system to monitor student attention during online video-based education. The system, called PyTtention, used computer vision with a 3D camera to estimate the visual focus of attention of students based on head pose values as they watched videos. It also incorporated Bayesian analysis to calculate attention estimates. The system allowed educators to evaluate the effectiveness of instructional content and adapt courses accordingly. Testing results varied, with better accuracy observed when users had more movement during calibration and their head positions matched their calibration data. Future work included enhancing the robustness of the system and incorporating eye gaze tracking.

Due to numerous distractions and the lack of student-teacher connections in the current educational environment, student disengagement is a serious problem. Recent research investigates real-time analysis of the facial expressions of the students to identify emotional states such as boredom, bewilderment, attention, frustration, yawning, and tiredness. Traditional techniques of engagement monitoring have drawbacks. This strategy outperformed student self-reported engagement levels and yielded encouraging results (Pabba & Kumar, 2022). In order to enhance online higher education courses during the COVID-19 epidemic, SCUT adopted IT solutions (Lu et al., 2020). Through platform updates, dedicated servers, and collaborations with social media sites, issues like system readiness and resource scarcity were resolved. To improve the quality of online education, distinct online teaching approaches and teacher IT tool proficiency were highlighted. In the face of the epidemic, the study presents suggestions for enhancing online education.

The research studies referenced in the aforementioned texts concentrate on implementing computer vision, machine learning, and artificial intelligence methodologies to monitor and evaluate student behavior, attentiveness, and involvement within educational contexts. Although these studies display promise, they encounter various limitations. Certain studies solely depend on facial expressions to detect learner engagement (Anh et al., 2019), disregarding elements such as body language and interactions, which may not always accurately reflect the level of engagement of a student. Other research efforts highlight challenges such as the necessity for substantial resources, student cooperation, and ethical considerations associated with privacy and consent (Reddy et al., 2022). Multiple studies highlight the significance of constructing and evaluating datasets to ensure precise predictions, given that certain research relies on self-reported student data that could be susceptible to biases and inaccuracies (Mindoro et al., 2020). The systems that track emotions, attention, and engagement are faced with issues of precision, false positives or negatives, and generalization across several environments (Gupta et al., 2023). Furthermore, the effectiveness of AI-driven learning analytics necessitates a broader focus and larger sample sizes (Ouyang et al., 2023). Various systems demonstrate potential biases and real-time monitoring can raise ethical concerns as well (May et al., 2011).

Motivated by the limitations outlined above, we have designed an attention monitoring system that incorporates behavioral analysis and the surrounding environment by observing factors like face detection, pose estimation, hand tracking, and mobile phone detection. In order to

overcome technological constraints such as privacy concerns, we have obtained consent from the students and have opted for a resource-efficient implementation that is feasible for different educational settings, utilizing solely a webcam. The system has been trained and tested using a dataset comprising 4000 records of students (Hossen & Uddin, 2023) who either pay attention or do not pay attention to online lectures. Biases and inaccuracies were addressed in the creation of this dataset. Remarkably, we have employed the XGBoost machine learning classifier, which exhibits the highest accuracy and precision of 99.75% and 99.71% respectively, in the case of attention detection. Additionally, it offers real-time monitoring of student engagement through a user-friendly web interface, providing insights into both attentive and inattentive behaviors. Moreover, this study addressed ethical and privacy concerns related to the use of facial recognition (Liu et al., 2021) in online learning by collecting data in raw format, consisting of the count and coordinates of detected faces within the frame, face width and height (in pixels), and face detection confidence. This proactive approach addresses potential privacy implications while ensuring valuable insights into student engagement.

3. Methodology

3.1. Overview of the system

The system comprises two distinct sections, designed separately for the teachers and the students. The teachers are capable of supervising an online class, while the students can attend a class. Utilizing the section of teachers, one may create a class of a course, complete with specific details that will be securely stored in the database, and invite students to attend a class. Additionally, the teachers have the option of monitoring the attentiveness or disruptive behaviors of a student. The section of students, on the other hand, enables them to register for the desired course. They can do so by providing their face image, which is subsequently verified before granting access. While in class, the system will collect data via the webcam of students, which will then be analyzed by the trained modules to determine the attentiveness levels or detect disruptive behaviors of the participants if any. Finally, the system generates a comprehensive anonymous report based on the behaviors exhibited by the students.

The system consists of several modules such as the registration module, face detection module, phone detection module, hand tracking module, and pose estimation module. These modules have been effectively integrated into a web interface. The workflow diagram of this system is illustrated in Fig. 1. Below is a brief overview of each of these modules.

3.1.1. Registration module

Prior to gaining access to the system, it is imperative that a student undergoes the registration process. Within the registration module, the student must provide the essential information, and in addition, photographs are promptly captured for documentation purposes. This significant data is subsequently stored in a centralized database for future utilization in user authentication. Fig. 2 depicts the comprehensive process of registration. Upon successful completion of registration, a student is granted access to the system by means of providing a facial image and other pertinent details.

3.1.2. Face detection

Upon attempting to access the system, the identity of the individual

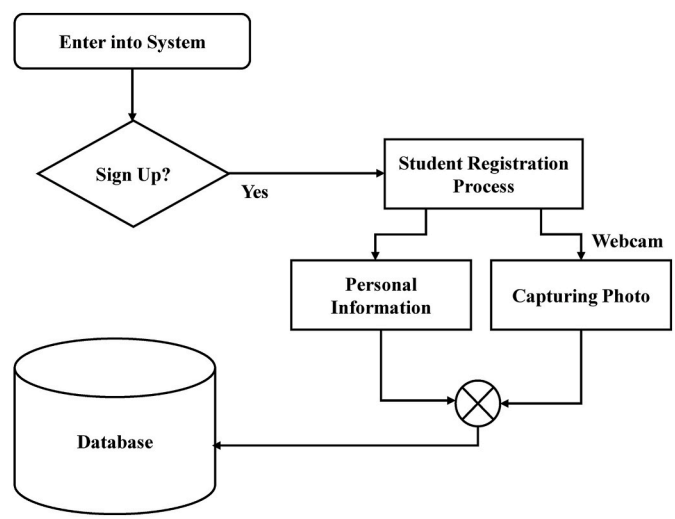


Fig. 2. The process flow of student registration.

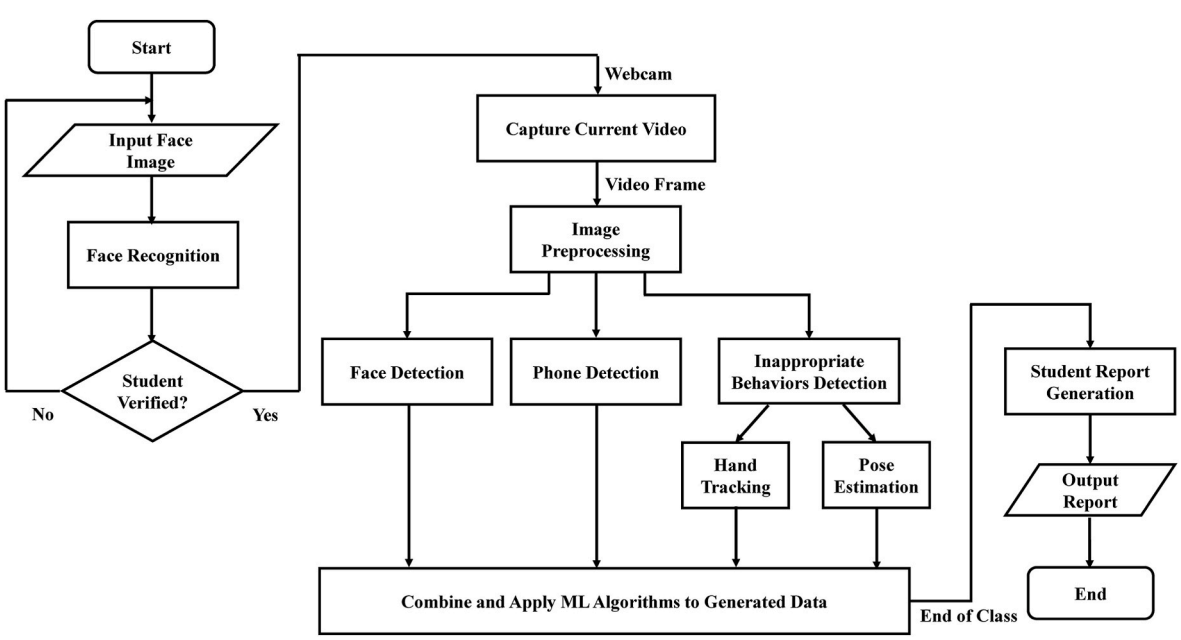


Fig. 1. The workflow diagram of the online class monitoring system.

Fig. 3. Steps of face recognition pipeline.

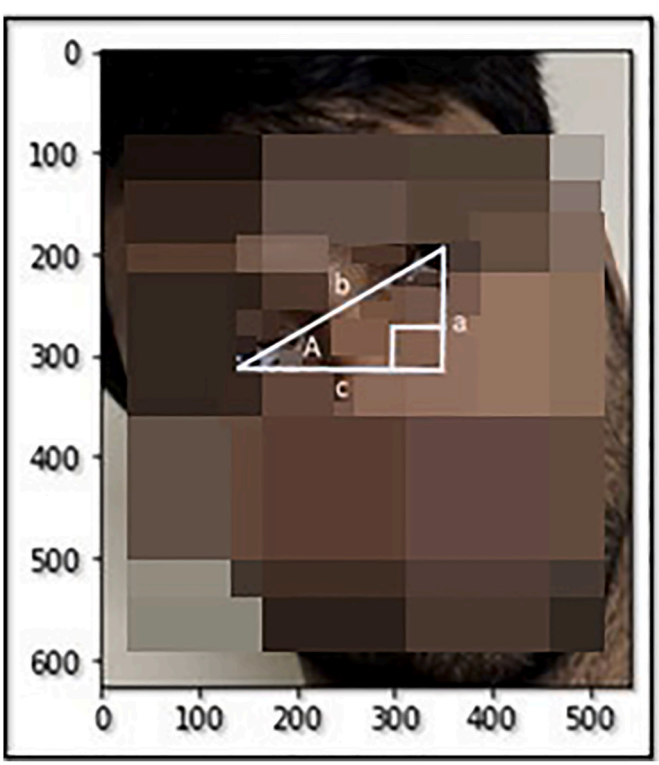


Fig. 4. Face alignment process.

is confirmed utilizing the facial detection module. Additionally, throughout the monitoring of a class session, the continuous face detection module captures facial images from the real-time footage, subsequently generating output that can be categorized into three distinct groups. The first group pertains to the identification of exclusively registered individuals present in front of the webcam, while the second group is concerned with detecting the absence of any person before the webcam. The third and final group concerns itself with recognizing the presence of multiple faces in front of the webcam.

3.1.3. *Phone detection*

During an online classroom session, a student may become preoccupied with employing their mobile device. This monitoring system has the capacity to continuously identify mobile devices that are being used during the online session. Should any be detected, prompt notification will be issued to the instructors of the situation.

3.1.4. *Disruptive behavior detection*

A comprehensive means of identifying disruptive or fraudulent behaviors can be achieved by employing the hand tracking and pose estimation module. A concise summary of these techniques is outlined herein:

3.1.4.1. Hand tracking. Continuous tracking of hand movements plays a crucial role in the monitoring of online classes. This is primarily due to the fact that students in online classes have the freedom to hold any notebook or mobile phone, which can potentially impede their learning. So it is essential to ensure that their hands are tracked at all times. The hand-tracking module consists of two distinct components. One of the components performs verification for the presence of both hands, whereas the other ascertains the absence of a single hand, such as the right hand for an individual who is left-handed. The system will attempt to ascertain if the person is holding an object in such circumstances.

3.1.4.2. Pose estimation. Rotating the head of a student to observe fellow students is a customary practice in conventional classes. However, such behavior is perceived as inattentive during online classes. The task of maintaining a straight head position for prolonged periods can prove to be arduous. Thus, a brief duration of head rotation is not deemed inappropriate conduct. However, consistent head rotation over extended periods

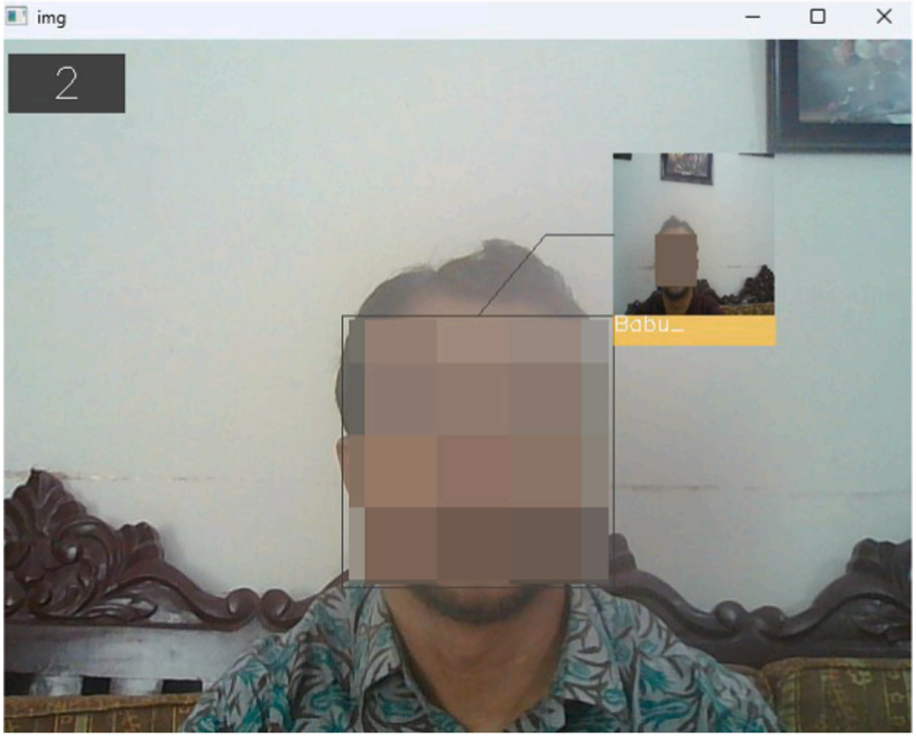


Fig. 5. Face recognition using the VGG-Face model.

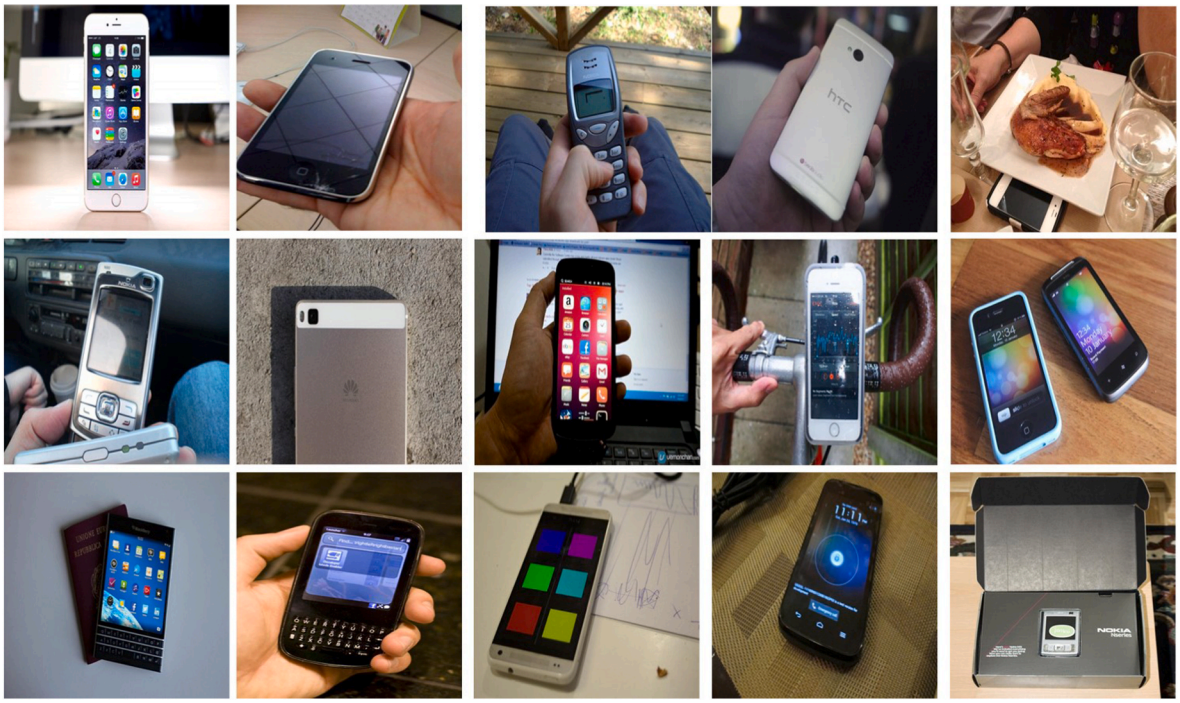


Fig. 6. Sample dataset for mobile phone detection.

or multiple rotations is regarded as disruptive behavior. The pose estimation module is utilized to detect head rotation. During pose estimation, the initial step involves detecting the face. Following face detection, we proceed to ascertain the degree of head rotation. Finally, the frequency or duration of head rotation is calculated.

3.1.5. Combine and apply ML algorithms to generated data

The system is designed to accumulate diverse data, including faces, multiple hands, poses, and phone detection data from their corresponding modules. Specifically, in the face data, the coordinates of the face boxes and detection confidence will be gathered. Likewise, in the pose estimation data, four-direction data (forward, left, right, and down), and angles based on the nose direction will be collected. With regard to phone detection, data on phone detection and the co-ordinate of the phone detection box, as well as the confidence of phone prediction, shall be acquired. The attentiveness score of the class will

subsequently be determined using machine learning algorithms.

3.1.6. Student report generation

When an online course commences, instructors have the ability to observe the present state of the learners. Upon completion of the course, a conclusive anonymous report shall be formulated and archived within the database, which may also be perused by instructors. The aforementioned report, which is system-generated, primarily furnishes the level of attentiveness of the students towards the online course. It also presents a summary of behaviors of the students anonymously throughout online courses, which is obtained through a machine learning algorithm that was trained to employ the outputs of rudimentary system components, such as face detection, phone detection, hand tracking, and pose estimation, amongst others. This report is exclusively viewable by teachers on a designated webpage.

3.2. Implementation details

3.2.1. Tools and technologies

To train attention detection classifiers, Google Colab was utilized. On a personal computer equipped with an Intel Core i5 8th Gen Processor and 16 GB RAM, the remaining implementation was conducted. Python 3.10.7, along with various tools and libraries such as Pandas, Numpy, Scikit-learn, and Django, etc., were employed in the development of the presented system.

3.2.2. Face recognition

The four major stages of the contemporary facial recognition pipeline, including detection, alignment, representation, and verification (Taigman et al., 2014), are depicted in Fig. 3. In our study, we employed a hybrid facial recognition framework, known as Deepface (Serengil & Ozpinar, 2020, October), which encompasses various cutting-edge models such as VGG-Face, Google Facenet, OpenFace, Facebook DeepFace, DeepID, ArcFace, and Dlib. The subsequent subsections briefly describe the aforementioned steps of the facial recognition pipeline.

The initial steps of the face recognition pipeline involve detection and alignment. Visually detecting faces and eyes is performed using the OpenCV library (Bradski, 2000), but the library does not include face alignment as a standard feature. Google researchers have discovered that incorporating face alignment algorithms into the preprocessing of source images leads to a notable increase in the accuracy of the face recognition model, from 98.87% to 99.63% (Parkhi et al., 2015). To correctly align faces, trigonometric techniques are utilized and a right-angled triangle is drawn based on the position of the eyes as depicted in Fig. 4. The angle A for alignment is calculated using the cosine rule of Eq. (1) and inverse trigonometric function, and the base image is rotated by the calculated angle A (Parkhi et al., 2015).

$$cosA = \left(b^2 + c^2 - a^2\right)/2ab \tag{1}$$

Facial recognition models make use of CNNs to convert facial images to vectors. These models are trained to identify people and extract raw feature vectors for faces (Taigman et al., 2014). The Deepface framework supports various CNN models with specific input-output shapes as specified in (Parkhi et al., 2015). In this paper, the VGG-Face Model is utilized, which takes in images of size (224 × 224 × 3) and produces a 2622-dimensional vector as an image representation.

In the representation stage, face images are represented as vectors. The verification module compares two vectors of the same dimensions and determine their similarities by using different distance measures such as cosine similarity, Euclidean distance, or its L2 form. The Euclidean distance is computed using the length of the line connecting two vectors, as specified in Eq. (2), with low distance indicating the same person and high distance indicating different persons (Parkhi et al., 2015).

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{2}$$

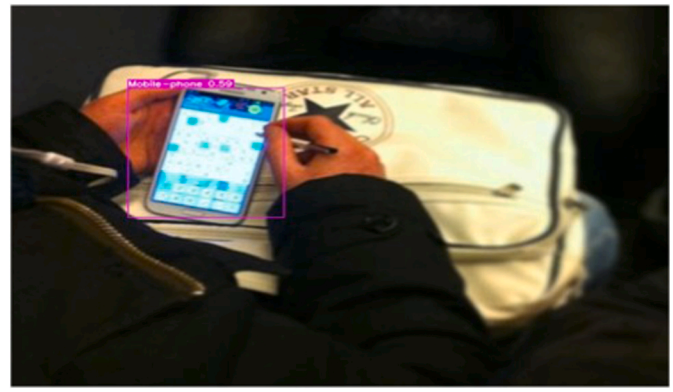


Fig. 7. Mobile phone detection using YOLOv7.

Upon logging into the system, the student is required to provide a facial image, which is subsequently processed and analyzed by the system. The analysis entails comparing the image to the stored representations in our database. A successful match is contingent on the smallest Euclidean distance between the two representations being below 0.5, thus facilitating successful login of the student. Fig. 5 provides visual illustrations of matching instance of facial image of the student.

3.2.3. Mobile phone detection

YOLOv7, the latest iteration in the YOLO family (Wang et al., 2023) of real-time object detection algorithms, was introduced in July 2022. Compared to its predecessors, including YOLOv5 and others, it stands out for its exceptional real-time object detection capabilities with improved speed, accuracy, and reduced computational requirements. YOLOv7 incorporates elements like E-ELAN for better learning, model scaling for enhanced performance, and a trainable bag-of-freebies for increased accuracy without added inference costs. Remarkably, it has been trained on the MS COCO dataset (Lin et al., 2014) without additional data sources or pre-trained weights, making it efficient for object detection on smaller datasets. The YOLOv7 paper (Wang et al., 2023) has received significant attention, and its source code is available on GitHub, making it accessible to a wide range of computer vision applications.

In order to implement the module for mobile phone detection, we utilized YOLOv7. YOLOv7 was pre-trained on the large-scale MS COCO dataset (Lin et al., 2014), consisting of over 200,000 labeled images and 80 classes including the "cell phone" class. Examples of the "cell phone" class are presented in Fig. 6. Currently, there exist three distinct YOLOv7 models, namely, YOLOv7, YOLOv7-tiny, and YOLOv7-W6. Our selection for this study was YOLOv7 as it is the fastest and most accurate basic model optimized for ordinary GPU computing.

An instance of the mobile phone detection module in action within the visual surroundings of the students is illustrated in Fig. 7. This module outputs six high-level features, which are as follows: (i) the presence of a mobile phone, (ii) x-coordinate of phone, (iii) y-coordinate of phone, (iv) width of phone, (v) height of phone, and (vi) phone detection confidence (in case of detection). These high-level features are subsequently put to use in the attention detection model.

3.2.4. Disruptive behaviors detection

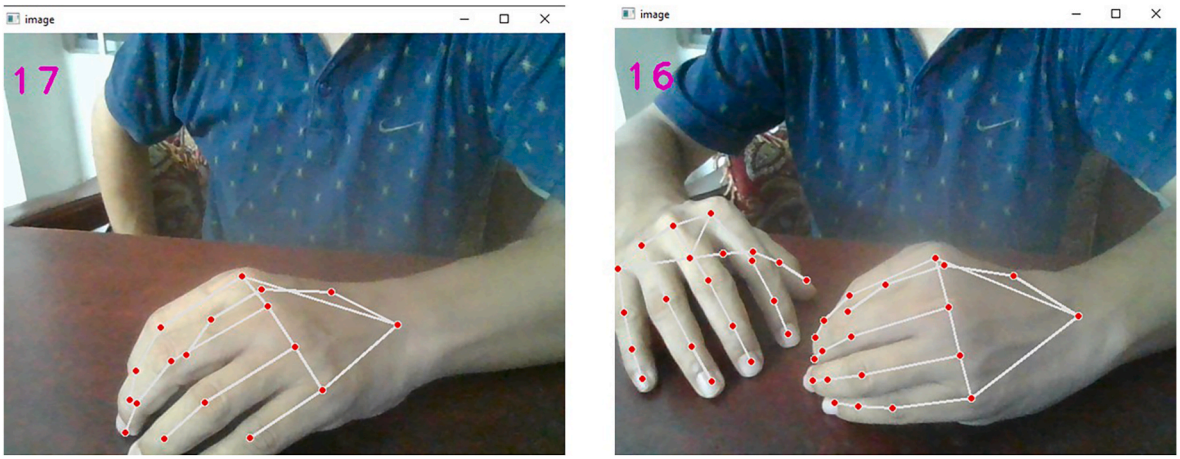
3.2.4.1. Hand tracking. The MediaPipe library (Lugaresi et al., 2019), developed by Google, is utilized by the system in order to detect and track a hand. Algorithm 1 outlines the method for hand-tracking tasks. It should be noted that both the detection confidence and the tracking confidence are set to 50 percent. As depicted in Fig. 8, an example is provided.

Algorithm 1. Hand tracking procedure using MediaPipe.

Input: Video frames with single or both hands

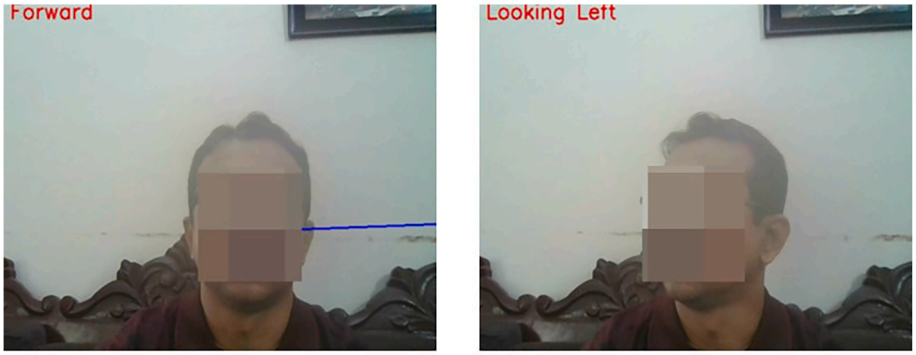
Output: Hand landmarks, Position and orientation of hands

i. Initialize the MediaPipe hand-tracking module.
ii. Capture video frames from the web camera by utilizing the openCV, and then convert the image frame from BGR to RGB format.
iii. Process each RGB frame using MediaPipe model to detect hand landmarks.
iv. If a hand is detected:
    a. Extract the 2D coordinates of the hand landmarks.
    b. Determine the position and orientation of the hand by using the hand landmarks.
    c. Display the hand landmarks and show the connection of the points on the hand.
v. Display the processed frame in the OpenCV window.
vi. Repeat steps (iii) – (v) until all frames have been processed.
vii. End.



(a) (b)

Fig. 8. Hand detection using MediaPipe: (a) Left hand detection, and (b) Left-right hand detection.



(a) (b)

Fig. 9. Pose estimation using solvePnP: (a) Looking forward, and (b) Looking left.

3.2.4.2. Pose estimation. In this particular system, pose estimations are characterized as the rotation of the head in a distinct direction. In order to perceive this head rotation, the initial step is to detect the face. Numerous machine-learning approaches exist for the detection of faces, and in our case, the MediaPipe library is utilized for this purpose. Once the face has been detected, the solvePNP method is employed to determine the rotation of the head. Algorithm 2 outlines the steps involved in attention analysis tasks, which entails using the face mesh module of MediaPipe to recognize facial landmarks, followed by the solvePnP() function of OpenCV to estimate the head pose. Ultimately, we obtain rotation angles in relation to the nose directions, which represent the primary axis. An example of pose estimation is depicted in Fig. 9.

Algorithm 2. Pose estimation procedure using MediaPipe and Perspective-*n*-Point (PnP) function

Input: Video frames with faces

Output: Face landmarks, Head pose (degree of rotation)

i. Initialize the MediaPipe face mesh module and OpenCV window for visualization.
ii. Capture video frames from the web camera by utilizing the openCV, and then convert the image frame from BGR to RGB format.
iii. Process every RGB frame using MediaPipe face mesh to identify all face landmarks.
iv. If any faces are detected, for each face:
    a. Extract the 2D coordinates of the face landmarks.
    b. Calculate the z-coordinate by the face landmarks and thus convert the 2D coordinates into the 3D model points, representing the nose tip, chin, eye corners, and mouth corners.
    c. Estimate the camera matrix using the properties of the image obtained from the web camera (focal length and center) and distortion coefficients.
    d. Obtain the rotation vector and translation vector, representing the head pose, using solvePnP() function of OpenCV, passing the 3D model points and 2D image points.
    e. Acquire the degree of rotation from the rotation matrices using decomposition functions of OpenCV to estimate the pose.
v. Display the processed frame in the OpenCV window.
vi. Repeat steps (iii) – (v) until all frames have been processed.
vii. End.

3.2.5. Dataset preparation

Each basic component of the system namely face detection, hand tracking, mobile phone detection, and pose estimation module outputs a set of distinct high-level features. For example, the face detection module produces the count of detected faces within the frame, the coordinates of the face, and the width and height (measured in pixels) of the face. The pose estimation component generates the head orientation, x-axis rotation, and y-axis rotation of the head. The hand-tracking module produces the number of hands identified in the current frame. These high-level features, collected from the fundamental components of the system, are combined to form the attention detection dataset.

In the process of collecting data from human subjects, adhering to ethical guidelines and ensuring informed consent is crucial to protect the rights and privacy of the participants. Approximately 30 undergraduate students volunteered for this research study, and data collection was conducted meticulously in line with proper protocols. This involved obtaining informed consent from the participants, prioritizing privacy and confidentiality, utilizing non-invasive data collection methods, and emphasizing the comfort and safety of the participants throughout the process.

The dataset comprises 17 columns, with 16 feature columns and one label column, and it encompasses 4000 records (Hossen & Uddin, 2023). The features of the dataset are described in Table 1, as outlined. The aforementioned features are then utilized to train a machine learning classifier, which makes continuous decisions regarding the level of attentiveness of students. The dataset is freely available in the Mendeley Data Repository (https://doi.org/10.17632/smzggbnkd2.1).

3.2.6. Integration and user interface

All of the modules have been developed, trained, and evaluated utilizing appropriate datasets. Our ensuing objective is to formulate a structure that effectively integrates these modules. To achieve this objective, we created a web interface for the system utilizing HTML and CSS in the front end and Django (Django, 2023) in the back end. During online classes, we fetch the feedback from Django, which generates a frame. This frame is imperative for the operation of all the modules. We take the frame and convert the color of the image from BGR to RGB prior to forwarding it to our designated module. Our system functions entirely on frames, thus the server (Django, 2023) transmits the image frame, processes it in the 'view.py' file, and returns the amended image to the server. The website displays the streaming video in its usual form. Once the image processing has been completed, we obtain an array that represents the attention status of the student. This array is subsequently stored in a CSV file, which is used for evaluating the attention detection model. Eventually, we integrate our model into the 'view.py' file of Django, combining and processing the array within the model to obtain the desired output, indicating the level of student attentiveness. The user interface is partitioned into teacher and student sections.

Table 1
Description of students' attention detection dataset columns.

Column	Description	Column	Description
no_of_face	Number of faces found in the frame	**pose_y**	Face orientation in y-axis
face_x	X-coordinate of the upper-left corner of the face	**phone**	1: Phone detected 0: No phone detected
face_y	Y-coordinate of the upper-left corner of the face	**phone_x**	X-coordinate of the upper-left corner of the detected phone
face_w	Width of the face (in pixels)	**phone_y**	Y-coordinate of the upper-left corner of the detected phone
face_h	Height of the face (in pixels)	**phone_w**	Width of the phone (in pixels)
face_con	Face detection confidence	**phone_h**	Height of the phone (in pixels)
no_of_hands	Number of detected hands	**phone_con**	Phone detection confidence
pose	Face orientation (Forward | Down | Left | Right)	**label**	Target column (0: Attentive, 1: Inattentive)
pose_x	Face orientation in x-axis		

3.2.6.1. Teacher section. In the teacher section, an instructor is afforded the opportunity to create and furnish their profile inclusive of their name, email, department, contact number, and other relevant information, as shown in Fig. 10(a), which is duly stored within the 'teachers' table. To initiate an online course, the instructor must complete the form for course creation by furnishing the pertinent information such as course title, course code, credits, class time, and other related data, as shown in Fig. 10(b), which will also be kept in the 'course' table. During the commencement of a virtual course, the educator is able to monitor the current progress of a student on the class report page. Upon conclusion of the course, an anonymous report of attentiveness will be produced and the instructor can easily access and review same from the same page.

3.2.6.2. Class report page. The duration of the class and the anonymous report of their attentiveness are all easily accessible through this webpage. The level of attentiveness demonstrated by the students is presented as a percentage and was previously mentioned to be generated through the utilization of a machine learning algorithm. Additionally, a dropdown menu will appear when the mouse is moved over the attentiveness percentage showcasing various student behaviors such as multiple face, hand tracking, pose estimation, and phone detection information. The generated reports can serve as a valuable feedback mechanism that empowers educators to make data-informed decisions, create more engaging learning experiences, and provide targeted support to students.

3.2.6.3. Student section. In the student section of web interface, various modules are integrated to execute diverse tasks. Multiple webpages have been devised to furnish registration details of students, exhibit course particulars, and enable class streaming video preview. In the ensuing sections, an overview of these pages has been presented.

3.2.6.4. Registration page. On this particular page, the student is required to provide essential credentials such as name, identification number, email address, password, and images of their face. The face registration module adeptly captures five images of the student via the webcam and subsequently saves them in a database. The registration interface is portrayed in Fig. 11(a), while the process of registering the face of student through the webcam is illustrated in Fig. 11(b).

3.2.6.5. Class view page. Herein, a student is enabled to perceive the specifications of forthcoming courses. Through this webpage, a student is afforded the opportunity to enroll in an online class i.e., join an online class through Zoom, Google Meet, etc. The page displaying the course information is illustrated in Fig. 12.

3.2.6.6. Class preview page. Upon enrollment of a student in an online class, the system promptly initiates the collection of data from the web camera, which is then transmitted to its fundamental modules, including face detection, pose estimation, hand tracking, and phone detection, subsequently yielding multiple outputs. The machine learning model then processes these outputs and predicts the level of attentiveness. Students can access a report on this page, which comprehensively outlines their attentiveness, encompassing details regarding their conduct, such as multiple face detection, hand tracking, pose estimation, and phone detection.

4. Results and discussions

The objective of the attention monitoring system implemented in online classes was to evaluate the level of student engagement and attentiveness during remote learning sessions. In this segment, we delve into the dataset utilized for the attention detection classifier. Afterward, a comprehensive evaluation of the performance of classifiers was conducted utilizing various evaluation metrics, including precision, recall,

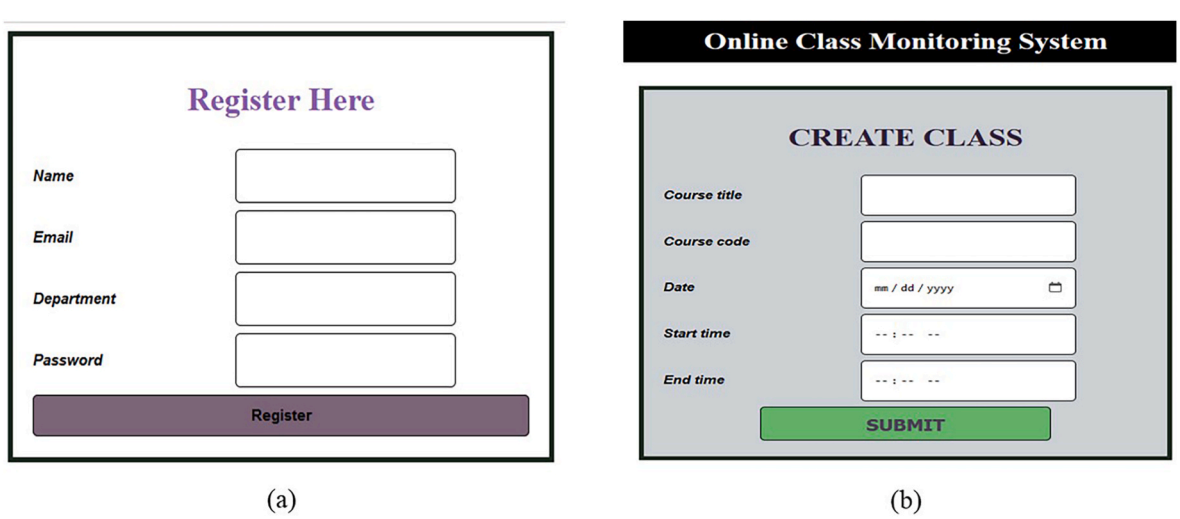


Fig. 10. Profile and course creation: (a) Teachers profile information, and (b) online course creation form.

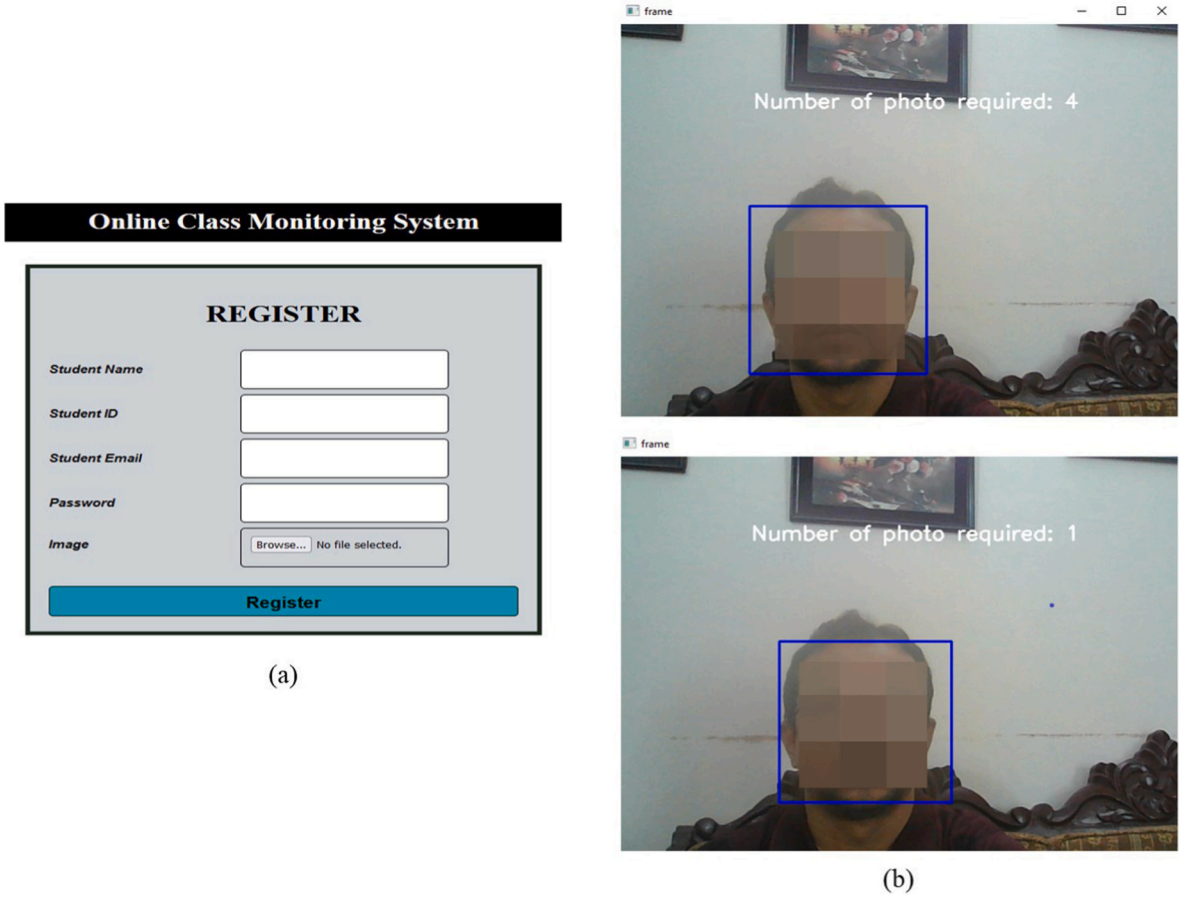


Fig. 11. The registration process of student with images: (a) Registration form, and (b) Face images of the student.

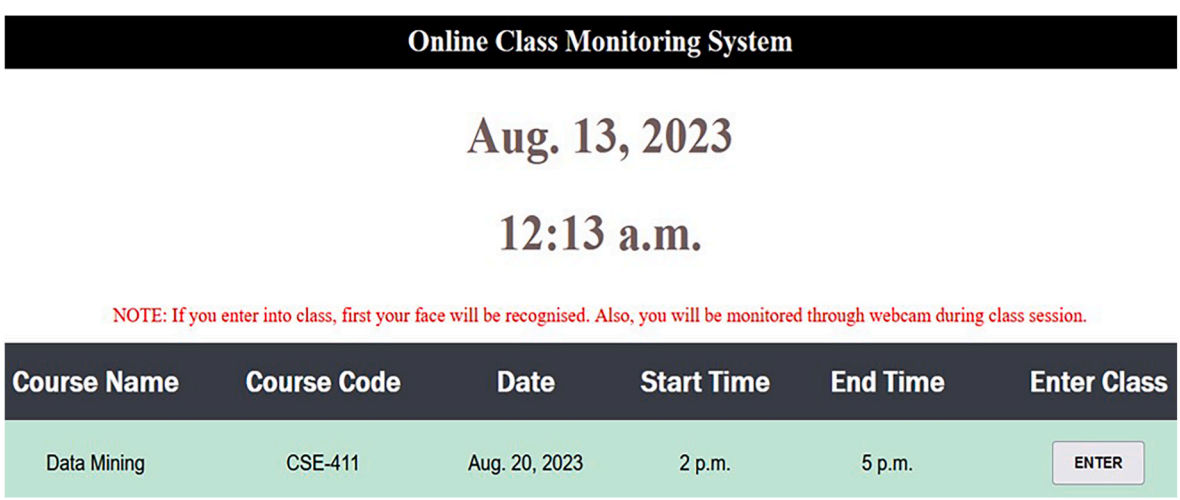


Fig. 12. Scheduled online class information.

accuracy, and F1-score. Lastly, we present the outcomes obtained from the monitoring system data collection and provide a comprehensive analysis of the findings.

4.1. Results analysis

4.1.1. Evaluation metrics

In this discourse, we delve into diverse evaluation metrics that are commonly implemented to gauge the efficacy of classification models. These include accuracy, precision, recall, and F1-score. Accuracy is a statistical parameter that amalgamates the performance of a model across all categories. Precision pertains to the capacity of a classification model to exclusively identify pertinent data points. Precision is widely used as a predictor of positive outcomes. Recall is a metric that measures the ability of a model to forecast a class. The F1 score is defined as the harmonic mean of precision and recall. Elaborate explication of these metrics is obtainable in (Powers, 2020). In the following equations, from Eq. (3) to Eq. (6), we establish accuracy, precision, recall, and F1-score

in terms of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$Precision=\frac{TP}{TP+FP} \tag{4}$$

$$Recall=\frac{TP}{TP+FN} \tag{5}$$

$$F1_score=\frac{2*Precision*Recall}{Precision+Recall}=\frac{2*TP}{2*TP+FP+FN} \tag{6}$$

Where,

True Positive (TP) - the classifier classifies the right class as right,
True Negative (TN) - the classifier classifies the wrong class as wrong,
False Positive (FP) - the classifier classifies the wrong class as right,
False Negative (FN) - the classifier classifies the right class as wrong.

4.1.2. Dataset analysis

These high-level features, collected from the fundamental components of the system, are combined to form the attention detection dataset, which is provided in Table 2 with only a few rows and major columns.

The dataset in its entirety comprises 4000 individual records and has been partitioned into two distinct subsets, utilizing an 80:20 ratio, subsequent to shuffling. The training dataset, which is the first subset, is employed to fit the model. On the contrary, the second subset known as the test dataset is not utilized to train the model. Instead, it is presented to the trained model to generate predictions, which are then compared against anticipated values. Following the division of our dataset into train and test sets, the training dataset encompasses 3200 records, while the test set contains 800 records. The primary objective of this partition is to evaluate the performance of the models on the test data.

In general, the scores assigned to feature importance reflect the level of usefulness or value that each feature contributes towards the construction of boosted decision trees within the model. A higher score assigned to the importance of a feature indicates that the respective attribute plays a more substantial role in facilitating critical decisions with the trees. Fig. 13 provides an illustration of the importance assigned to the top seven features in the dataset. The findings reveal that 'pose_y' has the highest importance score, while 'phone' has the lowest score among the top seven features.

The outcomes of k-fold cross-validation, where k = 3, as depicted in Table 3 (only the first and last three rows have been shown) for the XGBoost classifier demonstrate that the mean root-mean-square error (RMSE) for both the training and test datasets decreases with the increasing rounds of boosting. This affirms that the model is enhancing its performance and precision in forecasting the targeted variable. The standard deviations of RMSE values, however, remain almost stable, indicating a steady generalization ability. These findings suggest that the XGBoost classifier is proficiently learning from the data and holds the potential to perform well on unfamiliar data.

4.1.3. Performance analysis of classifiers

To accomplish the task of detecting attention, the utilization of six well-established machine learning classifiers was undertaken. Our attention detection dataset was used to train and test these classifiers, namely the Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), AdaBoost, Gradient Boosting (GB), and Extreme Gradient Boosting (XGBoost). A summary of the performance of these machine-learning models is presented in Table 4. Upon in-depth analysis of the performance of the classifier models, it was discovered that XGBoost outperformed the other models.

Fig. 14 demonstrates the evaluation of the machine learning algorithms on the test dataset in the form of confusion matrices. Analysis of the confusion matrices reveals that the conventional Support Vector Machine Classifier (SVC) demonstrates inferior performance when compared to the remaining ensemble-based classifiers, with 28 and 66 False Positive and False Negative values, respectively. In contrast, the Extreme Gradient Boosting (XGBoost) Classifier surpasses all other models, with True Positive, False Positive, True Negative, and False Negative values of 348, 1, 450, and 1, respectively.

In simpler terms, XGBoost can be characterized as a regularization technique for the pre-existing gradient-boosting algorithm. Owing to this attribute, it exhibits superior performance as compared to the typical gradient-boosting algorithm, thereby rendering it much faster. Additionally, it yields better accuracy in datasets containing numerical and categorical features. As a result, the XGBoost classifier was chosen to undertake the task of attention detection.

4.1.4. Key findings and contributions

In our research, we present significant findings bridging technology-enhanced learning and machine learning. We developed a specialized system with components like face detection, hand tracking, cell phone detection, and pose estimation, creating a robust platform for monitoring student behavior during online classes. Through rigorous evaluation, we identified the XGBoost algorithm as optimal for attention detection, achieving 99.75% accuracy on test data.

We identified various behavioral cues, enriching our understanding of student engagement, including facial expressions, hand movements, and pose variations. We recognize the contextual nature of student engagement, acknowledging that disengagement can be influenced by factors like dull content or individual circumstances. We've also created an intuitive, user-friendly interface for simplified attentiveness monitoring, enhancing accessibility for educators. Our system generates

Table 2
Sample of students' attention detection dataset.

no_of_face	face_x	face_y	no_of_hand	pose	pose_x	pose_y	phone	phone_x	phone_y	label
1	232.66	222.12	0	forward	5.28	3.59	1	214	349	1
1	309.56	282.00	0	forward	6.77	4.43	0	0	0	0
2	256.32	246.20	2	right	32.59	36.57	1	117	137	1
1	293.00	285.86	2	forward	−4.01	−9.91	0	0	0	1
1	307.53	210.03	0	left	3.74	−14.43	0	0	0	1
1	364.80	272.80	0	down	−15.00	10.00	0	0	0	1
1	265.47	262.19	2	forward	−0.17	−2.40	0	0	0	0
1	286.80	212.35	1	left	23.21	−20.21	1	125	300	1
1	293.75	277.36	1	forward	−7.57	−7.57	0	0	0	0
1	219.98	232.04	1	right	3.99	16.82	0	0	0	1
0	0.00	0.00	0	forward	−1.17	−9.46	0	0	0	1
1	271.28	245.71	2	forward	−6.68	−0.06	0	0	0	0
1	344.34	229.25	0	forward	10.71	−9.52	0	0	0	0
1	280.01	139.99	1	down	−10.97	−1.65	1	155	286	1

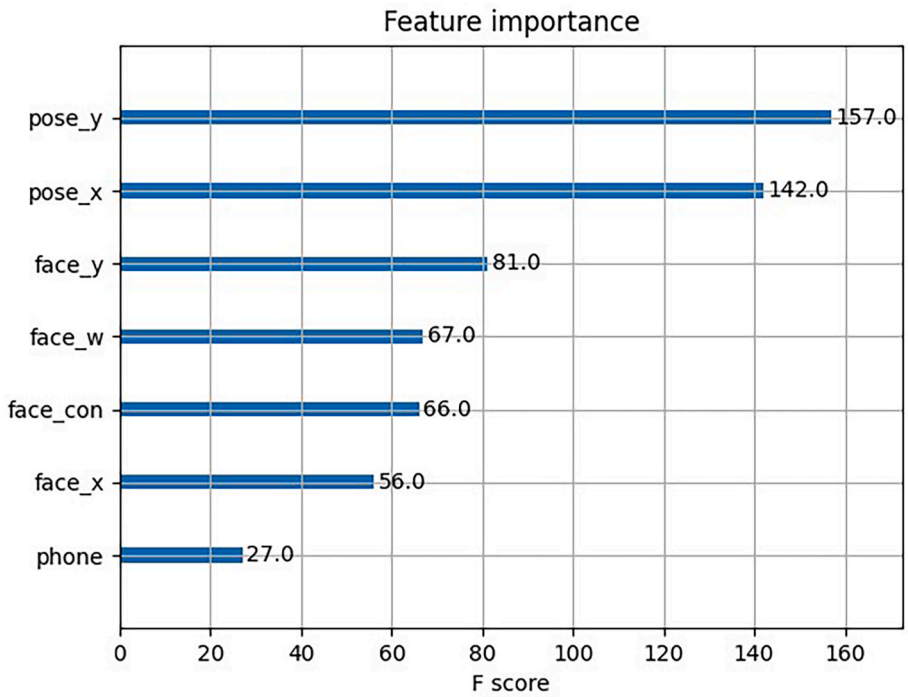


Fig. 13. Top seven features of the attention detection dataset.

Table 3
The results of k-fold cross-validation of XGBoost classifier on the dataset.

Round	train-rmse-mean	train-rmse-std	test-rmse-mean	test-rmse-std
0	0.484477	0.000237	0.484702	0.000221
1	0.452246	0.010012	0.452390	0.009420
2	0.437895	0.014059	0.437982	0.013975
.	.	.	.	.
.	.	.	.	.
47	0.156848	0.007522	0.165182	0.006169
48	0.155411	0.006988	0.163817	0.007024
49	0.154428	0.007724	0.162822	0.006492

Table 4
Performance of machine learning models on attention detection dataset.

Model	RMSE	Precision	Recall	F1-score	Accuracy	AP
Support Vector Machine	0.3427	0.9099	0.8108	0.8575	0.8825	0.8203
Decision Tree	0.1322	0.9772	0.9828	0.9800	0.9825	0.9679
Random Forest	0.1658	0.9969	0.9398	0.9675	0.9725	0.9632
AdaBoost	0.1274	0.9941	0.9684	0.9811	0.9837	0.9765
Gradient Boosting	0.1172	0.9941	0.9742	0.9840	0.9862	0.9797
XGBoost	**0.0500**	**0.9971**	**0.9971**	**0.9971**	**0.9975**	**0.9955**

detailed anonymous reports after each class, offering a comprehensive view of student attentiveness and behaviors.

4.2. Discussions

4.2.1. Privacy and ethical considerations

In the contemporary educational context, adopting the Foucaultian approach (Foucault, 1977) provides a crucial perspective for the examination of privacy and ethical considerations in attention monitoring. This approach highlights that surveillance extends beyond a technical process; it is a tool of power and control deeply embedded in the structure of educational institutions. Foucault's theory of power-knowledge dynamics (Foucault & Gordon, 1980) is pertinent as we consider the privacy and ethics of our monitoring system. Surveillance represents the exercise of power through knowledge. It's vital to recognize its potential to create power imbalances, enabling those wielding surveillance tools to exert influence over those monitored. In light of this, our ethical duty is to scrutinize how our system might affect power dynamics within educational settings, ensuring that students are not subjected to arbitrary surveillance that undermines their autonomy.

Drawing inspiration from Foucault's emphasis on self-determination and autonomy, our approach prioritizes informed consent and empowerment. We actively involve students in the monitoring process, ensuring they grasp the objectives, mechanisms, and implications of surveillance. In doing so, we treat students as active participants in their educational journey, allowing them to make informed choices regarding their engagement. This aligns with Foucault's ideas of self-awareness and active participation within governing systems. We acknowledge that data collected through surveillance can be a source of power. In response, we implement stringent data protection measures, including anonymization, to guard against any potential misuse of personal information.

Through the Foucaultian lens, our privacy and ethics approach is rooted in profound respect for individual autonomy, informed participation, rigorous data protection, and the recognition of surveillance as a system of power. Our commitment to transparent, ethical practices aims to harmonize educational surveillance with Foucault's principles of autonomy and self-awareness.

4.2.2. Teaching quality considerations

Indeed, the level of attention in a learning environment can be profoundly influenced by the quality and engagement level of the lecture. We concur that effective teaching methods and content delivery play a pivotal role in shaping student attention. In our research, we strive for a non-intrusive approach to comprehensively capture the intricate nature of attention patterns in online classes. We recognize that numerous factors, including the quality of instruction, can impact

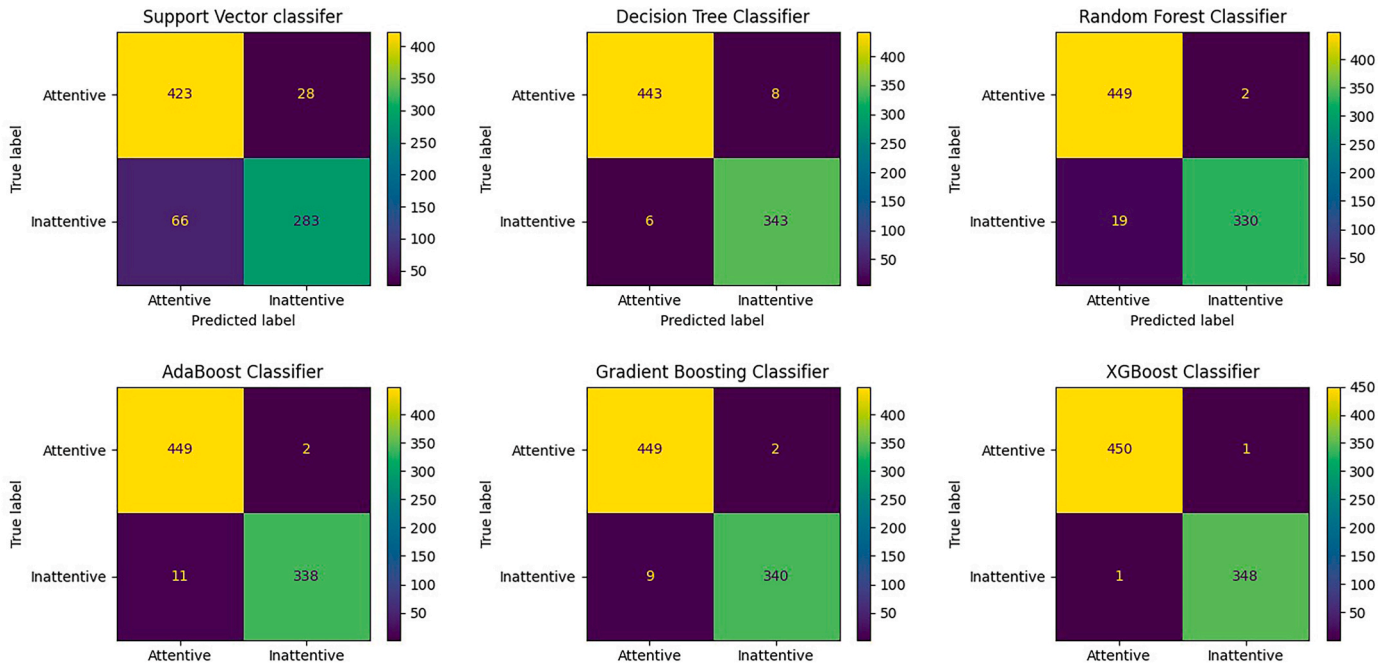


Fig. 14. Confusion matrices of machine learning classifiers on test data.

student engagement. Our emphasis lies in presenting attention monitoring as a valuable diagnostic tool for educators, offering insights to enhance their teaching methods and course materials.

It's noteworthy that the teacher involved in our evaluation has expressed significant appreciation for the support provided by our proposed tool. The implementation of an automatic system for improving student attention can help maintain focus during lectures and facilitate attention checks or solicit questions in instances of potential lapses in concentration. By adopting this perspective, our research seeks to contribute to a holistic understanding of attention in online learning, considering both learner behavior and the instructional context, ultimately enriching the online education experience.

4.2.3. Evaluation and practical implications

A conclusive experiment was carried out to analyze the conduct of two students who were not involved in the creation of the dataset, in order to generate an anonymous report on the degree of attention exhibited by the students. The students participated in a brief online class of approximately 5 min and displayed both attentive and inattentive behaviors during the session. The behaviors depicted in Fig. 15 provide a visual representation of some of the observed activities.

Utilizing the identified behaviors in each frame, the XGBoost machine learning model produced the level of attentiveness and saved it in a file named 'attention_stat.csv'. Upon processing all the frames, the percentage of attentiveness during the class was computed. The data was collated, and a class report containing all the behaviors identified was generated, as presented in Fig. 16. The initial student (Student-1) had lower attentiveness at 67%, while the second student (Student-2) had a higher degree of attentiveness at 94%. Such observation can benefit teachers and institutions in identifying the cause of student inattentiveness and making appropriate decisions.

The findings of this study bear practical implications to transform online teaching practices by offering a data-driven approach to understanding and enhancing student engagement. By proficiently identifying and addressing fluctuations in student attention, educators can tailor teaching strategies for superior outcomes. The system, which generates comprehensive anonymous reports on student attentiveness, offers invaluable insights for delivering personalized feedback and timely interventions, ultimately elevating the quality of the learning experience.

Moreover, this study underscores the pivotal role of data-driven decision-making in the realm of online education. Educators and institutions can harness these insights to fine-tune curriculum design and optimize resource allocation, thereby enriching overall instructional quality. Furthermore, the ability to identify at-risk students presents an opportunity for early intervention and support, curbing dropout rates. By leveraging data and machine learning, educators and institutions can craft online learning environments that are not only more engaging and responsive but also more effective. However, it remains essential to strike a balance between data-driven interventions and safeguarding students' autonomy and privacy within the online learning environment.

4.2.4. Limitations

Our work comes with several inherent limitations that warrant consideration. Recognizing these limitations is vital for a comprehensive understanding of the scope and applicability of our system in assessing student attentiveness. Firstly, akin to traditional face-to-face classes, our system predominantly relies on observable behaviors to gauge student attentiveness. Consequently, it may struggle to fully capture instances where students are physically present but mentally disengaged, a challenge common to both online and offline classrooms.

Secondly, our approach primarily quantifies attention levels through technological means, potentially neglecting the nuanced nature of student engagement, including legitimate peripheral participation as described by the community of practice theory (Lave & Wenger, 1991; Wenger, 1998). This focus on quantification may not encompass the richness of student interactions and involvement in the learning process. Moreover, our system heavily depends on observable behavioral cues for assessing student attentiveness, which may not account for nuanced scenarios such as students being drowsy due to medication or sleep deprivation. Additionally, the system's categorization of behaviors, such as head rotation, as "inappropriate," may not adequately accommodate context-specific variations in learning environments, limiting its ability to provide a comprehensive analysis of student engagement.

In our study, although we examined body language signifiers as indicators of attention, we acknowledge that these cues carry a broader range of potential meanings beyond inattention, encompassing various emotional states and cognitive engagement levels. This complexity

Fig. 15. Some examples of the observed activities detection: (a) Single face detection, (b) Double face detection, (c) Multiple face detection, (d) Looking right, (e) Looking down, (f) Right-hand detection, (g) Both hand detection and (h) Mobile phone detection in different orientations.

inherent in human communication through body language necessitates further exploration in future research. Another noteworthy limitation is the absence of real-time analysis of students' facial expressions to identify emotional states like boredom, frustration, yawning, and tiredness. This omission may compromise the effectiveness and efficiency of our study in capturing the full spectrum of student engagement.

Lastly, our system overlooks aspects like eye-gaze tracking and drowsiness detection, which could contribute valuable insights into the level of engagement of a student. Consequently, there may be instances

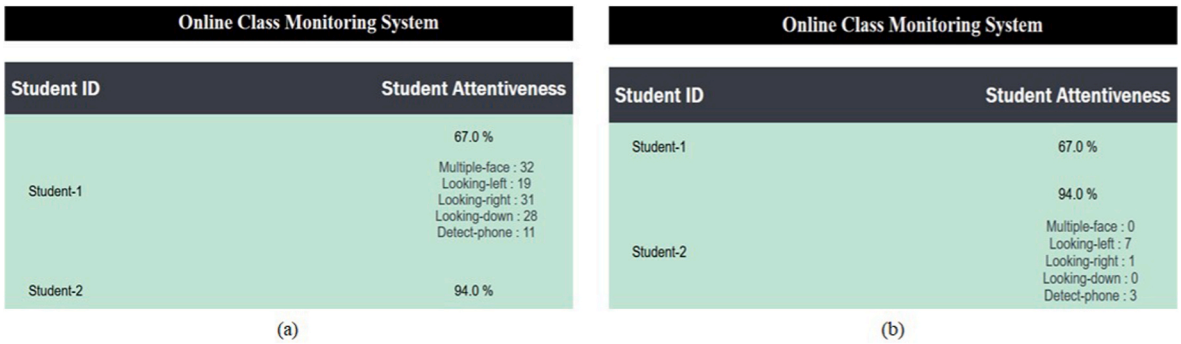


Fig. 16. Some examples of attention levels: (a) Less attentive, and (b) More attentive.

where the proposed system does not accurately reflect the true state of attentiveness of a student. These limitations underscore the need for ongoing refinement and expansion of our methodology to provide a more comprehensive and nuanced assessment of student engagement in various learning environments.

4.2.5. Future implementation and research

The future of attention-monitoring systems holds promising opportunities. Cutting-edge technologies like affective computing and multimodal analysis offer the potential for more precise attention evaluation. Multimodal data processing, incorporating facial expressions, body language, speech and vocal patterns, and interactions with learning materials, can enhance accuracy. For instance, speech recognition algorithms can assess audio inputs and active participation, enabling a comprehensive evaluation of student engagement. Real-time feedback integration is also feasible, facilitating swift intervention and re-engagement with disengaged students, thus benefiting education through immediate attention-level feedback or instructional adjustments.

Advancements in technology, including eye-tracking and drowsiness detection systems, may offer more accurate and unobtrusive methods to assess student participation. For instance, when a student's drowsiness surpasses a predefined threshold, real-time alerts can notify teachers and administrators. Future attention monitoring systems could further benefit from analyzing spoken or written responses to questions of the students in online courses, leveraging developments in natural language processing (NLP).

Additionally, future improvements in attention monitoring systems may include the analysis of emotions such as boredom, confusion, focus, frustration, yawning, sleepiness, and tiredness. Recognizing the emotional context can enhance the effectiveness of the system. Addressing the challenge of students physically attending while being inattentive, similar to face-to-face classes, future research and system refinement may explore integrating physiological measures or self-reporting mechanisms to gauge internal engagement levels of the students.

Furthermore, our study may pave the way for future research that comprehensively investigates the relationship between engagement, legitimate peripheral participation, and online learning outcomes. Our forthcoming research will broaden its scope to comprehensively examine the complexities of online learning. Beyond technical aspects, we will explore factors like instructional design, content relevance, and individual differences to understand their impact on student engagement and attentiveness. This research aims to refine attention-monitoring technology to improve educational outcomes in diverse learning environments. Lastly, as an iterative and adaptive surveillance system, continuous assessment and refinement based on student feedback are possible. This approach aims to address privacy concerns and enhance the supportiveness and non-intrusiveness of the system.

5. Conclusion

In this paper, we present an innovative multimodal analytic methodology that is tailored precisely for the rigorous monitoring of online classes. This approach clarifies a complex area by concentrating on the major issues related to the evaluation of student participation and attention in virtual learning environments. The proficiency of this system lies in its ability to meticulously record and scrutinize student behavior throughout online lessons. This is made possible by the seamless integration of pivotal components, which incorporate robust features such as face recognition, hand tracking, mobile phone detection, and pose estimation modules. A salient highlight is the exceptional accuracy demonstrated by the extreme gradient boosting (XGBoost) method, which achieves an impressive 99.71% precision in effectively assessing the attention levels of the students. The capabilities of the system go beyond accuracy, generating comprehensive and granular reports that expound on the nuances of attentiveness. These detailed anonymous reports equip educators with invaluable data-driven insights, fostering the implementation of targeted interventions to enhance the quality of student learning experiences. The precision and user-friendly architecture of this system make it an indispensable tool for educators who aspire to create immersive and captivating online learning environments. In the future, enhancing the system may involve integrating new input modalities for additional data dimensions. Also, evaluating the effectiveness of the system across different student groups and educational contexts could provide valuable insights. These explorations have the potential to reshape online learning, making it more engaging and effective.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethics statement

All procedures of data collection were performed in compliance with relevant laws and institutional guidelines, and informed consent was obtained for experimentation with human subjects.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abate, A. F., Cascone, L., Nappi, M., Narducci, F., & Passero, I. (2021). Attention monitoring for synchronous distance learning. *Future Generation Computer Systems, 125*, 774–784.

Al Yousufi, A., Naidu, V. R., Jesrani, K., & Dattana, V. (2023). Tracking students' progress using big data analytics to enhance student's employability: A review. In *SHS web of conferences* (Vol. 156)EDP Sciences, Article 07001.

Alhothali, A., Albsisi, M., Assalahi, H., & Aldosemani, T. (2022). Predicting student outcomes in online courses using machine learning techniques: A review. *Sustainability, 14*(10), 6199.

Aljabarti, E. A., & Othman, M. S. (2023). *Classification students' attention in E-learning using machine learning*. Available at SSRN 4320161.

Anderson, T., & Dron, J. (2011). Three generations of distance education pedagogy. *International Review of Research in Open and Distance Learning, 12*(3), 80–97.

Anh, B. N., Son, N. T., Lam, P. T., Chi, L. P., Tuan, N. H., Dat, N. C., Trung, N. H., Aftab, M. U., & Dinh, T. V. (2019). A computer-vision based application for student behavior monitoring in classroom. *Applied Sciences, 9*(22), 4729. https://doi.org/10.3390/app9224729

Arciniegas, D. F. T., Amaya, M., Carvajal, A. P., Rodriguez-Marin, P. A., Duque-Munoz, L., & Martinez-Vargas, J. D. (2021). Students' attention monitoring system in learning environments based on artificial intelligence. *IEEE Latin America Transactions, 20*(1), 126–132.

Bashir, A., Bashir, S., Rana, K., Lambert, P., & Vernallis, A. (2021). Post-COVID-19 adaptations; the shifts towards online learning, hybrid course delivery and the implications for biosciences courses in the higher education setting. *Frontiers in Education. Frontiers, 6*, Article 711619. https://doi.org/10.3389/feduc.2021.711619

Bedi, A. (2023). Keep learning: Student engagement in an online environment. *Online Learning, 27*(2).

Bradski, G. (2000). The openCV library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer, 25*(11), 120–123.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Das, S., Chakraborty, S., & Mitra, B. (2021). Quantifying students' involvement during virtual classrooms: A meeting wrapper for the teachers. In *Proceedings of the 12th Indian conference on human-computer interaction* (pp. 133–139).

Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist, 26*(3–4), 325–346.

Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems, 49*(1), 5–22.

Django. (2023). The web framework for perfectionists with deadlines. Retrieved from https://djangoproject.com. (Accessed 19 July 2023).

Durães, D., Carneiro, D., Jiménez, A., & Novais, P. (2018). Characterizing attentive behavior in intelligent environments. *Neurocomputing, 272*, 46–54.

Elias, C. (2020). *Analysis of speech parameters as indicators of engagement in conversation*. Trinity College: Doctoral dissertation.

Foucault, M. (1977). *Discipline and punish: The birth of the prison (trans. Sheridan, A.)*. Vintage Books.

Foucault, M., & Gordon, C. (1980). *Power/knowledge: Selected interviews and other writings 1972–1977*. Pantheon.

Guo, J., & Xu, T. (2022). *IT monitoring and management of learning quality of online courses for college students based on machine learning*. Mobile Information Systems, 2022.

Gupta, S., Kumar, P., & Tekchandani, R. K. (2023). Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. *Multimedia Tools and Applications, 82*(8), 11365–11394.

Hassed, C. (2016). Mindful learning: Why attention matters in education. *International Journal of School & Educational Psychology, 4*(1), 52–60.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111–127.

Hodges, C. B., Moore, S., Lockee, B. B., Trust, T., & Bond, M. A. (2020). *The difference between emergency remote teaching and online learning*.

Hossen, M. K., & Uddin, M. S. (2023). *Students attention detection dataset*. Mendeley Data. https://doi.org/10.17632/smzggbnkd2.1. Vol. 1.

Hu, Y. H. (2022). Effects of the COVID-19 pandemic on the online learning behaviors of university students in Taiwan. *Education and Information Technologies, 27*(1), 469–491. https://doi.org/10.1007/s10639-021-10677-y

Hwang, K. A., & Yang, C. H. (2009). Automated inattention and fatigue detection system in distance education for elementary school students. *Journal of Educational Technology & Society, 12*(2), 22–35.

Kainat, Ali, S., Iqbal, K. F., Avaz, Y., & Saiid, M. (2022). A review on different approaches for assessing student attentiveness in classroom using behavioural elements. In *2022 2nd international conference on artificial intelligence (ICAI)* (pp. 152–158). IEEE.

Kansal, A. K., Gautam, J., Chintalapudi, N., Jain, S., & Battineni, G. (2021). Google trend analysis and paradigm shift of online education platforms during the COVID-19 pandemic. *Infectious Disease Reports, 13*(2), 418–428.

Kavitha, S., Raghuraman, G., Kavithasri, A., Aishvarya, S., & Janani, B. (2023). *Learning behaviour analysis of online course learners using EEG and facial expression data*. Measurement: Sensors, Article 100669.

Khan, A., Egbue, O., Palkie, B., & Madden, J. (2017). Active learning: Engaging students to maximize learning in an online course. *Electronic Journal of e-Learning, 15*(2), pp107–115.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In *Computer vision–ECCV 2014: 13th European conference, Zurich, Switzerland, september 6-12, 2014, proceedings, Part V* (Vol. 13, pp. 740–755). Springer International Publishing.

Liu, T., Yang, B., Geng, Y., & Du, S. (2021). Research on face recognition and privacy in China—based on social cognition and cultural psychology. *Frontiers in Psychology, 12*, Article 809736.

Lu, F., Chen, X., Ma, X., Liu, Z., & Chen, Y. (2020). The exploration and practice of IT solutions for online classes in higher education during COVID-19 pandemic. In *2020 international symposium on educational technology (ISET)* (pp. 298–302). IEEE.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., & Grundmann, M. (2019). Mediapipe: A framework for perceiving and processing reality. In *Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR)* (Vol. 2019).

Lugrin, J. L., Latoschik, M. E., Habel, M., Roth, D., Seufert, C., & Grafe, S. (2016). Breaking bad behaviors: A new tool for learning classroom management using virtual reality. *Frontiers in ICT, 3*, 26.

Manca, F., & Meluzzi, F. (2020). *Strengthening online learning when school are closed: The role of families and teachers in supporting students during the COVID-19 crisis*. Organisation for Economic Co-operation and Development.

Martinovic, D., & Ralevich, V. (2007). Privacy issues in educational systems. *International Journal of Internet Technology and Secured Transactions, 1*(1–2), 132–150.

Mason, J. (2010). Attention and intention in learning about teaching through teaching. In *Learning through teaching mathematics: Development of teachers' knowledge and expertise in practice* (pp. 23–47). Dordrecht: Springer Netherlands.

May, M., & George, S. (2011). Using students' tracking data in e-learning: Are we always aware of security and privacy concerns?. In *2011 IEEE 3rd international conference on communication software and networks* (pp. 10–14). IEEE.

Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record, 115*(3), 1–47.

Mershad, K., & Said, B. (2022). DIAMOND: A tool for monitoring the participation of students in online lectures. *Education and Information Technologies, 27*(4), 4955–4985.

Mindoro, J. N., Pilueta, N. U., Austria, Y. D., Lacatan, L. L., & Dellosa, R. M. (2020). Capturing students' attention through visible behavior: A prediction utilizing YOLOv3 approach. In *2020 11th IEEE control and system graduate research colloquium (ICSGRC)* (pp. 328–333).

Moşteanu, N. R. (2021). Teaching and learning techniques for the online environment. How to maintain students' attention and achieve learning outcomes in a virtual environment using new technology. *International Journal of Innovative Research and Scientific Studies, 4*(4), 278–290.

Muthuprasad, T., Aiswarya, S., Aditya, K. S., & Jha, G. K. (2021). Students' perception and preference for online education in India during COVID-19 pandemic. *Social Sciences & Humanities Open, 3*(1), Article 100101.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In *Handbook of research on educational communications and technology* (pp. 125–143). Routledge.

Novita, V., Abdullah, F., Rosmala, D., Silvani, D., & Hidayati, A. N. (2022). Students' learning engagement in an online academic listening and speaking class. *J-SHMIC: Journal of English for Academic, 9*(1), 87–102.

Ong, S. G. T., & Quek, G. C. L. (2023). Enhancing teacher–student interactions and student online engagement in an online learning environment. *Learning Environments Research*, 1–27.

Ouyang, F., Wu, M., Zheng, L., Zhang, L., & Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education, 20*(1), 1–23.

Pabba, C., & Kumar, P. (2022). An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition. *Expert Systems, 39*(1), Article e12839.

Parambil, M. M. A., Ali, L., Alnajjar, F., & Gochoo, M. (2022). Smart classroom: A deep learning approach towards attention assessment through class behavior detection. In *2022 advances in science and engineering technology international conferences (ASET)* (pp. 1–6). IEEE.

Parkhi, O., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. In *BMVC 2015-proceedings of the British machine vision conference 2015*. British Machine Vision Association.

Parvathy, G. S., & Smitha, R. (2022). A review on students attention monitoring systems. *International Research Journal of Engineering and Technology, 9*(11), 688–691.

Peper, E., Wilson, V., Martin, M., Rosegard, E., & Harvey, R. (2021). Avoid Zoom fatigue, be present and learn. *NeuroRegulation, 8*(1), 47-47.

Pimenta, A., Carneiro, D., Neves, J., & Novais, P. (2016). A neural network to classify fatigue from human–computer interaction. *Neurocomputing, 172*, 413–426.

Pimenta, A., Gonçalves, S., Carneiro, D., Fde-Riverola, F., Neves, J., & Novais, P. (2015). Mental workload management as a tool in e-learning scenarios. In *2015 international conference on pervasive and embedded computing and communication systems (PECCS)* (pp. 25–32). IEEE.

Powers, D. M. (2020). *Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation*. arXiv preprint arXiv:2010.16061.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education, 93*(3), 223–231.

Rahul, R. K., Shanthakumar, S., Vykunth, P., & Sairamnath, K. (2020). Real-time attention span tracking in online education. In *2020 IEEE MIT undergraduate research technology conference (URTC)* (pp. 1–4). IEEE.

Reddy, A. R., Gandhi, K. A., Jaffer, S. T., & Srilekha, M. K. (2022). Student live behaviour monitoring during virtual class using artificial intelligence. In *Journal of physics: Conference series* (Vol. 2335)IOP Publishing, Article 012026. No. 1.

Riding, R. J., & Sadler-Smith, E. (1997). Cognitive style and learning strategies: Some implications for training design. *International Journal of Training and Development, 1*(3), 199–208.

Saha, D., Mukherjee, I., Roy, J., Sarkar, B., & Bhattacharjee, R. (2021). Attendance and attention monitoring-a perspective in digital pedagogy. In *Journal of physics: Conference series* (Vol. 1797)IOP Publishing, Article 012067, 1.

Sartika, N., Sukmana, Y., Eliviani, R., & Awaludin, R. (2022). Smart online class monitoring service design. In *2022 8th international conference on wireless and telematics (ICWT)* (pp. 1–7). IEEE.

Serengil, S. I., & Ozpinar, A. (2020). Lightface: A hybrid deep face recognition framework. In *2020 innovations in intelligent systems and applications conference (ASYU)* (pp. 1–5). IEEE.

Sharma, M. (2021). SOCMS: Smart online class monitoring system. *Journal of Statistics & Management Systems, 24*(2), 251–261. https://doi.org/10.1080/09720510.2020.1833466

Sharma, A. S., Amin, M. R., & Fuad, M. (2023). Augmenting online classes with an attention tracking tool may improve student engagement. In *International conference on human-computer interaction* (pp. 105–121). Cham: Springer Nature Switzerland.

Sharma, P., Joshi, S., Gautam, S., Maharjan, S., Khanal, S. R., Reis, M. C., … de Jesus Filipe, V. M. (2023). Student engagement detection using emotion analysis, eye tracking and head movement with machine learning. In *Technology and innovation in learning, teaching and education: Third international conference*. Lisbon, Portugal: TECH-EDU 2022.

Stewart, M. (2021). Understanding learning: Theories and critique. In *University teaching in focus* (pp. 3–28). Routledge.

Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1701–1708).

Useche, O., & El-Sheikh, E. (2016). An intelligent web-based system for measuring students' attention levels. In *Proceedings on the international conference on artificial intelligence* (p. 381). ICAI.

Villegas-Ch, W., García-Ortiz, J., Urbina-Camacho, I., & Mera-Navarrete, A. (2023). Proposal for a system for the identification of the concentration of students who attend online educational models. *Computers, 12*(4), 74.

Wang, R. (2022). Research on the digital monitoring system of online course teaching quality under the background of big data. In *2022 international conference on educational innovation and multimedia technology (EIMT 2022)* (pp. 734–741). Atlantis Press.

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7464–7475).

Wang, H., & Xia, Y. (2021). Research and design of an attention monitoring system based on head posture estimation. In *2021 6th international conference on intelligent computing and signal processing (ICSP)* (pp. 1456–1459). IEEE.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70.

Computers and Education: Artificial Intelligence 5 (2023) 100174

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

Graph convolutional network for automatic detection of teachers' nonverbal behavior

Shiyan Pang [a], Shuhui Lai [a], Anran Zhang [a], Yuqin Yang [a,*], Daner Sun [b]

[a] Faculty of Artificial Intelligence in Education, Central China Normal University, Hubei, China
[b] Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong

ARTICLE INFO

Keywords:
Nonverbal behaviors
Graph convolutional network
Object detection
Extraction of skeleton points
Teachers' professional development

ABSTRACT

Accurately discerning teachers' nonverbal behaviors is essential to facilitate their reflective practice and professional growth. Yet, conventional manual approaches to observe and assess these behaviors are exceedingly time-consuming and resource-intensive. Furthermore, limited research has examined the precise and automated detection of teachers' nonverbal behaviors in classroom videos. This problem is exacerbated by the absence of a universally accepted framework for categorizing such behaviors, as well as a dearth of standardized datasets and corresponding algorithms. To address these challenges, this study investigates methods for the automatic detection of teachers' nonverbal behaviors. First, we theoretically define five common categories encompassing teachers' nonverbal behaviors, furnishing a structured framework for their analysis. Subsequently, in our methodology, we focus on primary and secondary school teachers, utilizing classroom video recordings as our primary data source. This study meticulously gathers and annotates datasets comprising teachers' nonverbal behaviors and devises an innovative method grounded in graph convolutional networks for their automated detection. To evaluate the effectiveness of our approach, we compare it with other state-of-the-art methods, highlighting its advantages. In practice, our method holds significant potential for real-world applications within authentic teaching contexts, offering the capability to automatically detect teachers' behaviors. These automated insights can serve as a valuable tool to guide and refine teaching practices, fostering more productive teaching outcomes.

1. Introduction

Teachers' nonverbal behavior plays a pivotal role in the realm of education, influencing both teaching and learning outcomes (Simpson & Erickson, 1983; Wang et al., 2022; Won et al., 2014). The development of meaningful nonverbal communication skills is not only crucial for teachers' professional growth but also essential for fostering successful student learning (Chang et al., 2018; Özmen, 2010). Teachers' nonverbal behavior encompasses a spectrum of elements, including body movements, facial expressions, gestures, and various forms of silent or auxiliary language such as gaze, voice modulation, speed, and tone (Ambady & Rosenthal, 1992; Simpson & Erickson, 1983). Proficient nonverbal communication enhances teacher-student rapport, fosters engagement, and enriches the learning experience (e.g., Brey & Pauker, 2019; Brophy & Good, 1984; Gholamrezaee & Ghanizadeh, 2018; Wang et al., 2022; Won et al., 2014; Šerić, 2021). Therefore, helping teachers recognize, reflect upon, and refine their nonverbal behavior is a core objective in teacher education (Chang et al., 2018; Özmen, 2010). An effective means to achieve this is through the analysis of post-task teaching videos, which can be conveniently recorded in smart classrooms (Won et al., 2014; Wu et al., 2020).

However, teachers encounter several challenges when reviewing their teaching videos and analyzing their instructional behavior, including time constraints and a lack of expertise in detecting and analyzing nonverbal behavior (Authors, 2023; Won et al., 2014; Wu et al., 2020). Therefore, automating the detection and analysis of teachers' nonverbal behavior is vital to support their reflective practice and professional development. Previous studies have employed traditional machine learning techniques to automate this process (Raza et al., 2015; Won et al., 2014). However, these methods typically require manual feature extraction and selection, heavily relying on expert input. In recent years, some scholars have explored deep learning-based

* Corresponding author. Faculty of Artificial Intelligence in Education, Central China Normal University, No. 152 Luoyu Road, Wuhan, Hubei, 430079, China.
E-mail addresses: pangsy@ccnu.edu.cn (S. Pang), lsh1005@mails.ccnu.edu.cn (S. Lai), zhar@mails.ccnu.edu.cn (A. Zhang), yangyuqin@ccnu.edu.cn (Y. Yang), dsun@eduhk.hk (D. Sun).

https://doi.org/10.1016/j.caeai.2023.100174
Received 18 June 2023; Received in revised form 6 October 2023; Accepted 7 October 2023
Available online 2 November 2023
2666-920X/© 2023 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

approaches for automatic detection of teachers' nonverbal behavior (Chen, 2020; Meng, 2022; Wu et al., 2020; Zhao et al., 2021). These methods automatically extract features from data and iteratively refine models to enhance the accuracy and reliability of teachers' nonverbal behavior detection, showing promising results.

Nonetheless, current deep learning-based detection of teachers' nonverbal behavior faces several challenges: firstly, there is a lack of standardized categories, with varying scholars adopting different frameworks for analyzing nonverbal behavior; secondly, due to data privacy concerns and inconsistencies in categorization, publicly available datasets for teachers' nonverbal behavior are scarce, necessitating further research to improve detection accuracy and reliability; thirdly, the practical utilization of automatic detection results in teacher training and professional development remains underexplored. To address these issues, our study seeks to answer the following research questions:

(1) What categories of teachers' nonverbal behavior can be identified for automated detection?
(2) How can the accuracy and reliability of automatic detection of teachers' nonverbal behavior be effectively improved?

In the subsequent sections, we will review the existing research on teachers' nonverbal behavior and its automated detection and analysis. We will then present our categorizations of teachers' nonverbal behaviors and the methods employed for their automated detection. Subsequently, we will outline our experimental approach and discuss the findings, along with their implications.

2. Literature review

2.1. Teachers' nonverbal behavior

Research on teachers' nonverbal behavior has held a significant position within the field of education, as evidenced by several influential studies (e.g., Brey & Pauker, 2019; Gholamrezaee & Ghanizadeh, 2018; Won et al., 2014). Broadly speaking, teachers' nonverbal behavior encompasses their actions and gestures using their bodies to convey information. In a more specific context, it pertains to their use of nonverbal symbols, such as gestures, head movements, facial expressions, and spatial positioning, to communicate effectively. Teachers commonly employ both verbal elements, such as varying tones of voice, and nonverbal cues, including meaningful facial expressions, body movements, and gestures, to enhance the appeal, interest, and informativeness of their teaching (Keller et al., 2016; Simpson & Erickson, 1983). Richmond (2002) emphasizes the independent existence of nonverbal communication alongside verbal messages and underscores the pivotal role that nonverbal behaviors play in students' learning. It is important to note that nonverbal behaviors and verbal communication are interdependent, both playing crucial roles in facilitating successful student learning (Gholamrezaee & Ghanizadeh, 2018; Simpson & Erickson, 1983).

Meaningful nonverbal behaviors exhibited by teachers play a pivotal role in the learning experiences of both students and educators (Brey & Pauker, 2019; Won et al., 2014). They have been found to exert positive influences on students' motivation, interest, emotions, affective experiences, and overall learning outcomes (Wang et al., 2022; Yang et al., 2022). For instance, Brey and Pauker (2019) discovered that teachers' nonverbal behavior can significantly impact students' motivation to learn. Mazer (2013) explored the connections between teachers' nonverbal behaviors and students' interest and engagement in learning, revealing significant correlations and associations between these factors. Gholamrezaee and Ghanizadeh (2018) established that teachers' nonverbal behaviors predict students' learning outcomes positively. Wang et al., (2022) affirmed that teachers' meaningful nonverbal behaviors positively influence students' emotional experiences. Moreover, some prior research has also examined teachers' nonverbal behavior in the context of their professional development (e.g., Chang et al., 2018; Özmen, 2010). These studies collectively underscore the need for comprehensive investigations into teachers' nonverbal behaviors. However, it's noteworthy that the majority of studies tend to concentrate on how teachers' nonverbal behaviors impact students' learning, motivation, emotions, and engagement, while limited research addresses the role of nonverbal behaviors in teachers' own learning and professional development.

Research on teachers' nonverbal behavior has traditionally relied on methods such as questionnaires, classroom observations, and interviews (e.g., Özmen, 2010; Šerić, 2021). In parallel, several studies have employed video recordings to delve into teachers' nonverbal behaviors (e.g., Chang et al., 2018; Wang et al., 2022; Wang & Loewen, 2016). However, in these video-based studies, the analysis has predominantly been conducted manually, a process known for its labor-intensity and susceptibility to human-coder bias (Won et al., 2014). Recent advancements in artificial intelligence technologies have introduced automation into classrooms, facilitating the automatic detection and analysis of teachers' nonverbal behaviors (Won et al., 2014). For example, Won et al. (2014) examined teacher-student nonverbal behaviors using vision hardware and machine learning algorithms. Wu et al. (2020) investigated teachers' nonverbal behaviors through automated detection methods. Building upon this trajectory of research, the current study similarly explores teachers' nonverbal behaviors through automatic detection and analysis of video recordings.

2.2. Classification of teachers' nonverbal behavior

The classification of teachers' nonverbal behavior forms the foundational framework for the study. Table 1 provides an overview of the prevalent classification systems employed in defining teachers' nonverbal behavior.

The Student-Teacher (S-T) Analytical Method encompasses a comprehensive range of factors, including language usage and the teaching process. The Flanders Interaction Analysis System (FIAS) (Flanders, 1970) meticulously categorizes teacher-student language interactions in classroom settings into three primary groups: teacher language, student language, and silence (representing a lack of effective language activity). These categories are further divided into ten subcategories. Building upon the foundations of FIAS, the Information-Technology-Based Interaction Analysis System (ITIAS) (Gu & Wang, 2004) introduced further refinements. For instance, the

Table 1
Categories of teacher's behavior according to different classification standards.

Classification Basis	Category
S-T Analytical Method	Explanation, question, prompt, suggestion, praise, criticism, summary and demonstration
FIAS (Flanders, 1970)	Accept emotion, praise, accept opinions, ask questions, teach, command, criticize
ITIAS (Gu & Wang, 2004)	Accept emotion, praise, accept opinions, ask open questions, ask closed questions, teach, instruct, criticize, proactive questioning, piscuss with peers, teacher manipulation technology, student manipulation technology, and technology acting on students
Zhou (2003)	Illustrative action, symbolic action, instrumental action, explicit action, regulatory action, adaptive action, distance action
He (2009)	Comprehensive gesture, pictographic gesture, indicative gesture, evaluative gesture, affective gesture, emphatic gesture
Ding (2020)	Blackboard writing, teaching, instructions, patrol guidance, and learners' demonstration on the stage
Chen (2020)	Symbolic gestures, schematic gestures, evaluative gestures, and indicative gestures
Meng (2022)	Indicative gestures, descriptive gestures, and rhythmic gestures

original FIAS category of *asking questions* within teacher language was subdivided into *asking open questions* and *asking closed questions*. The student language category was expanded to include subcategories like *proactive questioning* and *discussion with peers*, with the addition of a technology-related category. Additionally, Zhou (2003) approached teachers' nonverbal behavior from the perspective of observer recording and identified seven distinct categories: adaptive, illustrative, symbolic, explicit, instrumental, distance, and regulatory actions. Expanding on this work, He (2009) conducted a specialized classification of finger gestures employed by Chinese teachers, delineating six categories: indicative, understanding, evaluative, affective, pictographic, and emphatic gestures. Studies focused on automatic classification of nonverbal behaviors in the classroom have also surfaced, although the specific categories of nonverbal behaviors have exhibited variations. For instance, Ding (2020) considered categories like teaching, blackboard writing, instructions, patrol guidance, and learners' stage demonstrations. Chen (2020) classified gestures into symbolic gestures, conscious gestures, evaluative gestures, and indicative gestures. Meng (2022) divided gestures into indicative gestures, descriptive gestures, and rhythmic gestures. In summary, the current classification of teachers' nonverbal behavior exhibits variability, contingent on diverse data sources and distinct research objectives.

2.3. Automatic detection and analysis of nonverbal behavior

The application of nonverbal behavior detection in the field of education has wide-ranging implications. Several studies have explored automatic detection methods for nonverbal behavior, employing various machine learning algorithms such as Support Vector Machine (SVM) (Zhang, 2016), Decision Tree (Rodríguez et al., 2015; Won et al., 2014), Random Forest (Aydin et al., 2016), Logistic Regression (Nguyen et al., 2014; Won et al., 2014), and Naive Bayes (Lechuga Redondo et al., 2022). These algorithms, including SVM, Decision Tree, Random Forest, Logistic Regression, and Naive Bayes, are well-established supervised machine learning techniques commonly utilized for classification and regression tasks. In these studies, these classifiers are primarily employed to categorize nonverbal behavior based on manually extracted features from data, such as images.

With the emergence of deep learning algorithms, the automatic detection of nonverbal behavior through deep learning approaches has become predominant. Deep learning methods for nonverbal behavior detection typically encompass Convolutional Neural Networks (Ding, 2020; Meng, 2022; Tu et al., 2018; Wei et al., 2019), Faster Region-based Convolutional Neural Networks (Faster R–CNN) (Bai et al., 2020; Liu et al., 2020; Pang et al., 2022, 2023), YOLO (You Only Look Once) (Yan et al., 2019), and 2D/3D Pose Estimation (Cao et al., 2017; Chen, 2020; Wu et al., 2020).

The detection of nonverbal behavior mainly focuses on two key subjects: students and teachers. For the examination of students' nonverbal behaviors, Wei et al. (2019), Bai et al. (2020), and Liu et al. (2020) have utilized images as data sources to recognize various classroom behaviors exhibited by students. Wei et al. (2019) adapted the pre-trained VGG16 deep network model from ImageNet for the detection of students' classroom behaviors, successfully identifying seven typical classroom behaviors. Bai et al. (2020), leveraging the Faster R–CNN network, improved the accuracy of detecting students' classroom behaviors by integrating multiple features and applying transfer learning. Liu et al. (2020) introduced an enhanced method for detecting the classroom behavior of multiple students, addressing challenges posed by diverse postures, dense goals, and high occlusion rates. Additionally, Zeng (2020) and Xie et al. (2022) employed video data for recognizing students' classroom behavior. Zeng (2020) combined object detection with behavior recognition to enhance the effectiveness of detecting students' classroom behavior. Xie et al. (2022) proposed a YOWO (You Only Watch One) network model for real-time detection of students' classroom behavior, observing spatial positions through detection boxes. Due to the complexity of students' behavior in classrooms, directly using images or videos as data sources for detection can be affected by background noise, necessitating extensive training datasets and posing challenges in model training. To address this, Xu et al. (2020) and He et al. (2020) introduced pose estimation technology to enhance the accuracy of students' classroom behavior detection. The pose estimation technology involves the use of a pre-trained pose estimation model to extract student skeleton points from images and subsequently train a classification network for identifying students' classroom behavior.

Research focused on teachers' nonverbal behaviors primarily utilizes classroom videos as data sources. Yan et al. (2019) evaluated teachers' performance by detecting their body movements, with the behaviors of teachers mainly categorized into pointing to the blackboard and standing in front of the podium. Ding (2020) classified teachers' nonverbal behaviors by analyzing teacher images from classroom videos using VGG-based convolutional neural networks. Wu et al. (2020) collected and constructed a dataset of teacher behaviors using online open classroom teaching videos and explored behavior recognition methods based on RGB video and skeleton points. Chen (2020) extracted contour features and skeleton points of teachers from smart classroom video data, employing an RNN algorithm to classify gestures into symbolic gestures, intentional gestures, evaluative gestures, and indicative gestures. Meng (2022) created the CCNU-Hand dataset, comprising eight types of static hand gestures for detecting teachers' descriptive gestures. This dataset enabled hand gesture detection in single-frame images. Subsequently, a training dataset based on NTU-RGB + D was constructed for continuous gestures, and multi-layer perceptron and CNN were employed to classify continuous teaching gestures.

In terms of implementation types, the aforementioned nonverbal behavior detection methods can be broadly categorized into two groups: (1) direct detection of teacher behavior from images using convolutional neural networks or object detection methods; and (2) initial extraction of teacher skeleton points from images, followed by behavior classification based on these skeleton points.

The first type typically demands extensive training data and presents training challenges, while the second type leverages skeleton point information as input, often overlooking the interplay between different skeleton points. This study aligns with the second category, utilizing object detection to first localize the teacher's area. Subsequently, the OpenPose algorithm is employed to extract teacher skeleton points. To achieve precise teacher behavior detection, an appropriate adjacency matrix is designed based on the extracted skeleton points and their relationships, forming the basis for the graph convolutional network.

3. Methods

3.1. Definition of teachers' nonverbal behavior

This study relied on an extensive collection of teaching videos to scrutinize teachers' nonverbal behavior, concentrating solely on those behaviors that were discernible through visual observation. In the process of categorizing teachers' nonverbal behavior, we deliberately excluded categories associated with vocal expressions or spatial proximity. Drawing from pertinent literature on teacher behavior and teaching analysis, we delineated five distinct types of teacher nonverbal behavior from a visual standpoint, as detailed in Table 2: symbolic action, conscious action, indicative action, evaluative action, and adaptive action.

3.2. Detection of teachers' nonverbal behavior

Fig. 1 shows a flowchart of the automatic detection of teachers' nonverbal behavior in this study. First, the faster R–CNN object detection algorithm (Ren et al., 2015) was used to locate each teacher's position and obtain the teacher's image. Second, the human pose

Table 2
Definitions of teachers' nonverbal behaviors.

Class	Meaning
Symbolic action	Teachers use gestures to indicate figures or numbers
Conscious action	Teachers motion students to stand, sit, raise their hands or keep quiet
Indicative action	The teacher points to the blackboard or screen
Evaluative action	The teacher claps or gives a thumbs up
Adaptive action	Body movements or postures teachers gradually develop over time to adapt to certain psychological, physical or objective environmental needs, e.g., hands on your back, hands together, etc.

estimation algorithm was applied to extract the coordinates of the teacher's skeleton points from the image. Third, based on these skeleton points, a graph convolutional network was built to classify the teacher's behaviors.

3.2.1. Extraction of teachers' image

As the extraction of the teacher's skeleton points and classification of the teacher's nonverbal behavior in this study were based on the extracted teacher's image, the accuracy of this first step greatly influenced the subsequent experiments. Extracting the teacher's image was modeled as an object detection problem. The faster R–CNN algorithm was selected for object detection because it offers several advantages. First, its detection is accurate and performs well with multi-scale and small objects and crowded targets, making it suitable for the complexity of the classroom setting. Second, it performs well with multiple object detection datasets and is efficient for transfer learning, as it uses a model pre-trained on a large-scale dataset to initialize the model parameters so that only the target classification network is modified according to the number of targets, thereby reducing dataset usage and the difficulty of model training. Therefore, to locate the teacher's positions, the faster R–CNN algorithm was selected, and a pre-trained model on the ImageNet dataset was used for transfer learning to improve the accuracy of the teacher image extraction.

The faster R–CNN network has five parts: the feature extraction module, region proposal network (RPN), region of interest (RoI) pooling, classification module and regression module, as shown in Fig. 2. The feature extraction module uses a set of conv + ReLU + pooling layers to extract the feature maps of the image for the subsequent RPN layer. In this study, ResNet-50 was chosen as the feature extraction network, and the output was the multi-scale feature maps of ResNet-50 at each stage. RPN mainly uses the anchor mechanism to generate candidate boxes, replacing the traditional candidate box generation algorithm to improve the speed of object detection. The RoI pooling layer is responsible for collecting the region proposals. The classification module classifies all region proposals into predefined categories through the full connection layer and softmax function. The regression module performs a bounding-box regression on the region proposals to obtain the final predicted box with higher accuracy.

As the aim of this step was to extract the teacher's image, a binary classification task was constructed for the two datasets to distinguish only the teacher and the background without subdividing the teacher's nonverbal behavior. In this step, extracting the teacher's image from the original image was a relatively simple detection task. Thus, based on the pre-trained model on the ImageNet dataset, we constructed a small number of teacher target detection datasets to train the model to perform the binary classification task, which produced relatively perfect teacher's images.

3.2.2. Extraction of the teacher's skeleton points

After the teacher's target images were acquired, the nonverbal behavior had to be identified from the images. As the differences between the characteristics of some teachers' nonverbal behavior in images can often be small, the strategy in this study was to extract the teacher's skeleton points and use this information to more accurately detect and classify their nonverbal behavior. Human skeleton point information is composed of joint points that form an abstract representation of human joint connections. This can eliminate much redundant information unrelated to behavior, such as background, thus reducing the complexity of behavior detection. At present, with the extensive application of deep learning, the pose estimation algorithm based on the convolutional neural network can reliably and robustly extract human skeleton information. Thus, in this study, the OpenPose algorithm (Cao et al., 2017) was used to extract the teacher's skeleton points.

OpenPose is a human skeleton point extraction algorithm based on the convolutional neural network. This algorithm is currently one of the best and fastest solutions for obtaining skeleton points. It has three parts: (1) part confidence map generation, (2) skeleton point affinity field calculation and (3) skeleton point matching, as shown in Fig. 3. The part confidence map generation module first uses VGG, ResNet and other networks to extract high-order features from the image, then these high-order features are used to obtain the part confidence map. Part confidence maps are used to represent the position of the skeleton points, which are not isolated points on the image but circular patterns created by Gaussian kernels that occupy a certain area. The skeleton point affinity field is used to describe the affinity between the different skeleton points, and is mainly calculated based on the angle and direction between skeleton points. Points belonging to the same object show large affinity, whereas points belonging to different targets show small affinity. The skeleton point matching module mainly classifies the skeleton points of the same object into one class according to the affinity, and the Hungarian algorithm was used for skeleton point matching in this paper.

The structure of the human body can be formed from a combination of skeleton points. Commonly used skeleton models include an 18-point model that follows the COCO format standard and a 25-point model that follows the Body_25 format standard, as shown in Fig. 4.

The 18-point model is sequentially represented by the numbers 0 to 17, corresponding, respectively, to the nose, neck, right shoulder, right elbow, right wrist, left shoulder, left elbow, left wrist, right hip, right knee, right ankle, left hip, left knee, left ankle, right eye, left eye, right ear, and left ear. The 25-point model adds the middle hip, left big toe, left little toe, left heel, right big toe, right little toe and right heel.

In this study, because the application scenario was the offline classroom and the teacher's feet were not visible in the image, the 18-point model was used. Each skeleton point in the model had its own label. Each extracted skeleton point contained x, y values and a confidence score. In addition, as the size of the teacher's image obtained from object detection was not fixed, after obtaining the skeleton points, regularization was performed to convert the coordinates from the coordinate position of the entire image to the relative position coordinate. The formula was

$$P[i]^{'} = P[i]\frac{1}{\|P[1] - P[i]\|^2}, \tag{1}$$

where $P[i]$ represents the coordinates of the i points in the original image, where $i \in [2, 3, \ldots, 18]$, $P[i]^{'}$ represents the coordinates of the i points after regularization of the original coordinates and $P[1]$ is the position of the first point in Fig. 4 (left).

3.2.3. Detection of teachers' nonverbal behavior

After the teacher's skeleton points were extracted from each image, the nonverbal behavior was classified according to the 18 sets of coordinates in the skeleton point model. As the inputs from subsequent models were discrete coordinates and were adjacent to the points, the mainstream graph convolutional network (GCN) was used to build a nonverbal behavior detection model for a specific teacher.

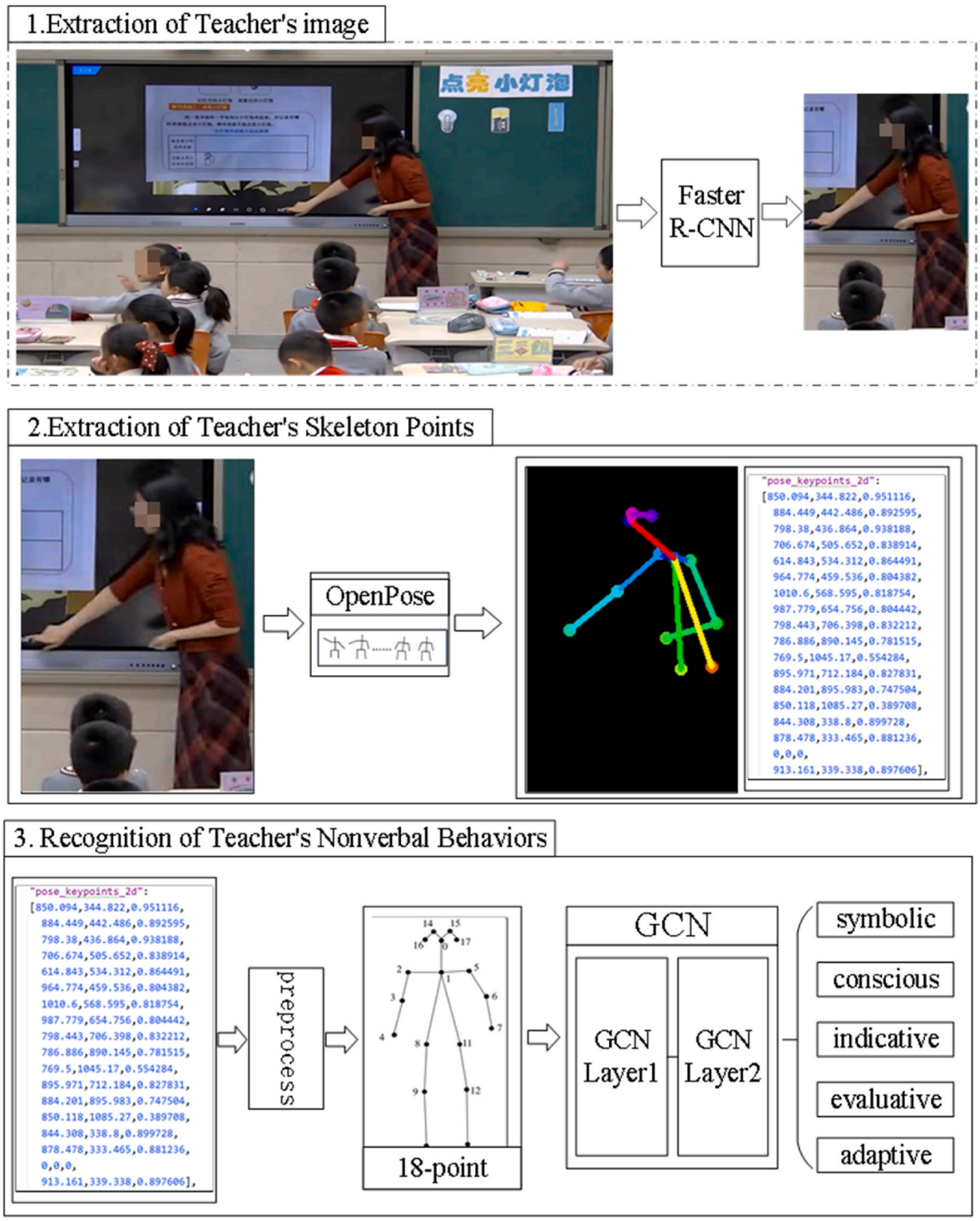


Fig. 1. Flowchart of teachers' nonverbal behavior detection.

3.2.3.1. Use of the GCN to detect nonverbal behavior. The GCN generalizes the convolution operation from Euclidean space data (such as images) to non-Euclidean space data with a graphic structure. The core idea of using the GCN was to obtain a function map through the back-propagation algorithm and then, through the relationships between the nodes in the map, aggregate its features and the neighboring features to represent the newly generated nodes.

Any graph convolutional layer can be represented by a nonlinear function, and the difference between different models lies in the implementation of the function f. The formula for the graph convolutional layer is

$$H^{l+1} = f\left(H^l, A\right), \tag{2}$$

where $H^0 = x, x \in R^{N\times M}$ is the input of the first layer, l is the current layer number, N is the number of nodes, M is the dimension of the feature vector of the node and A is the adjacency matrix.

The formula for the graph convolution in this study is

$$H^{l+1} = \sigma\left(L^{sym} H^l W^l\right), \tag{3}$$

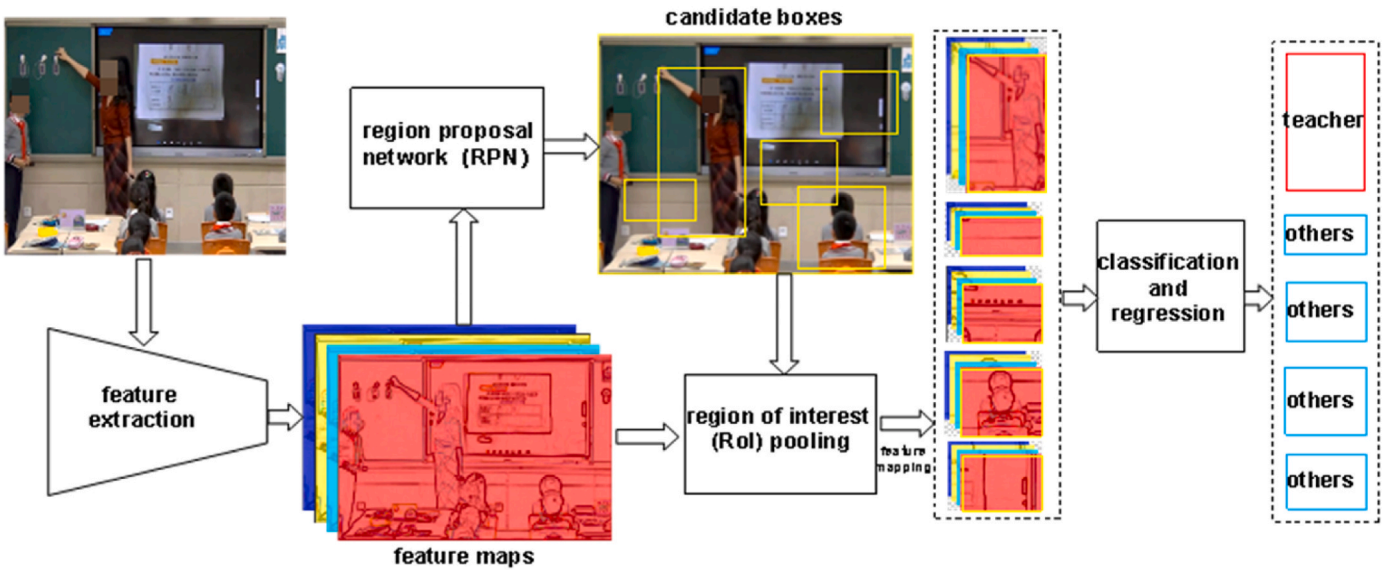


Fig. 2. Flowchart of our extraction of teachers' image with faster R–CNN.

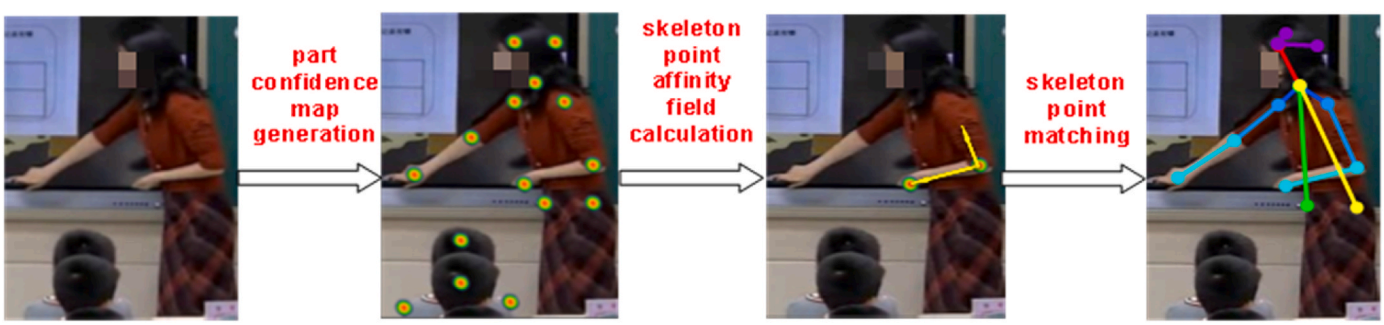


Fig. 3. Flowchart of our extraction of the teacher's skeleton points with OpenPose.

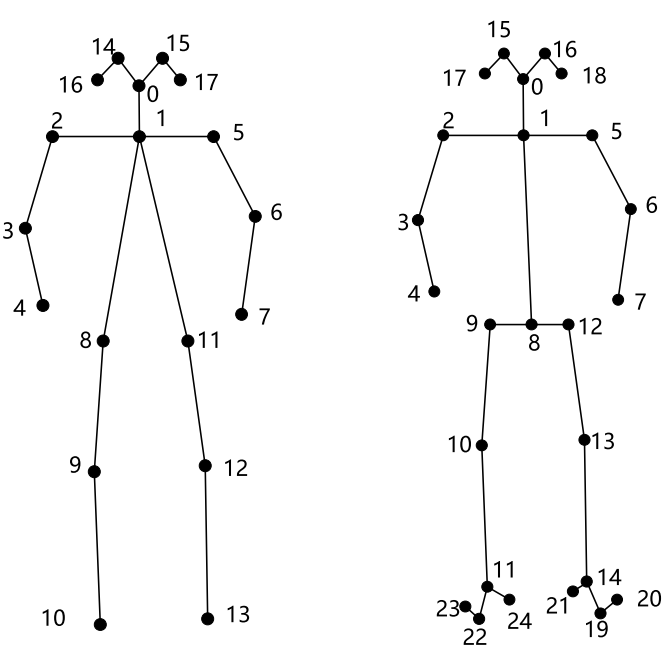


Fig. 4. 18-point skeleton model (left) and 25-point skeleton model (right).

where $\sigma(\cdot)$ is the nonlinear activation layer, W^l is the weight matrix of the lth layer, L^{sym} is the symmetric normalized Laplacian matrix $L^{sym} = D^{-1/2}\widehat{A}D^{-1/2}$, where D is the degree matrix, and $\widehat{A} = D - A$. Thus, the above formula can be written as

$$H^{l+1} = \sigma\left(D^{-1/2}\widehat{A}D^{-1/2}H^l W^l\right), \tag{4}$$

3.2.3.2. Nonverbal behavior detection with GCN. According to the theory of GCN, the input of the teachers' nonverbal behavior detection model based on GCN needs to include the coordinates and the adjacency matrix of the skeleton points. The skeleton point coordinates were obtained using the skeleton point extraction algorithm. The improvement in the GCN algorithm in this study mainly lay in the design of the adjacency matrix of the skeleton points.

In the teachers' nonverbal behavior detection task, the human body can be regarded as an articulated system composed of articulated joints and rigid bones. Its inherent structure is based on the graph shown in Fig. 4. According to Fig. 4 (left), the edge information of the human skeleton graph is as follows: E = [(4, 3), (3, 2), (7, 6),(6, 5), (13, 12), (12, 11), (10, 9), (9, 8), (11, 1), (8, 1), (5, 1), (2, 1), (0, 1), (15, 0), (14, 0), (17, 15), (16, 14)].

According to the edge information, the adjacency matrix can be obtained as follows:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

A skeleton model based on a graph $G(x, A)$ is constructed, where $x \in R^{N\times 3}$ represents the two-dimensional coordinates containing N skeleton points and A is the adjacency matrix of the skeleton points with $N\times N$ dimensions. Here, we set the diagonal elements of the adjacency matrix A $a_{ii}= 0$ to discard the self-connection of each skeleton point.

However, in the process of behavior detection based on skeleton points, some skeleton points are not directly connected in the above adjacency matrix A, but show certain connections in different nonverbal behaviors, which is also an important factor in behavior detection. For example, the left and right hands are physically disconnected from each other, but their relationship is of great significance in recognizing the clapping action. Even in the clapping action, the logical connection between the left and right hands changes much more than the physical connection, as shown by the dotted line in Fig. 5.

Therefore, the relationships between skeleton points were divided into physical connections and logical connections, and an adjacent matrix A with different weights was constructed. The physical connection refers to the hard connection between human joints, which is mainly determined by the skeleton point extraction algorithm and has a strong connection; that is, the weight of the edge information does not change under any action, as shown by the solid line in Fig. 5. Logical connection was defined in this study as a connection that is important for teachers' nonverbal behavior detection but was not a hard connection. As a supplement to physical connection, logical connection is shown by the dotted line in Fig. 5. For each element a_{ij} in the adjacency matrix A, the weights of the physically connected edges are denoted by α and the weights of the logically connected edges are denoted by β, as shown in the following formula.

$$a_{ij} = \begin{cases} 0, & i = j \\ \alpha, & \text{physical connection}, \\ \beta, & \text{logical connection} \end{cases} \tag{5}$$

After the adjacency matrix A was determined, the input skeleton point coordinates $x \in R^{N\times 3}$ were converted into the spatial structure features through the GCN using Equation (4). Finally, a softmax classifier was used to classify the spatial structure features obtained by the graph convolution (e.g. H^{l+1}) into one of the five categories: symbolic action, conscious action, indicative action, evaluative action and adaptive action.

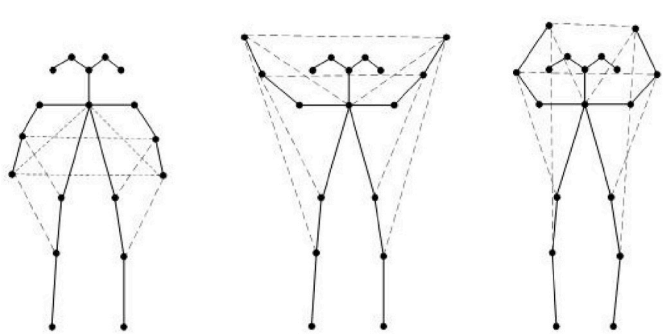


Fig. 5. Schematic diagram of physical connection and logical connection.

4. Experiments

4.1. Experimental datasets

Currently, while deep learning methods generally outperform traditional algorithms across various domains, they necessitate a substantial amount of training data to develop an optimal model. Consequently, for the detection of teachers' nonverbal behavior, the availability of a suitable training dataset is imperative. Despite the efforts of many scholars and research institutions in publishing behavior detection datasets, such as the Kinetics Human Action Video Dataset (Kay et al., 2017), the UCF series database (Soomro et al., 2012), the HMDB behavior database (Kuehne et al., 2011), and the NTU RGB + D dataset (Shahroudy et al., 2016), a noticeable gap remains in the realm of open training datasets specifically tailored for teachers' behavior detection.

To address this gap and align with our research objectives, we identified five distinct categories of teachers' nonverbal behavior from relevant literature. Subsequently, we proceeded to construct two separate datasets for teachers' nonverbal behavior detection. The first dataset was designed to simulate teachers' nonverbal behavior within a classroom setting. The second dataset was drawn from a selection of ministerial-level "excellent lessons" and classroom videos featuring science teachers, sourced from the National Public Service Platform of Education Resources. These two datasets were utilized to train and validate the proposed algorithm, encompassing both simulated and real-life scenarios. Below, we provide a detailed description of these two datasets.

Simulated Classroom Dataset. This dataset was recorded according to five standard nonverbal behaviors of teachers. After the recording was completed, the video of each action was cut every 30 frames (1 s) to produce 9410 images. Samples from the dataset are shown in Fig. 6.

In this dataset, the training and test sets were divided at a 7:3 ratio, and the number of images in each action category is shown in Table 3.

Real Classroom Dataset. To address the lack of teachers' behavior detection datasets in real scenarios, this study selected the classroom videos of science teachers and ministerial "master classrooms" from the National Public Service Platform of Education Resources to collect nonverbal video data of teachers in science classes. However, most of the real classroom videos collected consisted of a mixture of teacher, student and multimedia classroom images. We therefore used video processing tools to edit and remove all video clips that excluded the teacher, and we obtained 1806 images. Fig. 7 shows samples of the dataset.

In this dataset, the training and test sets are also divided according to the 7:3 ratio, and the number of images in each action category is shown in Table 4.

4.2. Experimental details

Parameters of the Experimental Environment. Given the high arithmetic power demanded by the deep learning algorithm in this study, a high-performance environment of Graphic Processing Unit was selected for model training. The parameters of the experimental environment are shown in Table 5.

Dataset Annotation. The annotation tool for the datasets used in this study was LabelImage on the Windows platform. We labeled 11,216 classroom images from the two datasets. Taking the annotation of one image as example, as shown in Fig. 8, there is only one teacher taken as the target in each image and marked with a rectangular box. The

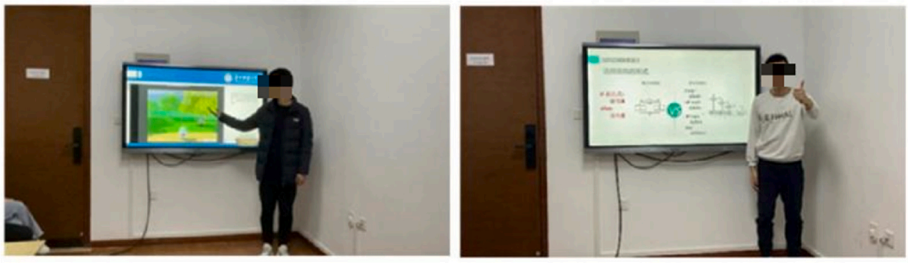


Fig. 6. Samples of simulated classroom dataset.

Table 3
Details of simulated classroom dataset.

Category	Total/Train/Test
Symbolic Action	1055/739/316
Schematic Action	3632/2543/1089
Indicative Action	1164/815/349
Evaluative Action	2044/1431/613
Adaptive Action	1515/1061/454

students, tables and chairs, blackboards and whiteboards, etc. in the classroom are regarded as background. After annotation, using the LabelImage annotation tool, a file in XML format was generated for subsequent automatic processing.

Dataset augmentation. Small sample datasets are prone to overfitting, and generalization ability of the resulting model is poor. In this study, the dataset was enhanced by image scaling, random rotation, random adjustment of color, contrast, sharpness, to reduce the overfitting problem caused by less data. At the same time, considering the common noise in the video acquisition process, Gaussian noise and salt and pepper noise are simulated and superimposed to the dataset to enhance the robustness of the proposed algorithm, as shown in Fig. 9.

Experimental Groups. The experiments were divided into two groups: machine learning and deep learning. For the machine learning group, the K-Nearest Neighbors, decision tree, random forest and support vector machine algorithms were selected. For the deep learning group, multilayer perception and GCN were selected. Each group of experiments was performed on two datasets as shown in Table 6.

Training Details. The machine learning method used in this study was implemented based on the Scikit-learn library and evaluated using a fivefold cross-validation method with a data input of 18 sets of point coordinates (i.e., 36 items of data). The multilayer perception method used three hidden layer structures, and each node of the hidden layer was activated using the ReLU function after computation. Adam was selected as the optimizer algorithm with a 0.01 learning rate, and the training comprised 100 epochs. A two-layer GCN method was used in this study.

4.3. Experimental results and analysis

The experimental comparison results of the different methods are shown in Table 7.

Table 7 shows that the accuracy of the machine learning algorithm is lower than that of the deep learning method. It can be seen that with a

Table 4
Details of real classroom dataset.

Category	Total/Train/Test
Symbolic Action	205/144/61
Schematic Action	225/158/67
Indicative Action	470/329/141
Evaluative Action	192/135/57
Adaptive Action	714/500/214

Table 5
Parameters of the experimental environment.

CPU	Intel CoreI9-10900K CPU@3.70GHZ
Memory	64 GB
GPU	NVIDIA RTX2080Ti 11 GB
Operating System	Windows 10
Programming Language	Python 3.6.0
Deep Learning Framework	Pytorch 1.6.0

Fig. 7. Samples of real classroom dataset.

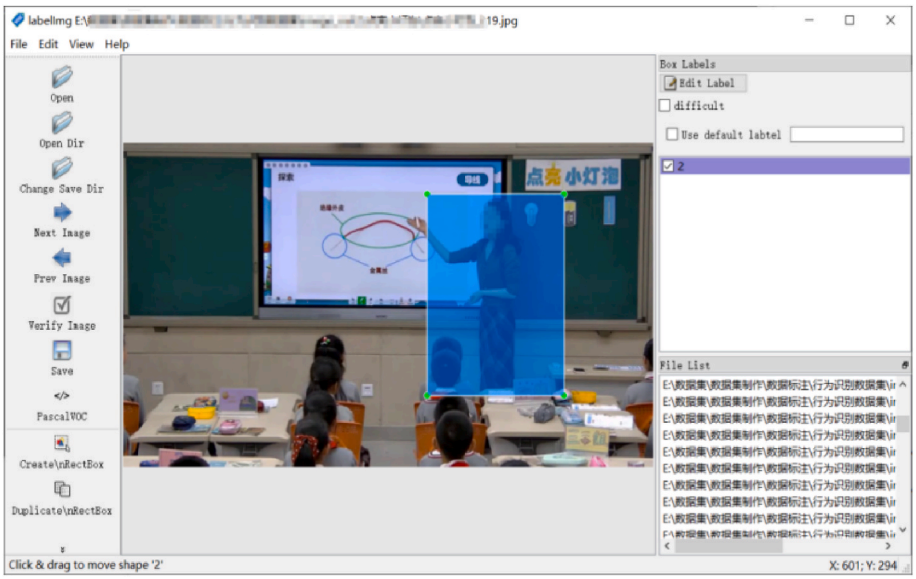


Fig. 8. Example of annotated dataset.

Fig. 9. Examples of an image superimposed with Gaussian noise (left) and salt and pepper noise (right).

Table 6
Grouping of experiments.

Method	Model	Input
Machine Learning	K-Nearest Neighbor decision tree random forest support vector machine	Coordinates of skeleton points and corresponding ground truth (e.g., labels of teachers' nonverbal behavior)
Deep Learning	multi-layer perception GCN	

Table 7
Comparison of experimental results of different methods.

Method	Model	Accuracy of simulated dataset(%)	Accuracy of real dataset(%)
Machine Learning	K-Nearest Neighbor	72.17	67.96
	decision tree	76.95	65.72
	random forest	81.94	73.15
	support vector machine	82.35	72.64
Deep Learning	multi-layer perception	90.65	78.83
	GCN	94.68	81.64

certain amount of data, the deep learning methods perform better than the machine learning methods. The accuracy rates on the simulated classroom dataset and the real classroom dataset are 94.68% and 81.64%, respectively.

For a more in-depth analysis of the experimental outcomes, we extracted accuracy statistics for the proposed GCN method pertaining to the five nonverbal behavior categories, which were presented in Table 8. Notably, indicative action exhibited the highest detection accuracy, whereas conscious action proved to be the most challenging to identify. This discrepancy can be primarily attributed to the fact that each category comprises multiple behaviors, as outlined in Table 2. Consequently, when there is a greater divergence among behaviors within the same category, the task of identification becomes inherently more intricate.

Table 8
Statistical results of accuracy of different categories.

Classes	Accuracy of simulated dataset (%)	Accuracy of real dataset (%)
Symbolic action	89.18	78.45
Conscious action	89.31	74.16
Indicative action	96.65	81.12
Evaluative action	94.02	74.46
Adaptive action	92.09	78.65

5. Discussion

5.1. Comparison of different methods

Skeleton point extraction. In this study, we initiated the process by using the Faster R–CNN object detection algorithm to pinpoint the position of each teacher and extract their images. Subsequently, we utilized OpenPose, a human posture estimation algorithm, to extract skeleton points from these teacher images. In contrast to traditional machine learning techniques, the method proposed herein offers several advantages: (1) it does not rely on manual feature extraction, (2) it exhibits lower time and space complexity, and (3) it does not compromise the subsequent behavior detection performance. Furthermore, in comparison to the YOLO algorithm (Yan et al., 2019) and the VGG-16 convolutional neural network algorithm (Ding, 2020; Meng, 2022), both of which employ images as direct inputs, our proposed method effectively mitigates background interference in complex classroom settings and enhances the reliability of skeleton point extraction.

Teachers' nonverbal behavior detection. In this study, we employed GCN for teachers' nonverbal behavior detection and differentiated between physical and logical connections to optimize the adjacency matrix. When an ample dataset is available, deep learning methods such as GCN outperform traditional machine learning techniques, including K-Nearest Neighbors, decision trees, random forests, and support vector machines, in detecting teachers' nonverbal behavior. Deep learning methods also offer the advantage of convenience and speed by eliminating the need for manual feature design. Our results indicate that GCN achieved higher accuracy than the multilayer perceptron deep learning method. This discrepancy arises because the multilayer perceptron method cannot process data in non-Euclidean space, with its network input being simple coordinates that overlook the spatial connectivity between different points. Conversely, GCN can simultaneously process coordinates and spatial connections among skeleton points, thereby more effectively representing the action characteristics of the human skeletal structure.

5.2. Accuracy analysis of different types of teachers' nonverbal behavior

As depicted in Table 8, indicative and evaluative actions exhibited the highest detection accuracy among the five nonverbal action categories, while symbolic and conscious actions proved to be the most challenging to identify. This result can be attributed to the inherent characteristics of these five nonverbal behaviors. According to the definitions provided in Table 2, indicative actions typically involve the teacher pointing at the blackboard with one arm, making this action more distinct from the other categories and thus easier to identify. Evaluative actions attained the second-highest accuracy due to their inclusion of double-gesture actions such as clapping and arm crossing, in contrast to the other four actions that primarily consist of single-gesture actions. However, symbolic and conscious actions exhibited lower detection accuracy and were susceptible to misidentification. Notably, there are certain similarities in the definitions of symbolic and conscious actions. For instance, the gestures associated with symbolic actions, used to indicate numbers like 1, 2, 3, etc., bear some resemblance to the conscious actions instructing students to raise their hands or stand.

5.3. Implications

5.3.1. Theoretical implications

This study represents an amalgamation of various theoretical approaches to accurately detect and analyze teachers' nonverbal behavior. In the realm of applying artificial intelligence to education, relying solely on a single artificial intelligence theory may not yield practical results. Instead, we integrated multiple theories, taking into account their diverse applicability, to identify the most effective algorithm for discerning teachers' nonverbal behavior. This approach serves as an illustrative example of the integrated utilization of various artificial intelligence theories to accurately detect teachers' nonverbal behaviors. Furthermore, we incorporated skeleton point adjacencies into the graph convolutional network and optimized the adjacency matrix through physical and logical connections, effectively distinguishing different nonverbal behaviors among teachers. Additionally, this study establishes five common categories of teachers' nonverbal behavior: symbolic action, schematic action, indicative action, evaluative action, and adaptive action. These categories provide a structured framework for analyzing teachers' nonverbal behaviors in future research endeavors.

5.3.2. Practical implications

Artificial intelligence technologies are gradually finding their place in smart classrooms, ushering in a new era of education. However, there remains a scarcity of studies focusing on the application of artificial intelligence in analyzing teacher behavior, particularly teachers' nonverbal behavior, which permeates the entire teaching process. In this study, we harnessed artificial intelligence technologies, including object detection networks and graph convolutional networks, to detect teachers' nonverbal behaviors. These identified nonverbal behaviors can aid teachers in conducting a detailed analysis of their teaching processes and enhancing their awareness of using meaningful nonverbal cues for effective teaching. Our experimental results underscore a significant enhancement in the accuracy of detecting teachers' nonverbal behaviors. As many schools have embraced smart classrooms equipped with multimedia tools and comprehensive teaching resources, our methods can be readily applied in authentic teaching settings for automatic detection of teachers' behavior. This empowers teachers to analyze and regulate their teaching behaviors, fostering a conducive environment for effective teaching based on automatic analysis results.

5.3.3. Methodological implications

Traditional participatory observation is a widely utilized method for detecting and analyzing teachers' nonverbal behaviors. This approach involves experts entering the classroom to observe, evaluate, and guide teachers' nonverbal behaviors, enabling teachers to improve their teaching based on expert feedback. However, this method presents a short-term limitation as experts cannot provide continuous feedback, and ongoing self-reflection requires teachers to consistently acquire classroom teaching information. Yet, teachers often grapple with heavy teaching loads, time constraints, and limited energy, making it challenging to thoroughly review and analyze teaching videos. In response, this study introduces an artificial intelligence method for the automatic detection and analysis of teachers' nonverbal behaviors using video recordings. Compared to traditional machine learning methods (e.g., SVM, Decision Tree, Random Forest, Logistic Regression, and Naive Bayes), our approach eliminates the need for manual feature extraction to classify teachers' nonverbal behavior, rendering it convenient and efficient. In contrast to deep learning methods that directly analyze teachers' behavior from images (e.g., Faster R–CNN, YOLO, and convolutional neural networks), our method integrates object detection and skeleton point extraction techniques, effectively eliminating background interference. Moreover, we enhance the establishment of adjacencies between different skeleton points and optimize the adjacency matrix through physical and logical connections, simplifying the distinction between these categories. Extensive experiments underscore that artificial intelligence technology can compensate for the shortcomings of traditional manual methods in terms of efficiency and consistency, making more effective use of resources like smart classrooms, reducing time and labor costs, and providing valuable support for teachers' self-reflection and improvement in teaching.

6. Conclusion

In recent years, there have been rapid advancements in computer vision and machine learning technologies. However, despite these

advances, accurately and automatically identifying teachers' nonverbal behavior remains a challenge due to the complex filming conditions and the substantial variations in nonverbal behaviors within the same category exhibited by different teachers. Consequently, this study introduces an approach for the automatic detection of teachers' nonverbal behavior, leveraging Graph Convolutional Networks (GCN) with an extensive dataset of teaching videos. The method used in this study involves several key steps. Firstly, the Faster R–CNN algorithm is employed to ascertain the precise location of the teacher within the frame and extract their image. Subsequently, the OpenPose algorithm is utilized to analyze the teacher's image, extracting the coordinates of their skeleton points. Finally, leveraging these skeleton point coordinates and the physical as well as logical connections between them, a GCN model is constructed to detect and classify teachers' nonverbal behavior.

To evaluate the efficacy of the proposed method, two distinct datasets were meticulously collected and curated, following predefined criteria for five categories of nonverbal behaviors. One dataset was created to simulate classroom settings, while the other was derived from real classroom scenarios, drawing from publicly available educational resources. A comparative analysis was conducted between the traditional machine learning approach and the multilayer perception method based on these two datasets. The experimental findings demonstrate that video image data from classrooms can effectively identify teachers' nonverbal behavior. Moreover, the deep learning GCN method significantly outperforms traditional machine learning techniques in terms of detection accuracy. These results underscore the potential of leveraging advanced machine learning technologies, like GCN, to enhance the automatic identification of teachers' nonverbal behavior, thereby offering valuable insights into the realm of education.

6.1. Limitations and future research

This study has achieved some progress in the automatic recognition of teachers' nonverbal behavior, highlighting the immense potential of artificial intelligence technology within the domain of teachers' nonverbal behavior detection. Nonetheless, given the time constraints and the scope of our capabilities, this study does have several limitations, encompassing the following three primary aspects.

First, the nonverbal behavior dataset of teachers in this study comprised solely of teaching images, limiting the detection and analysis to visually distinguishable nonverbal behaviors. This approach excluded other essential categories, such as language and body distance information, which precludes a comprehensive analysis of teachers' nonverbal behavior.

Second, the extraction of teacher's skeleton points was conducted in two stages: teacher detection and skeleton point extraction. The accuracy of teacher detection significantly impacts subsequent steps and may result in false detections, as illustrated in Fig. 10. In the process of identifying the target teacher, instances of misidentification, where a student may be erroneously identified as the teacher, can lead to incorrect skeleton point extraction and consequently, inaccurate teacher behavior detection.

Third, due to the limited sample size and the specific context of this study, the model employed for teachers' nonverbal behavior detection may lack generalizability. One of the datasets used for training the behavior detection model was generated from the research team's simulated classroom videos of teachers' nonverbal behavior. These videos sampled a single frame every 30 frames (at 1-s intervals), potentially oversimplifying the actual complexity of real teachers' nonverbal behaviors. Variations in smart classroom camera angles, teacher standing habits, and individual characteristics can result in substantial differences in body shape and movement angles for each behavior. Additionally, the applicability of the teachers' nonverbal behavior detection model across diverse scenarios, the optimization of model architecture, the expansion of sample size, and the enhancement of model robustness in various settings need further consideration.

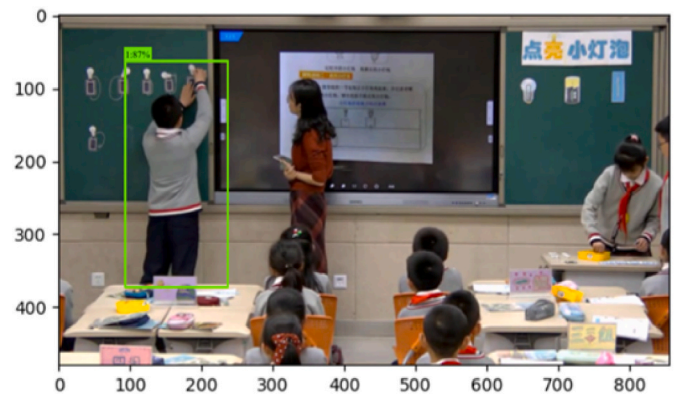


Fig. 10. An example of teacher's false detection.

Furthermore, the method proposed in this study is geared towards classifying teachers' nonverbal behavior from single-frame images. However, nonverbal behaviors often occur sequentially, with a series of consecutive image frames providing a more accurate depiction of a teacher's behavior category. Future research endeavors will involve modeling sequential images to enhance the precision of teacher's nonverbal behavior detection.

Statements on open data and ethics

The Simulated Classroom Dataset is not available due to the nature of the consent form at the time.

The video data of the Real Classroom Dataset comes from the National Public Service Platform of Educational Resources, and it is open for all.

The overall project research had research ethics approval from the Human Research Ethics Committee of Central China Normal University.

Author contributions

All the authors made substantial contributions to this paper. The first author designed the study, collected and analyzed data, and wrote the paper. The second author analyzed data, assisted in conducting literature review and wrote the paper. The third author designed and analyzed the data, and wrote the paper. The fourth assisted in conducting the study and wrote the paper. The fifth author helped with the framing of the paper and provided comments for improving the paper. All authors read and approved the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported in part by Ministry of Education of the People's Republic of China (Grant No. 22YJC880058 and 21YJA880078), the National Natural Science Foundation of China (Grant No. 62107020), Knowledge Innovation Program of Wuhan-Shuguang Project (Grant No. 2022010801020281), University-level Educational Reformation Research Project for Undergraduate Education, Central China Normal University, China (Grant No. 202143, CCNUAI&FE2022-03-15), the Fundamental Research Funds for the Central Universities (Grant No. CCNU23KYZHSY17). The authors would also like to thank the National Public Service Platform of Educational Resources for providing video data of the Real Classroom Dataset.

References

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*(2), 256–274. https://doi.org/10.1037/0033-2909.111.2.256

Aydin, B., Kindiroglu, A. A., Aran, O., & Akarun, L. (2016). Automatic personality prediction from audiovisual data using random forest regression. December. In *2016 23rd international conference on pattern recognition (ICPR)* (pp. 37–42). IEEE.

Bai, J., Gao, H. L., Wang, Y. Z., Yang, L. B., Xiang, X. H., & Lou, X. W. (2020). Detection of students' classroom performance based on faster R-CNN and transfer learning with multi-channel feature fusion. *Journal of Guangxi Normal University, 5*, 1–11. https://doi.org/10.16088/j.issn.1001-6600.2020.05.001

Brey, E., & Pauker, K. (2019). Teachers' nonverbal behaviors influence children's stereotypic beliefs. *Journal of Experimental Child Psychology, 188*, Article 104671. https://doi.org/10.1016/j.jecp.2019.104671

Brophy, J. E., & Good, T. L. (1984). Teacher behavior and student achievement. *Handbook Of Research On Teaching*, 340–370.

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7291–7299).

Chang, C. F., Gröschner, A., Hall, N. C., Alles, M., & Seidel, T. (2018). Exploring teachers' emotions via nonverbal behavior during video-based teacher professional development. *AERA Open, 4*(4), Article 2332858418819851.

Chen, S. H. (2020). *Research and application of quantitative computational methods for teachers' non-verbal behaviors in intelligent classroom environment*. Central China Normal University. https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CDFDLAST2021&filename=1020117152.nh.

Ding, N. (2020). *Intelligent Analysis and Recognition of teachers' body Movements in middle school classroom teaching videos*. Central China normal university. https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202101&filename=1020119488.nh.

Flanders, N. A. (1970). *Analyzing teacher behavior* (Vol. 107). Massachusetts: Addison-Wesley.

Gholamrezaee, S., & Ghanizadeh, A. (2018). EFL teachers' verbal and nonverbal immediacy: A study of its impact on students' emotional states, cognitive learning, and burnout. *Psychological Studies, 63*(4), 398–409. https://doi.org/10.1007/s12646-018-0467-5

Gu, X. Q., & Wang, W. (2004). Classroom analytics techniques for teachers' professional development: A new exploration. *China Educational Technology*, (7), 4.

He, C. H. (2009). *A study on language teachers' classroom gestures*. Xiangtan University.

He, X. L., Fan, Y., Chen, Z. Z., Fang, J., & Li, Y. Y. (2020). The recognition of student classroom behavior based on human skeleton and deep learning. *Modern Educational Technology, 11*, 105–112.

Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., … Zisserman, A. (2017). *The kinetics human action video dataset. arXiv preprint arXiv: 1705.06950*.

Keller, M. M., Hoy, A. W., Goetz, T., & Frenzel, A. C. (2016). Teacher enthusiasm: Reviewing and redefining a complex construct. *Educational Psychology Review, 28*(4), 743–769.

Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., & Serre, T. (2011). HMDB: A large video database for human motion recognition. November. In *2011 International conference on computer vision* (pp. 2556–2563). IEEE.

Lechuga Redondo, M. E., Niewiadomski, R., Francesco, R., & Sciutti, A. (2022). Comfortability recognition from visual non-verbal cues. November. In *Proceedings of the 2022 international conference on multimodal interaction* (pp. 207–216).

Liu, X. Y., Ye, S. P., & Zhang, D. H. (2020). Improved multi-objective regression student classroom action detection method. *Computer Engineering and Design*, 2684–2689.

Mazer, J. P. (2013). Associations among teacher communication behaviors, student interest, and engagement: A validity test. *Communication Education, 62*(1), 86–96. https://doi.org/10.1080/03634523.2012.731513

Meng, B. H. (2022). *Research on intelligent recognition of teaching gestures for smart classroom*. Central China Normal University. https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFDTEMP&filename=1022090628.nh.

Nguyen, L. S., Frauendorfer, D., Mast, M. S., & Gatica-Perez, D. (2014). Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior. *IEEE Transactions on Multimedia, 16*(4), 1018–1031.

Özmen, K. S. (2010). Fostering nonverbal immediacy and teacher identity through an acting course in English teacher education. *Australian Journal of Teacher Education, 35*(6), 1–23.

Pang, S. Y., Zhang, A. R., Lai, S. H., & Yang, Y. Q. (2023). Automatic recognition of teachers' nonverbal behaviors based on graph convolution neural network. In *International conference on education technology and computers (ICETC)* (pp. 429–435). ACM.

Pang, S. Y., Zhang, A. R., Lai, S. H., & Zuo, Z. Q. (2022). Automatic recognition of teachers' nonverbal behavior based on dilated convolution. In *2022 international conference on information systems and computer aided education (ICISCAE)* (pp. 162–167). IEEE.

Raza, A., Yousaf, M., Sial, H., & Raja, G. (2015). HMM-Based scheme for smart instructor activity recognition in a lecture room environment. *The Smart Computing Review*, 578–590. https://doi.org/10.6029/smartcr.2015.06.008

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems, 28*.

Richmond, V. P. (2002). Teacher nonverbal immediacy: Use and outcomes. In J. L. Chesebro, & J. C. McCroskey (Eds.), *Communication for teachers* (pp. 65–82). Boston, MA: Allyn & Bacon.

Rodríguez, C. V., Lavalle, M. M., & Elías, R. P. (2015). Modeling student engagement by means of nonverbal behavior and Decision trees. November. In *2015 international conference on mechatronics, electronics and automotive engineering (ICMEAE)* (pp. 81–85). IEEE.

Šerić, M. (2021). The relationship between teacher non-verbal communication and student behavior: A cross-national perspective. *Journal of Communication Inquiry, 45*(4), 383–410.

Shahroudy, A., Liu, J., Ng, T. T., & Wang, G. (2016). Ntu rgb+ d: A large scale dataset for 3d human activity analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1010–1019).

Simpson, A. W., & Erickson, M. T. (1983). Teachers' verbal and nonverbal communication patterns as a function of teacher race, student gender, and student race. *American Educational Research Journal, 20*(2), 183–198.

Soomro, K., Zamir, A. R., & Shah, M. (2012). A dataset of 101 human action classes from videos in the wild. *Center for Research in Computer Vision, 2*(11).

Tu, Z., Xie, W., Dauwels, J., Li, B., & Yuan, J. (2018). Semantic cues enhanced multimodality multistream CNN for action recognition. *IEEE Transactions on Circuits and Systems for Video Technology, 29*(5), 1423–1437.

Tu, Z., Xie, W., Qin, Q., Poppe, R., Veltkamp, R. C., Li, B., & Yuan, J. (2018). Multi-stream CNN: Learning representations based on human-related regions for action recognition. *Pattern Recognition, 79*, 32–43.

Wang, M., Chen, Z., Shi, Y., Wang, Z., & Xiang, C. (2022). Instructors' expressive nonverbal behavior hinders learning when learners' prior knowledge is low. *Frontiers in Psychology, 13*, Article 810451.

Wang, W., & Loewen, S. (2016). Nonverbal behavior and corrective feedback in nine ESL university-level classrooms. *Language Teaching Research, 20*(4), 459–478. https://doi.org/10.1177/1362168815577239

Wei, Y. T., Qin, D. Y., Hu, J. M., Yao, H., & Shi, Y. F. (2019). The recognition of students' classroom behaviors based on deep learning. *Modern Educational Technology, 29*(7), 87–91.

Won, A. S., Bailenson, J. N., & Janssen, J. H. (2014). Automatic detection of nonverbal behavior predicts learning in dyadic interactions. *IEEE Transactions on Affective Computing, 5*(2), 112–125.

Wu, D., Chen, J., Deng, W., Wei, Y., & Wei, Y. (2020). The recognition of teacher behavior based on multimodal information fusion. *Mathematical Problems in Engineering, 2020*(1), 1–8.

Xie, W., Tao, Y. P., Gao, J., Zhou, D. B., & Wang, W. (2022). YOWO Based real-time recognition of classroom learning behaviors. *Modern Educational Technology, 32*(6), 107–114.

Xu, J. Z., Deng, W., & Wei, Y. T. (2020). Automatic recognition of student's classroom behaviors based on human skeleton information extraction. *Modern Educational Technology*, (5), 108–113.

Yang, J. M., Yang, W. D., Chen, H., Wu, C. C., & Pi, Z. L. (2022). Do teacher gestures in instructional videos work? A meta-analysis based on 40 experimental and quasi-experimental studies from 2000-2021. *Modern Distance Education Research, 34*(1), 92–103.

Yan, X. W., Zhang, Z. H., Zhao, X. Y., & Jia, L. Y. (2019). Teachers' body movement detection for classroom teaching evaluation. *The Chinese Journal of ICT in Education, 16*, 88–91.

Zeng, J. L. (2020). *Research on the algorithm of students' classroom behavior detection based on video*. Harbin Institute of Technology.

Zhang, H. Y. (2016). *Design and implementation of learning behavior measurement system for classroom environment*. Huazhong University of Science and Technology.

Zhao, G., Zhu, W., Hu, B., Chu, J., He, H., & Xia, Q. (2021). A simple teacher behavior recognition method for massive teaching videos based on teacher set. *Applied Intelligence, 51*, 8828–8849.

Zhou, P. S. (2003). *A quantitative study of teachers' classroom nonverbal behavior*. Guangxi Normal University.

Shiyan Pang is an associate professor at the Faculty of Artificial Intelligence in Education at Central China Normal University. Her research interests include K12 AI education, application of artificial intelligence in education, data mining, classroom data analysis. Her publications have appeared in international journals such as ISPRS Journal of Photogrammetry and Remote Sensing, and International Journal of Remote Sensing.

Shuhui Lai is a graduate student at the Faculty of Artificial Intelligence in Education in Central China Normal University. Her research interests focus data mining, and AI literacy.

Anran Zhang is a graduate student at the Faculty of Artificial Intelligence in Education in Central China Normal University. His research interests focus data mining, deep learning and AI literacy.

Yuqin Yang is an associate professor of the learning sciences at the Faculty of Artificial Intelligence in Education at Central China Normal University. Her research interests include pedagogy and assessment of Knowledge Building, data science education, learning analytics, and collaborative learning. Her publications have appeared in international journals such as American Educational Research Journal, Internet and Higher Education, International Journal of Computer-Supported Collaborative Learning, British Journal of Educational Technology, Teaching and Teacher Education, International Journal of Educational Technology in Higher Education, Thinking Skills and Creativity, and Journal of Computer Assisted Learning.

Daner Sun is assistant professor at the Department of Mathematics and Information Technology, The Education University of Hong Kong. Her research interests are ICT-

supported science learning, seamless learning and teacher education. Her publications have appeared in international journals such as Journal of Research in Science Teaching, Thinking Skills and Creativity, Computers & Education, International Journal of Computer-Supported Collaborative Learning, and Education Technology and Society.

Computers and Education Open 4 (2023) 100139

Contents lists available at ScienceDirect

Computers and Education Open

journal homepage: www.sciencedirect.com/journal/computers-and-education-open

Defining and exploring online engagement fatigue in a university context

Suzanne Maloney [a,*], Megan Axelsen [a], Cathy Stone [b], Linda Galligan [a], Petrea Redmond [a], Alice Brown [a], Joanna Turner [a], Jill Lawrence [a]

[a] University of Southern Queensland West Street, Toowoomba, Qld, 4350 Australia
[b] University of Newcastle University Drive, Callaghan NSW 2308 Australia

ARTICLE INFO

Keywords:
Engagement
Online engagement fatigue
Regional university
Online engagement

ABSTRACT

This paper seeks to define online engagement fatigue. The need for a definition is twofold. First, the increased reliance on the internet for the full delivery of higher education courses could result in online engagement fatigue. Second, a clear definition is required for the theoretical construct of online engagement fatigue to ensure meaningful research and to advance knowledge. Thus, the challenge is to understand what online engagement fatigue is and what are its consequences for students and educators. A social constructionist approach was used to develop an emergent definition of online engagement fatigue, which was then refined through a qualitative exploration of how it is perceived by students and educators. The perceptions that students and educators hold about online engagement fatigue were collected through interviews with 18 students and ten educators. There were differences in how the students and educators perceived online engagement fatigue, with students more likely to confirm its existence and more likely to render examples of its effect. A refined definition is offered which gives a basis upon which future research can investigate this phenomenon in divergent settings.

1. Introduction

1.1. Online education

Before the COVID-19 pandemic, digital technology was fast becoming a central aspect of higher education. Spurred on by the pandemic, the dominant mode of delivery is now through online resources where, around the world, courses are increasingly being designed to be delivered through technology — 'digital first' — and supplemented by face-to-face human support [1]. Affecting all aspects of the student experience, digital technologies and the move to online learning has also been linked to an increase in behavioural, affective, and cognitive student engagement [2]. During the period when universities were pivoting to online learning, research by the OECD [3] indicated that most universities were inadequately equipped for the swift shift to online teaching and learning. These institutions were poorly equipped in terms of technology and training; however, they were particularly underprepared for the changes in the format of resources and pedagogical approaches.

As learning has moved into the online context, the facilitation of online student engagement has become a central concern of educators [2]. The ability to analyse learners' 'digital footprints' (trace data from their interactions with technology) using learning analytics data also means that student access to online resources – and 'counting clicks' – has become a simple measure (rightly or wrongly) of their engagement [4–6]. The push to ensure students are genuinely engaged in their learning, and to show evidence of that engagement thus raises the question of whether institutions are placing too much emphasis on the importance of visible student engagement in the online environment [7]. In doing so, are these engagement practices, by consequence, pushing students to become fatigued from those very engagement strategies designed to support students' learning?

While a large and growing body of research has explored the concept of online engagement [8–12] and the factors that may lead to increased engagement, a concept that has not been explored, and which may equally impact on student's learning outcomes, is *online engagement fatigue*. The purpose of this paper is to explore the concept of online engagement fatigue – whether it exists, how it can be defined, what are its defining elements, and to what extent does it play a role in online student disengagement. Finding answers to these questions may

* Corresponding author.
E-mail addresses: suzanne.maloney@usq.edu.au (S. Maloney), megan.axelsen@usq.edu.au (M. Axelsen), cathy.stone@newcastle.edu.au (C. Stone), linda.galligan@usq.edu.au (L. Galligan), petrea.redmond@usq.edu.au (P. Redmond), alice.brown@usq.edu.au (A. Brown), joanna.turner@usq.edu.au (J. Turner), jill.lawrence@usq.edu.au (J. Lawrence).

https://doi.org/10.1016/j.caeo.2023.100139
Received 5 April 2022; Received in revised form 17 March 2023; Accepted 30 April 2023
Available online 1 May 2023
2666-5573/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

ultimately assist educators to design and teach online courses in ways that are better able to build and maintain student engagement.

1.2. Student online engagement

Student engagement has long been seen as a precursor to learning [13], and something that is linked to improved graduation rates, classroom motivation, course achievement, retention, and persistence [14–17]. This is based on the acceptance that engagement is "a positive work-related state of fulfilment that is characterised by vigour, dedication and absorption" (Schaufeli et al., 2002 as reported in [18], p.3). In contrast, disengagement has been found to have a profound effect on student learning outcomes, cognitive development, and the quality of the student experience ([19]; Higher Education Standards [20–22]). From an online learning perspective, student engagement is recognised as crucial to student learning and satisfaction in online courses [23].

Even though the concept of engagement in education has been widely studied and written about [24–26], there remains a considerable variation in how student engagement is defined. It is therefore important for research projects to begin with a clear definition of their own understanding [27], upon which future research could contribute. In this project, the definition of student engagement offered by Bond et al. [2] was used as a starting point to understand the concept:

> Student engagement is the energy and effort that students employ within their learning community, observable via any number of behavioural, cognitive or affective indicators across a continuum ([2], p. 3).

While educational technology has long been seen as a way to improve student engagement, and a well-designed online course can assist in promoting student engagement [28], there is "no guarantee of active student engagement as a result of using technology" ([2], p. 2). Online learning environments need to be intentionally designed. This is because without careful planning and sound pedagogy, technology can actually lead to disengagement and impede rather than help learning [29–31]. For example, designers of online learning environments can fail to recognise that the online space is not a classroom and fall into a content-publishing mentality [32]. In addition, some of the online learning strategies that are being used to engage students – and perhaps to also provide evidence of engagement – may have the opposite effect by negatively affecting cognitive load [33], or overwhelming students with too many learning tools, too many links [34], or too much online content.

In an academic context, students often face study stress associated with high study demands and concerns about academic progress [35]. Study stress can also develop because of excessive academic load, the classroom environment, faculty interaction [36], concerns regarding the nature of current academic and social circumstances [37], or from merely being tired of learning [38]. If study stress is prolonged and exceeds the student's adaptive resources, it can result in high levels of *study-related fatigue* [39]. de Vries, et al. [39] found that many university students experience high levels of study-related fatigue brought about by prolonged stress due to concerns about academic achievement and the increased study demands that achievement requires.

2. Research question

While the concept of online engagement fatigue in higher education is not a term currently used in the literature, there has been some research undertaken that explores the relationship between fatigue and the academic attainment and wellbeing of students [40]. Some research has also explored the perspective of 'learning weariness'. Wang [41] defines learning weariness as a student's loss of interest in studying and tiredness towards learning, which can often embody indifferent attitudes, boredom, psychological conflict, and the potential for poor behaviour as a reaction to learning. Fatigue has also been recognised as something students experience because of family responsibilities [42], employment commitments [43], academic schedule [44], or poor sleep patterns [45]. This research seeks to extend this literature by examining fatigue in an online learning context. The overarching educational problem is whether online engagement fatigue is real. Specifically, the research questions are 1. Does online engagement fatigue exist? and 2. If it does exist, what are its defining elements and effect?

3. An emerging definition of online engagement fatigue

Online engagement fatigue is a phenomenon that the research team (which is mostly comprised of teaching academics) had anecdotally noticed while undertaking both their teaching responsibilities and their corresponding research addressing methods to engage students online. During their online teaching experiences, the research team noticed that at different times during the semester some online students' engagement seemed to diminish, for example, they participated less in discussions, shared fewer posts, interacted less with peers, educators, and the Learning Management System (LMS). This led to the team asking: does this lower participation/interaction reflect students' becoming very tired of being engaged? Or are they perhaps getting tired of showing their engagement? Maybe they are simply getting on with their work, with less evidence of engagement through the usual measures; or are they finding it very hard to keep going (for whatever reasons) and are therefore at risk of attrition? To inform the research and understanding of the phenomenon, the research team sought to first define online engagement fatigue.

The need to define online engagement fatigue was considered an important first step in seeking its existence. This allowed the development of a mutual understanding amongst the researchers of the identified factors encompassing online engagement fatigue. This definition was then used to develop the interview schedule and it was able to be interrogated during the qualitative analysis. To define online engagement fatigue the team adopted a social constructionist approach [46] that affords a subjectivist view of knowledge to be embraced, where knowledge is the result of social interchange. In this study the researchers moved through a process of filtering and adapting existing knowledge (including insights from existing research and literature) and combining this with additional contextual and experiential knowledge. The emergent definition for online engagement fatigue was based on agreed categories and themes in order to construct new knowledge [47].

The process, elaborated on in the next section and depicted in Diagram 1, commenced with an investigation of existing literature of key terms and phrases that the team determined to be associated with the concept of online engagement fatigue. These terms included: engagement in higher education; engagement in online learning; the concept of fatigue (fatigue, burnout, emotional exhaustion, cognitive/mental fatigue); and the effect of fatigue on online engagement (disengagement, academic achievement, interest). Once the review of the literature was complete the team moved through a constant comparison approach to collate emerging themes, key concepts, and terms [48]. Finally, the researchers individually, and then collectively, used a deductive approach to further filter concepts and implications from the literature, and identify any additional themes. This included a three stage-phase filtering process of: a) discussing shared understandings of what was meant and understood by the term online engagement fatigue; b) ensuring that the wording was succinct and relevant; and c) ensuring that the term had broad application.

The emerging definition of online engagement fatigue was defined by the research team as:

> *A reduction in online students' enthusiasm and motivation for engaging in course activities as a result of overexposure to online coursework and associated interactions.*

This emergent definition, developed through a social constructivist approach, was used to inform the qualitative stage of the project that

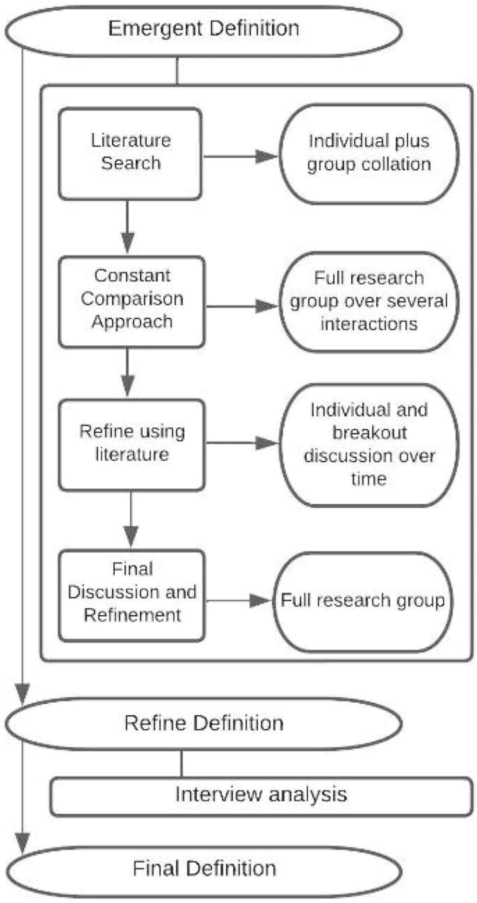


Diagram 1. Process Undertaken for Emergent Definition to Final Definition.

sought to empirically investigate the existence and defining elements of online engagement fatigue through the perceptions of students and educators.

4. Methodology

4.1. Method

The purpose of this stage was to undertake a qualitative process of interviewing students and educators to enable the research questions to be addressed. The process of developing a final definition is shown in Diagram 1. This qualitative method was chosen as it was important to explore the phenomenon through those who would experience it in a learning context and those who may witness it or have an ability to affect it in some way. That is, the aim was to explore whether students and educators attributed meaning to the concept of online engagement fatigue and, if so, how did they describe it? Given the exploratory nature of the research aim, participants were sourced from multiple disciplines. This was to allow the widest capture of themes from the key stakeholders (students and educators) in which to explore the existence of online engagement fatigue and refine its definition. The study received ethics approval at the university where the study took place.

4.2. Context

The study was undertaken at a regional university that has long offered courses in both on-campus and fully online modes. Before online learning became popular and a commonplace offering in the tertiary sector, the university was a leading Australian higher education provider of distance education, and in 2019, before the COVID-19 pandemic, two-thirds of its students were enroled as fully online students. In fact, most students enroled at the university undertake at least some of their courses in the online mode, even if they are enroled as on-campus students. The regional university also prides itself on delivering education to low socio economic (low SES) and regional, rural and remote (RRR) learners; and as such, the research also focussed on these students because online engagement fatigue issues could be more elevated for them given their increased vulnerability (isolation studying online, intermittent connectivity, unfamiliarity with university and discipline discourses and ongoing disadvantage in higher education). Indeed, research has found that students from less advantaged groups struggle more than students from higher economic class households to feel connected and engaged in their classes, coursework, faculty, and peers; and this disparity is heightened when it comes to online learning [49].

4.3. Participant selection

The university administration system was used to identify potential students for inclusion in the study and to extract contact information. Potential participants for the study were first-year students who were enroled in undergraduate courses at the university in semester 1, 2020, who had passed at least one course previously, and who were identified as domestic and rural, regional, or remote. The research design limited the target population to this group for two reasons. First, to put a boundary around what could reasonably be achieved given the exploratory nature of the research aim. Second, to help control confounding factors that could influence the meaning, experience, and effect of online engagement fatigue across various cohorts, including post-graduate and graduate students, and of years of online study. Whilst this limits the generalisability of the findings in the first instance, it was important to gather a base line view that could be meaningfully expanded in future research. Using these parameters, the researchers identified 203 potential participants who were then contacted via text message to determine their interest in participating. Eight responses were received, each expressing interest, while the remainder were followed up via email, generating a further ten expressions of interest. These 18 students were sent a participant information sheet and consent form in line with the ethics protocol, and each agreed to participate. Hence, this was an opportunistic, or convenience sample rather than a strictly representative one. Participants were then contacted to arrange an interview time and sent a $20 department store voucher in appreciation of their assistance.

Semi-structured, in-depth interviews, of between 40 mins and one hour duration, were conducted with all 18 student participants, who were given the choice of a face-to-face (three students) or online interview via Zoom (15 students). All interviews were conducted by one researcher not personally known to the students, using an interview schedule. All were recorded with the students' consent.

Members of the teaching staff from the host institution were also recruited for interviews, using researcher networks within the university and their teaching areas to invite staff with more than three years' online teaching experience, and from a broad range of disciplines, to participate. Those who expressed interest were given a participant information sheet and consent form, resulting in ten educators agreeing to participate. For consistency and to reduce bias, all interviews were conducted by one researcher who was external to the university and therefore not personally known to the participants. As with the student interviews, staff interviews were semi-structured and in-depth, using an interview schedule, and each was recorded with the staff members' consent.

4.4. Participant information

The students who participated in the interviews were enroled in a range of programs, including nursing (5), education (4), business, commerce and law (3), science (3), creative arts (2), and aviation (1). All

were enroled full-time, and most (13) of the student interviewees were aged under 25 with two aged over 55. As well as studying full-time, thirteen of the student interviewees also undertook part-time work (average of 20 h per week) and one had full-time carer's duties. Eight were male and ten were female.

The educators who participated in the interviews taught into a range of disciplines, including law, commerce, business, accounting, nursing and midwifery, engineering, computing science, food science, sport and exercise science, and education. Between them, the educators had an average of 16.5 years of higher education teaching experience. They taught an average of 3.5 courses per year which consisted of an average of 270 students per course. One of the educators taught fully online, three taught mostly online, five taught both online and face-to-face, and one taught mostly face-to-face. Seven were male and three were female. Detailed participant information is contained in Appendix 1.

4.5. Interview schedule

The data were collected via in-depth semi-structured interviews as reflecting the exploratory nature of the research. This approach provided flexibility in terms of questions and prompts [50] and was suitable for uncovering perceptions of the varying elements that may shape online engagement fatigue. This was in line with the two research questions seeking to reveal the existence of online engagement fatigue and to capture its defining elements and effects.

Two interview schedules, one for students and the other for staff, were developed from the first stage of the project. Both contained initial demographic questions followed by questions to explore interviewee perceptions of their experience, as either an online learner or online teacher, and to solicit personal stories and insights regarding any experiences of online engagement fatigue from an empathetic position [51]. The student interview schedule contained questions seeking to capture their idea of 'engagement', what behaviour they perceived an engaged student would show, the types of activities that help or hinder engagement, and the extent they feel they engage. They were then given the research team's definition of online engagement fatigue, asked if they had ever experienced this type of fatigue as a result of engaging with their online studies and, if so, how frequently. They were asked to describe such occurrences, such as when it occurs, what contributes or triggers this type of fatigue, and what impact they felt it has had on them and their studies. This line of questioning helped establish whether students experience online engagement fatigue (research question one – does online engagement fatigue exist?), what characterises it, and with what effect (research question two – what are its defining elements and effect?). The educator interview schedule followed a similar development but relevant to the teacher perspective, capturing a macro-level view and giving the educators the opportunity to discuss any interventions or strategies they employed to engage students or to diminish student fatigue. They were asked if they witness online engagement fatigue in their students and, if so, to describe what they observe, such as when it happens, whether they notice it in one course more than another, and how it impacts students both individually and from a student cohort perspective.

4.6. Qualitative data analysis

In-depth interviews generally result in large volumes of rich data, and this was certainly the case in this research with 28 in-depth interview transcripts. The main purpose of the qualitative analysis of the transcripts was to identify themes that represent the ways the students describe and experience online engagement fatigue in their online studies and the ways the educators witness and describe this fatigue in their online courses.

The interview transcripts were content analysed for both semantic and latent themes (thematic analysis) [52,53] using two stages of analysis. First, the data was analysed using the computer-assisted qualitative data analysis software, NVivo. This analysis helped to identify the explicit, surface meanings of the data, that is semantic themes. Specific phrases and words were identified using this first level of analysis. For example, for the theme *overwhelmed*, specific instances of this term were identified:

> *For me it's feeling overwhelmed. I mightn't have been able to get my lectures done in the previous week. The worst I've ever got behind was about six weeks behind with my lectures.* (Bella)

> *Probably overwhelmed because everything just came at once in every subject* (James)

To ensure, and increase, data dependability (i.e., by use of two analysis methods: computer software and manual) the manuscripts were then manually analysed to identify further instances where the semantic themes might be apparent in a more latent manner. As argued by Clarke and Braun [52], a common drawback of qualitative content analysis (especially when utilising qualitative analysis software) is the use of the main interview questions as the themes rather than interpreting the themes to make sense of them. The manual analysis was therefore important for looking beyond what was said [54,55] to identify and examine the underlying ideas, assumptions, and conceptualisations that informed and built upon the semantic content of the data., This manual analysis also considered the semantic themes in relation to the full transcripts, thus ensuring all instances of the semantic theme were identified. For example, using the theme *overwhelmed* again, the manual analysis identified instances where participants described feelings of being overwhelmed without the use of the word 'overwhelmed:

> *Then the assignments start piling up. You start stressing and then you start to overload with more information trying to catch up.* (Liam)

> *In literature courses you have a lot of books and novels, plays, things you need to read but then you've also got extra reading activities on top of those things that you need to read. Then the education course you have lots of theorists and theories and stuff that you need to read on. So, I guess when you first start in the course it's like, oh we're going to be looking at these things but you're not actually reading them yet. But then by the time you get to the middle of semester you've got two novels you've got to finish, you've got eight theories – eight theorists you need to read up on and you've got all of the readings that you might have missed that you need to read up on.* (Leah)

The manual thematic analysis also allowed the researchers to explore all the possible nuances and relationships, to view data from a variety of perspectives, and to move from micro- to macro-view in order to "support the analytic imagination necessary for understanding and theory generation" ([56], p. 12). The manual analysis was undertaken by two researchers who individually analysed the data and then compared findings to review, modify and refine the themes that emerged. This collaborate refinement of the themes was also important for identifying the 'essence' of each theme, that is, what the theme was about (Braun & Clarke, 2017), as well as to ensure the emergent findings were reliable and valid through a process of checking, discussion and agreement. A thematic map was then developed to illustrate the relationships between the themes (see the discussion).

The analysis found several themes that can be related to triggers of online engagement fatigue such as overexposure due to the quantum and timing of content and activities, as well as the effect of online engagement fatigue such as feelings of being overwhelmed, loss of motivation and reduced enthusiasm. These themes are discussed in the next section. In reporting the data, pseudonyms have been used.

5. Findings

The concept of online engagement fatigue and the way it impacts on student learning was perceived quite differently by educators and students. While the educators had mixed views about whether online

engagement fatigue was a real phenomenon, seventeen (of the eighteen) students felt that online engagement fatigue was something they had experienced in their online learning. Twelve students also indicated it was something they had noticed amongst their peers.

Interestingly, whilst the educators were uncertain about whether the phenomenon was real, they demonstrated a high level of awareness and empathy for the varying and demanding circumstances of online students and spoke of the quantum of course artefacts and related interactions. They took steps to manage and respond to the impact of this on students and spoke about 'checking in' with students, trying not to overload them with tasks and messages, keeping the technology demands as simple as possible, and 'chunking content' to make it more interesting and manageable. These are perhaps responses that can be understood as helping to prevent online engagement fatigue. So, while educators may be uncertain whether online engagement fatigue has been experienced by their students, they signalled an awareness of its possibility.

The students, on the other hand, spoke at length about experiencing the phenomenon and several themes emerged that reflect the ways students described and perceived online engagement fatigue. These showed that students defined online engagement fatigue in terms of feeling overwhelmed, as a loss of motivation, and as something that occurred when they had been online so often and for so long that it became 'too much'.

5.1. Overexposed and overwhelmed

More than half of the students (10) described online engagement fatigue in terms of feeling overwhelmed. Mostly they spoke about feeling overwhelmed due to the amount of learning content and how much coursework they thought they were expected to complete. In feeling overwhelmed, some of the students described wanting to 'shut down' or quit (in response to feeling that way).

> *I physically could not do all of the content and that was quite overwhelming especially when you hit that week six or seven and you're just like, oh my goodness, I haven't done anything, but that's not the case. That's definitely – then it starts to feel like you're not going to do well for the next bit so I would say that is fatigue …I just didn't do anything for about three weeks. That included my break, so I just didn't. This semester I didn't do anything because I was just like, this is too hard.* (Jenny, arts student)

> *I shut down personally. I say, no that's it. I'm done. I'm quitting. I – there was lots of times where I actually reached out to UNIVERSITY ACRONYM and said, can I please defer?* (Charlotte, aviation student)

Three of the educators also mentioned observing students become overwhelmed by content generated by communication. All three attributed this to the online learning context and the contemporary push to 'engage' students. As explained by the educator Rosemary when discussing the universities virtual classroom called StudyDesk:-

> *StudyDesk is overwhelming. The students get a million messages from the university now, like emails and messages and stuff. There's more and more stuff that we're encouraged, as educators, to put on our StudyDesk. So, they become more and more busy.*

To cater to a variety of learning preferences, educators are often encouraged to provide a variety of resources on the LMS. In addition, there are online student forums, a welcome video, an orientation video, videos to encourage connection between theory and practice, course announcements, extra readings, and the standard weekly lecture recordings. The LMS can easily become a place where students might find themselves overwhelmed by the complexity and the amount of content offered. As suggested by educator William, with all the activities educators can do to try and get students engaged and involved in their online learning, it is easy to "*overdo it*". Providing a semester's worth of online learning content from day one of semester may also contribute because students "*tend to read way ahead and then get overwhelmed*" (educator Amber).

Another six of the students suggested that being online so often became too much. The amount of time they were required to be online to complete learning tasks contributed to their propensity to suffer from online engagement fatigue and hindered their engagement with the course. They used words like 'overloading', 'tiresome', 'too much' and 'sick of it'.

> *I found that I couldn't sit as long in front of the computer so our sessions on Monday and Tuesday would be nine till four and I couldn't cope with nine till four. [It's that long] so I would do the first lecture or even half the lecture, go and make myself a cup a tea or go for a walk and listen to another lecture. Then I wouldn't do anything for the rest of the day because it would be too much – like too overloading.* (Charlotte, aviation student)

> *The online just started to become a little bit tiresome for me…. I think it's just having to* log *on, on a regular basis and just do the work, it can get a bit exhausting.* (James, psychology student)

> *You're just too much online, at some point it just gets too much.* (Bella, nursing student)

These students also spoke about specific online learning activities that contributed to their online engagement fatigue. For example, three students talked about being faced with too many forum posts; three mentioned that online lectures were too long (more than two hours); one talked of too much reading; and another of inadequate resources (due to not being on-campus) resulting in looking at a very small screen all day. A real struggle for many was spending such a significant amount of time at a computer and being required to 'be online' so often and so much, indicating that this in itself may be a trigger for online engagement fatigue.

The combination of having multiple course activities or assessment items across several courses occurring 'all at once' was raised by seven students as something they felt was a trigger for online engagement fatigue. As explained by Bella (nursing student), online engagement fatigue "*especially* [happens] *when you see your calendar and it says, lecture, lecture, lecture, lecture, assignment due, assignment due, quiz*". Related to this is feeling overloaded by the amount of course content or by the difficulty of that content and the perception of there being too little time.

> *I definitely think it is a content thing probably more than likely. They've released a bunch of – all the stuff that you need to do and then it just starts piling up at that point if you are not on top of it and no one really is. You can't be on top of everything …* (Jenny, arts student)

> *I was with a bunch of people and they were saying there was way too much knowledge …. Their brains were like – they were going gooey after – after a couple of hours they – some even said they weren't even considering coming to the last class just because there was way too much knowledge.* (Liam, exercise science student)

The above highlights the enormity of the task for students in putting boundaries and strategies in place to minimise their overexposure and reduce the likelihood of being overwhelmed.

5.2. Timing

There were differing views from the students about 'when' during semester they most often experienced online engagement fatigue. Two students said it happened at the end of semester once the work was completed, four said it occurred when assessments were due, while seven described experiencing online engagement fatigue and a related reduction in enthusiasm around mid-semester, at which point there was also a loss of productivity. As explained by Leah (education student),

I have noticed in myself a drop in interactions within the forum posts and things like that just before mid-semester breaks. Also, after mid-semester break, I find that it just continues to trail off until the end of semester.

Of the four students who felt their online engagement fatigue occurred when assessments were due, one discussed the increase in forum posts at this point. He indicated he spent significant amounts of additional time reading all the extra posts that were put up by his fellow students in case there was something pertinent in those posts related to the assessment. He felt he was negatively impacted by this extra reading and it contributed to his fatigue. Another student felt that when assessments were due, her focus swapped to completing that assessment rather than completing coursework activities; while another felt that she experienced online engagement fatigue just after a 'round' of assessments were completed. In her words,

When my first assignments were due, and then between that and the next round of assignments I just felt really tired of how much work I put into those assignments. (Isobel, education student)

The impact of assessment on online engagement was also a theme that emerged from nine of the 10 educator interviews. While three of the educators observed increases in participation just before exams and when assignments were due, six noted the opposite, perceiving that participation drops off when assessments are due. These six also noted the need to be mindful that students will often have multiple assessment items for multiple courses, impacting their level of observed involvement in a course and potentially leading to online engagement fatigue. As William commented,

This is a student I know, a really good student, very committed and doing quite well, but they said they were so tired because they just had all these assignments in this one week.

Nine educators spoke about witnessing a 'drop off' in engagement at specific points in the semester, but many questioned whether this was due to online engagement fatigue. They also noted that the drop off in engagement was generally only an issue they witnessed in first-year students; for students in subsequent years of study, engagement was more stable throughout the semester, and this was especially the case for post-graduate students.

5.3. Other factors driving reducing involvement

There was uncertainty amongst educators about whether the drop off in participation they witnessed occurred because of online engagement fatigue or was simply an indication that students' involvement was impacted by other factors. Five of the educators suggested this drop off reflected the impact that outside influences (such as work, family, health) had on students' time. Interestingly, the impact of outside influences on students' propensity to engage with the course was also discussed by half (9) the student participants. These students described outside influences as something that both impacted on their ability to engage with their learning and something that triggered their online engagement fatigue. As explained by Emily (education student), "*I was so busy with work, where I'd get home from work, and I'm like I'm so tired, I really don't want to do uni. I really don't want to have to talk to people on the screen*". Outside influences mentioned by the students included family or caring commitments, work, health-related issues, and lacking the technological skills required to study fully online.

Four of the educators suggested the drop off in engagement they witnessed was perhaps an indication that students were simply getting on with the work and being "*quite strategic about what they do and don't do*" (Rosemary). To provide students with a flexible study environment that might support their ability to undertake their studies in a manner that enables lifestyle integration, it is now policy at the university where this research took place that all course materials (for the entire semester) are available to students on the LMS from day one of semester. The provision of all online course resources from day one means that students often do not need to engage with the course or their educators and can be 'strategic' about those learning activities in which they engage. As observed by educator Simon, many students do not necessarily want to interact with anyone and are quite happy to undertake their learning independently. Spoken by Simon from a student's voice:

I just want to get my course curriculum, submit my whatever I've got to submit, get 100 per cent, because that's what I like to do in these courses, and to heck with having to interact with anyone.

Simon also noted that for these students, there may be little evidence of their engagement with the learning materials online through the LMS; however, as the quote below indicates, he felt that this does not mean the student is not engaged:-

You can't make the universal assertion that a non-clicking student is a poor student … Universities have this nasty habit - well, not a nasty habit, but I routinely see it where they consider a student with only 10 clicks on your page for a semester a poor student. [Whereas] the student just doesn't want to have anything to do with you. They're a high-performing student who's got their act together, and they just want the curriculum, the assessments, and then put their assignments back into the system. There's nothing wrong with that.

5.4. Reduction in enthusiasm and motivation

In describing online engagement fatigue, seven of the students associated it with a loss of motivation. All these students described feeling enthusiastic at the beginning of the semester when everything is novel and new but then losing motivation as the semester progressed. As articulated by Amelia (nursing student), "*you reach the point where you go, I think I've had enough of this*". Many of the students could not articulate why this loss of motivation occurred but suggested it could have been a result of feeling 'stretched' by the amount of work they had to do and thus feeling online engagement fatigue.

While the educators all spoke about strategies they use to motivate students (e.g., designing authentic learning tasks and activities [mentioned by 8 educators]; being responsive to students [35]; and providing support to struggling students [50]), there was no consensus about how things could be done better to prevent students from disengaging or losing motivation. Ideas suggested by individual educators (note, there were no consistent ideas) included improving university communication strategies so students are not overloaded with repeat messaging, improving transition to university (and hence student expectation management) through preparation programs, and trialling a block teaching approach where students focus on one course/subject at a time before moving onto the next one.

6. Discussion

The purpose of this research was to define online engagement fatigue by addressing the educational problem of whether online engagement fatigue is real; and commenced with developing an emergent definition. The first research question sought to confirm the existence of online engagement fatigue using the emergent definition and the qualitative evidence shows that from a student perspective, online engagement fatigue is a recognisable phenomenon. The second research question sought to understand the defining elements of online engagement fatigue and consider its effect. While the aspects developed in the emergent definition were present, the evidence points to a refinement in the definition. This is depicted in Diagram 2.

There were two aspects to the emergent definition developed in the first phase of the research process. The first aspect was the overexposure to course activities and related activities, and the second aspect was its consequence – that of a resultant loss of motivation and enthusiasm. According to the students, their experience of feeling overwhelmed is a

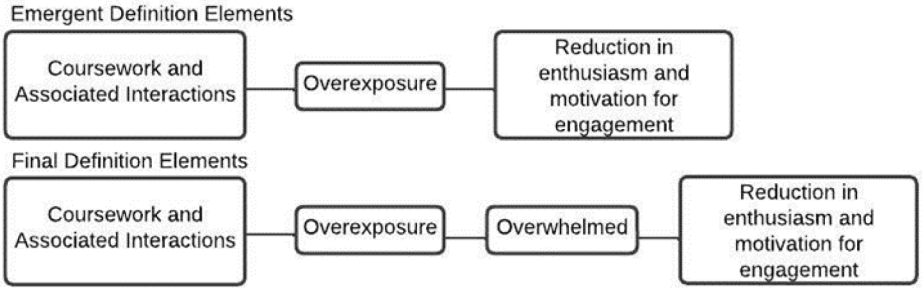


Diagram 2. Elements of Online Engagement Fatigue.

key characteristic of online engagement fatigue and therefore needs to be incorporated into the definition. Interestingly, although the educators overall were less inclined than the students to view online engagement fatigue as a phenomenon, many related observations of students being overwhelmed and shared strategies they used to overcome this, thus indicating that both students and educators agree that this is a risk factor for disengagement. Being overwhelmed is different to being overexposed. Overexposure is the repeated contact with little respite that may eventually lead to being overwhelmed. Overexposure could be managed with student psychological capital [18] to mitigate becoming overwhelmed; however, those students who became overwhelmed attached to it a strong emotional effect. This led to the final definition for online engagement fatigue:

> *A reduction in online students' enthusiasm and motivation for engaging in course activities as a result of overexposure to online coursework and associated interactions that lead to strong feelings of being overwhelmed.*

The findings here also suggest that some students may be discerning about the types of online artifacts and interactions they engage with, while others are not. Other researchers also report that some students learn to be selective with, or economise, learning resources [57]. With the increase in students juggling dual demands of work, study, and other responsibilities [18], the ability for students to be discerning in how much of the online material they actively engage with may help prevent online engagement fatigue. In a report by KPMG [1], it was suggested that in the future, as the higher education sector is increasingly disrupted by contemporary forces (technology change, demographic change, consumerism, competition), more students will be working part-time, undertaking family responsibilities, and wanting to integrate learning into their lives rather than interrupt their lives for an immersive university experience. This means the potential impact of these outside influences will increasingly become important considerations in online course design, as will be the student skill of discernment in navigating the institution's learning management system.

The findings also suggest that the timing of course activities and assessment was seen as a trigger for online engagement fatigue, as is the quantum of online communication. There are clear implications for appropriate online course design given that in online learning contexts the content and communication generated may increase rapidly, thus increasing pressure on students. This supports previous research warning of the use of technology to encourage student engagement [2, 28–31]. This means that as every educator in every course increasingly communicates through written, online channels – and as students respond via their own 'posts' (i.e., texts and messages) – the volume of content and associated messages that students are required to read in their courses correspondingly increases. As such, it is not surprising that students are feeling overwhelmed. As articulated by educator Simon, the consequence is "a *standard fight or flight response … where some people just* [say] *it's too much trouble, and they turn off*". Further, the increased online activity from every section of the university wanting to engage with students and the pressure for educators to conform to 'online study desk' minimal requirements could lead to behavioural change in the teaching team at the course level. This research contributes to this literature by uncovering the existence of online engagement fatigue in the perception of students and suggesting that further research is needed. In particular, what are educators currently doing to reduce the risk of online engagement fatigue in students; is facilitation and content changing as a result; and are educators themselves at risk of online engagement fatigue?

Emerging from the exploratory findings are several factors that improve our understanding of antecedents and consequences of online engagement fatigue and point to future research. This is presented in Diagram 3. Care should be taken in interpreting the diagram in relation to the direction of influence. Whilst arrows have been included to help understanding, it is acknowledged that the direction of influence could be bi-directional. For example, being overwhelmed could be an antecedent and a consequence of online engagement fatigue. Future research could explore these relationships as the inherent limitation in the research design of this project, exploratory qualitative analysis, reduces any claim to causation between factors.

There are several limitations of this study. Firstly, the students were first year students from regional, remote, or rural areas. While this restricts generalisability, it did provide a relatively homogenous group within which to examine the concept. Future research can seek to investigate whether the risk of online engagement fatigue changes as they progress through the years of their degree, and whether there are remedies being employed by more experienced students. Secondly, the opportunity sampling used in this research (students recruited via a general call through text messages and emails and educators approached by other academics) means that we cannot claim that the samples were representative of the wider body of students and educators. While this does not invalidate the findings, it needs to be acknowledged that these cohorts and their experiences may be different from their counterparts at other universities. Further research is needed to determine to what extent these findings are transferable to other cohorts. A broader study could further develop the concept of online engagement fatigue and the factors contributing to it, perhaps enabling the concept to be operationalised on a wider scale for future research. Finally, the normal caveat of researcher bias needs to be acknowledged. The researchers came to the project with their own experiences and understanding of student engagement that could influence their interpretations.

It was interesting to find that educators and students had conflicting perceptions and expectations concerning online engagement and what they consider 'enough' engagement. While several educators felt that perhaps students were just not engaging enough with online materials, the students strongly agreed that online engagement fatigue was a real phenomenon and something they regularly experienced. This leads to a likely expectation gap which could benefit from future research. This is especially important given that many higher education institutions are expanding online education and, in some cases, the move online that occurred because of the pandemic has intensified the integration of digital technologies and online education experiences for the long term [3]. This study emphasises some areas that those educators new to online learning should consider, such as ensuring suitable activities and

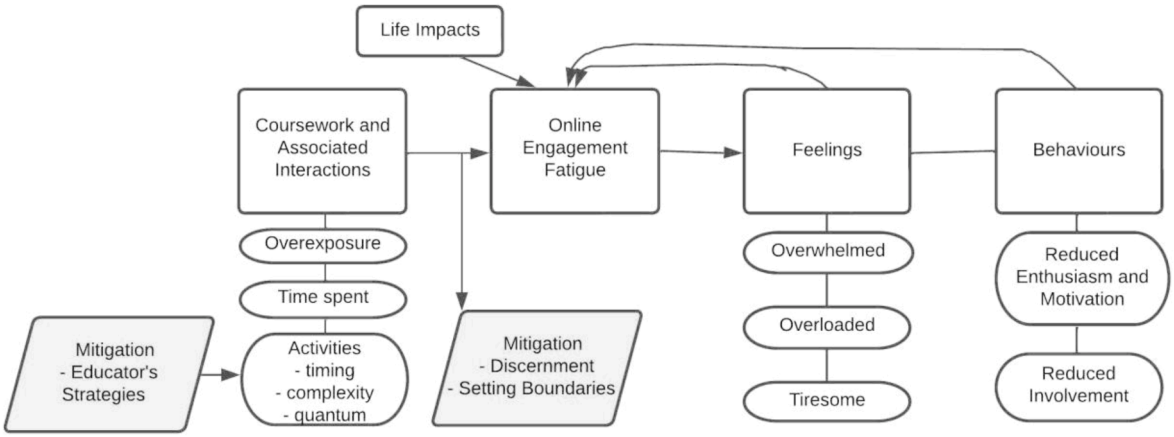


Diagram 3. Map of Elements and Factors of Online Engagement Fatigue.

facilitation to support student engagement while also ensuring online engagement fatigue is reduced. Future research could explore how the learning management system design and content impacts on students' propensity to experience online engagement fatigue.

As the emphasis on student engagement and observable engagement outcomes is currently high on the agenda in higher education, it is timely to consider whether students are being provided with too much learning content, whether educator expectations are too concerned with what is visible online, and whether an emphasis on course learning analytics is realistic? Alternatively, perhaps students are simply getting on with their work offline, rather than disengaging with the online content. Future research could examine student strategies for coping with the increasing pressure to visibly engage online, including the extent to which they are able to discern between the various demands and become more selective about when and with which to engage.

If, as indicated by the students in this research, online engagement fatigue is a phenomenon that students experience because of the way online learning content is structured and the time they are required to spend online, then this is important information for educators. Recognition of this phenomenon has potential implications for the ways educators design their online learning environment, perceive student engagement/non-engagement and the ways they communicate with students. It also has implications for how support is designed for students in the online learning context. These are important areas for future research into online pedagogies that support student engagement and success.

7. Conclusion

The student participants in this research almost unanimously agreed that online engagement fatigue exists and is regularly experienced by them. Exploring this phenomenon with both students and educators led to a clearer understanding of its elements that, in turn, refined the original definition. This refined definition more clearly illustrates the relationship between overexposure to online content/activities and feelings of becoming overwhelmed, which in turn can result in student disengagement. One key implication of this research is that higher education institutions need to consider the demands of online courses, how to reduce online overexposure and help students to avoid becoming overwhelmed by online demands. Further research is needed to explore possibilities for more easily recognising online engagement fatigue, the relationship between its antecedents and consequences, and taking action to reduce its impact on student motivation, involvement and success. This could include developing a scale measure to capture the dimensions of online engagement fatigue and exploring the student-educator expectation gap with respect to online engagement.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by a Higher Education Participation and Partnerships Program (HEPPP) Grant (https://www.dese.gov.au/heppp).

Appendix 1. Participant Information

Student Interviewees							
Pseudonym	M/F	Degree Program	Age	Lives with	Paid Work	Unpaid Work	Study Load
Jenny	F	Creative Arts	26	Alone	20 hrs pw	1 day per week	F/T
Gary	M	Nursing	53	Elderly parents	Nil	Daily	F/T
Amelia	F	Nursing	55	Husband & children	30 pw	3 h per week	F/T
Mark	M	Health Science	19	Parents & siblings	Nil		F/T
Olivia	F	Nursing	37	Alone	30 hrs pw	2 hrs per week	F/T
Sophie	F	Nursing	20	Sister (older)	20 hrs pw	Nil	F/T
Liam	M	Health Sciences	19	Parents	12–20 hrs pw	Yes	F/T
Bella	F	Nursing	22	Aunt & her family	20 hrs pw	1 day per week & some evenings	F/T

(*continued on next page*)

(*continued*)

Student Interviewees							
Pseudonym	M/F	Degree Program	Age	Lives with	Paid Work	Unpaid Work	Study Load
Lesley	M	Health Sciences	25	Friends	Nil	Nil	F/T
Mia	F	Creative Arts	62	Alone	Nil	Nil	F/T
Charlotte	F	Aviation	19	Parents	20–25 hrs pw	Nil	F/T
Harry	M	Business, Commerce, Law	23	Share house	25–30 hrs pw	Nil	F/T
Leah	F	Education	22	Mother & grandmother	25 hrs pw	Nil	F/T
Claudia	F	Business, Commerce, Law	20	Share house	10 hrs pw	Nil	F/T
Emily	F	Education	19	Friends	12 hrs pw	2.5 hrs per week	F/T
Isobel	F	Education	18	On campus	5–15 hrs pw	Nil	F/T
Eve	F	Education	18	Friends	20 hrs pw	Nil	F/T
James	M	Health Sciences	18	on campus	Nil	Nil	F/T

Educator Interviewees							
Pseudonym	**M/F**	**Discipline Area**	**Higher Ed. Teaching Experience (yrs)**	**Experience Across Universities (number)**	**Mode of Teaching**	**No. Students per Study Period**	**No. Courses per Study Period**
Jonathan	M	Law	30	2	Online	120	2
William	M	Business	20	1	Online and on-campus	200	3
Matthew	M	Education	15	yes (15 years as primary school principal before)	Online and on-campus	700	3
Rosemary	F	Nursing & Midwifery	13	1 (TAFE teaching previously)	Mostly online, some on-campus	500	3
Elizabeth	F	Accounting	10	3	Online and on-campus	200	4
Geoffrey	M	Commerce	25	2	Mostly	300	3
Simon	M	Engineering	25	1	Online and on-campus	80	2
Sam	M	Computing	5	1	Mostly online, some on-campus	150	2
Amber	F	Food Science	15–20	2	Mostly on-camps, some online	300	9
Bob	M	Sport & Exercise Science	6	1	Online and on-campus	130	2

Appendix 2. Interview Schedule

Student Questions

Demographic, personal circumstances & general engagement questions:

1. What is your age?
2. Who lives with you in your household? (e.g. partner, children, any others; relationships and ages of any children)
3. Do you live in a city, regional town or a rural environment?
4. Are you employed in any paid work? How many hours per week?
5. Are you employed in any unpaid work (e.g. volunteer work and/or unpaid caring work outside the immediate household, such as caring for parents)? How many hours per week?
6. Are you studying full- or part-time? (if part-time, what load?)
7. As an online student, what does 'being an engaged student' mean to you? How would describe the behaviour of an 'engaged' online student?
8. To what extent do you feel that you are an 'engaged' online student? – please describe.

9a. What types of things within your studies help or hinder your online engagement (e.g. teacher behaviour, study load, content, delivery, etc.)

9b. What types of things within your studies help or hinder your online engagement (e.g. teacher behaviour, study load, content, delivery, etc.)

10a. What types of things outside of your studies help or hinder your online engagement (e.g. work, family, paid/unpaid work etc.)

10b. What types of things outside of your studies help or hinder your online engagement (e.g. work, family, paid/unpaid work etc.)

CVExtra1: This was your first year at university and it may have been different than you had expected due to covid-19. Looking back how do you see your year compared to what you thought it would be?

CVExtra2: Assuming that all goes back to normal, how do you envision 2021? What do you see in your mind's eye for your studies next year? Much the same, better, worse?

Engagement fatigue questions:

We would like to talk with you now about something that we are calling 'online engagement fatigue'. What we mean by this is that we have noticed that at different times of the semester, some students seem to be less engaged, e.g. participating less in discussions, fewer posts, less interaction. One of the things that we are wanting to explore in this research is the notion of 'engagement fatigue' – i.e. whether online students ever find that they get very tired of being engaged or maybe showing engagement. We wonder if they are simply getting on with their work; or are they are finding it very hard to keep going? If so, we wonder what are the types of things that are contributing to engagement fatigue - making it very hard for them to keep going. I would be very interested in hearing your thoughts about this. E.g.

11. Do you think that you ever get tired of engaging with your studies online? If so, could you please tell me a little more about this?

If yes,

12. How frequently have you experienced 'online engagement fatigue'?
13. Does it tend to happen at any particular times of the semester more than others?
14. What tends to contribute to or trigger your online engagement fatigue?
14. What other circumstances – either within your studies or external to these - contribute to this?
15. How do you tend to respond to online engagement fatigue? E.g. not engage at all with your studies? Work on assessments instead of engaging with teacher/other students/ other activities?
16. What impact does this have on your studies more broadly?
17. Is there anything that helps to diminish online engagement fatigue? Something you do? Others do?
18. Is there anything that you think the university – your teachers or others – could do to help students avoid/diminish online engagement fatigue? (What activities do you find most useful in your studies; what do you find least useful?)
19. Any other general thoughts on what we have discussed?

If no,

11. How do you manage to keep yourself energised with your studies? What helps/hinders this for you?
12. What other circumstances – either within your studies or external to these – help you to maintain your energy/motivation for your studies?
13. Is there anything that others do – within university or outside of it – that contribute to this?
14. Any other general thoughts on what we have discussed?

Educator Questions

1. Gender?
2. Discipline/Faculty?
3. Years teaching experience?
4. At what institutions?
5. Do you teach only online or face to face and online?
6. How many students do you teach at present and across how many courses?

For your undergraduate courses…

7. Are you the only academic involved in the course/s or is there a larger teaching team?

Online Engagement fatigue questions:

Within this research we are wanting to explore the notion of 'online engagement fatigue' – defined as:

A reduction in the enthusiasm and motivation of online students for engaging in course activities as a result of overexposure to online coursework and associated interactions.

We are exploring whether online students become fatigued by the demands of the online environment and, if so, how is this related to their engagement. For example, at different times of the semester, some students do become less engaged, e.g. participating less in discussions, fewer posts, less interaction. We wonder if they are simply getting on with their work; or are they becoming fatigued by too much emphasis/expectations on and about online interactions? Or are finding it very hard to keep going for other reasons? I would be very interested in hearing your thoughts about this.

8. Have you witnessed a drop-off in student online engagement?

If yes,

9. Does it occur at a particular point in the semester?
10. Does it tend to happen across the whole student cohort or just a subsection of students?
11. To what extent do you think that this may be related to online engagement fatigue?
12. Do you notice it in one course and not another?

If yes,

13. Why do you think that is so? (type of course, number of activities, type of students, etc.)
14. What circumstances – either within the course or external to the course – contribute to this?
15. How do you respond to a drop in student online engagement?
16. Do you think a drop in student online engagement affects a student's outcome?

If so, how?

17. Is there anything that helps to diminish student online engagement fatigue? Something you do? Others do? Have you witnessed an intervention or action or behaviour of others?
18. Any other general thoughts.

References

[1] KPMG. The future of higher education in a disruptive world. London: KPMG International; 2020. Retrieved from, https://assets.kpmg/content/dam/kpmg/xx/pdf/2020/10/future-of-higher-education.pdf.

[2] Bond M, Buntins K, Bedenlier S, Zawacki-Richter O, Kerres M. Mapping research in student engagement and educational technology in higher education: a systematic evidence map. Int J Educ Technol High Educ 2020;17(2):1–30. https://doi.org/10.1186/s41239-019-0176-8.

[3] OECD. *OECD digital education outlook 2021: pushing the frontiers with artificial intelligence, blockchain and robots*. Paris: OECD Publishing; 2021. https://doi.org/10.1787/7fbfff45-en.

[4] Atherton M, Shah M, Vazquez J, Griffiths Z, Jackson B, Burgess C. Using learning analytics to assess student engagement and academic outcomes in open access enabling programmes. Open Learn: J Open, Distance e-Learn 2017;32(2):119–36. https://doi.org/10.1080/02680513.2017.1309646.

[5] Beer C, Clark K, Jones D. *Indicators of engagement.* Paper presented at the Curriculum, technology and transformation for an unknown future. In: Proceedings ASCILITE Sydney. 2010; 2010.

[6] Fincham E, Whitelock-Wainwright A, Kovanović V, Joksimović S, van Staalduinen J-P, Gašević D. Counting clicks is not enough: validating a theorized model of engagement in learning analytics. In: LAK19: Proceedings of the 9th International Conference on Learning Analytics & Knowledge. New York, NY: Association for Computing Machinery; 2019. p. 501–10. https://doi.org/10.1145/3303772.3303775.

[7] Dyment JJ, Stone C, Milthorpe N. Beyond busy work: rethinking the measurement of online student engagement. High Educ Res Develop 2020;39(7):1440–53. https://doi.org/10.1080/07294360.2020.1732879.

[8] Chen PSD, Lambert AD, Guidry KR. Engaging online learners: the impact of Web-based learning technology on college student engagement. Comput Educ 2010;54(4):1222–32. https://doi.org/10.1016/j.compedu.2009.11.008.

[9] Green RA, Whitburn LY, Zacharias A, Byrne G, Hughes DL. The relationship between student engagement with online content and achievement in a blended learning anatomy course. Anat Sci Educ 2018;11(5):471–7. https://doi.org/10.1002/ase.1761.

[10] Hampton D, Pearce PF. Student engagement in online nursing courses. Nurse Educ 2016;41(6):294–8. https://doi.org/10.1097/NNE.0000000000000275.

[11] Muir T, Milthorpe N, Stone C, Dyment J, Freeman E, Hopwood B. Chronicling engagement: students' experience of online learning over time. Distance Educ 2019;40(2):262–77. https://doi.org/10.1080/01587919.2019.1600367.
[12] Redmond P, Abawi LA, Brown A, Henderson R, Heffernan A. An online engagement framework for higher education. Online Learn 2018;22(1):183–204. https://doi.org/10.24059/olj.v22i1.1175.
[13] Lashari TA, Alias M, Kesot MJ, Abidin Z. An affective-cognitive teaching and learning approach for enhanced behavioural Engagements among engineering students. Eng Educ 2013;8(2):65–78. https://doi.org/10.11120/ened.2013.00011.
[14] Flynn D. Baccalaureate attainment of college students at 4-year institutions as a function of student engagement behaviours: social and academic student engagement behaviours matter. Res High Educ 2014;55(5):467–93. https://doi.org/10.1007/s11162-013-9321-8.
[15] Kuh GD, Cruce TM, Shoup R, Kinzie J, Gonyea RM. Unmasking the effects of student engagement on first-year college grades and persistence. J High Educ 2008;79(5):540–63. http://www.jstor.org/stable/25144692.
[16] Lee J-S. The relationship between student engagement and academic performance: is it a myth or reality? J Educ Res 2014;107(3):177–85. https://doi.org/10.1080/00220671.2013.807491.
[17] Trowler V. Student engagement literature review. York: The Higher Education Academy; 2010. Retrieved from, https://www.heacademy.ac.uk/system/files/StudentEngagementLiteratureReview_1.pdf.
[18] Barratt JM, Duran F. Does psychological capital and social support impact engagement and burnout in online distance learning students? Internet High Educ 2021;51. https://doi.org/10.1016/j.iheduc.2021.100821.
[19] Crampton A, Ragusa AT, Cavanagh H. Cross-discipline investigation of the relationship between academic performance and online resource access by distance education students. Res Learn Technol 2012;20(1):1–14. https://doi.org/10.3402/rlt.v20i0/14430.
[20] Higher Education Standards Panel. Final report: improving retention, completion and success in higher education. higher education standards panel. Canberra: Australian Government, Department of Education and Training; 2017. Retrieved from, https://docs.education.gov.au/system/files/doc/other/final_report_for_publishing.pdf.
[21] Ma J, Han X, Yang J, Cheng J. Examining the necessary condition for engagement in an online learning environment based on learning analytics approach: the role of the instructor. Internet High Educ 2015;24(26–34). https://doi.org/10.1016/j.iheduc.2014.09.005.
[22] Pittaway SM, Moss T. Initially, we were just names on a computer screen": designing engagement in online teacher education. Austr J Teach Educ 2014;39(7):140–56. https://doi.org/10.14221/ajte.2014v39n7.10.
[23] Martin F, Bolliger DU. Engagement matters: student perceptions on the importance of engagement strategies in the online learning environment. Online Learn 2018;22(1):205–22. https://doi.org/10.24059/olj.v22i1.1092.
[24] Alrashidi O, Phan HP, Ngu BH. Academic engagement: an overview of Its definitions, dimensions, and major conceptualisations. Int Educ Stud 2016;9(12):41–52. https://doi.org/10.5539/ies.v9n12p41.
[25] Fredricks JA, Filsecker M, Lawson MA. Student engagement, context, and adjustment: addressing definitional, measurement, and methodological issues. Learn Instr 2016;43:1–4. https://doi.org/10.1016/j.learninstruc.2016.02.002.
[26] Pérez-López R, Gurrea-Sarasa R, Herrando C, Martín-De Hoyos MJ, Bordonaba-Juste V, Utrillas-Acerete A. The generation of student engagement as a cognition-affect-behaviour process in a Twitter learning experience. Austr J Educ Technol 2020;36(3):132–46. https://doi.org/10.14742/ajet.5751.
[27] Boekaerts M. Engagement as an inherent aspect of the learning process. Learn Instr 2016;46:76–83. https://doi.org/10.1016/j.learninstruc.2016.02.001.
[28] Buck S. In their own voices: study habits of distance education students. J Libr Inform Serv Distance Learn 2016;10(3–4):137–73. https://doi.org/10.1080/1533290X.2016.1206781.
[29] Howard SK, Ma J, Yang J. Student rules: exploring patterns of students' computer-efficacy and engagement with digital technologies in learning. Comput Educ 2016;101:29–42. https://doi.org/10.1016/j.compedu.2016.05.008.
[30] Stone C. Online learning in Australian higher education: opportunities, challenges and transformations. Stud Succ 2019;10(2):1–11. https://doi.org/10.5204/ssj.v10i2.1299.
[31] Stone C. Opportunity through online learning: improving student access, participation and success in higher education. equity fellowship final report. Perth: Curtin University, National Centre for Student Equity in Higher Education; 2017. Retrieved from, https://www.ncsehe.edu.au/publications/opportunity-online-learning-improving-studentaccess-participation-success-higher-education/.
[32] Hewson ERF. Students' emotional engagement, motivation and behaviour over the life of an online course: reflections on two market research case studies. J Interact Med Educ 2018;1(10):1–13. https://doi.org/10.5334/jime.472.
[33] Sweller J. Cognitive load during problem solving: effects on learning. Cogn Sci 1988;12(2):257–85. https://doi.org/10.1016/0364-0213(88)90023-7.
[34] Nastaran Z. The important elements of LMS design that affect user engagement with e-learning tools within LMSs in the higher education sector. Austr J Educ Technol 2017;33(1). https://doi.org/10.14742/ajet.2938.
[35] Bitsika V, Sharpley CF, Rubenstein V. What stresses university students: an interview investigation of the demands of tertiary studies. J Psycholog Counsell Sch 2010;20(1):41–54. https://doi.org/10.1375/ajgc.20.1.41.
[36] Kizhakkeveettil A, Vosko AM, Brash M, Philips MA. Perceived stress and fatigue among students in a doctor of chiropractic training program. J Chiropract Educ 2017;31(1):8–13. https://doi.org/10.7899/JCE-15-27.
[37] Khojasteh F, Marziyeh A, Pourghaz A. Predicting fatigue in university students via subtypes of perfectionism. Int J Psychol 2019;13(2):8–42. https://doi.org/10.22034/IJPB.2018.115467.
[38] Lian P, Sun Y, Ji Z, Li H, Peng J. Moving away from exhaustion: how core self-evaluations influence academic burnout. PLoS One 2014;9(1):e87152. https://doi.org/10.1371/journal.pone.0087152.
[39] de Vries JD, van Hooff MLM, Geurts SAE, Kompier MAJ. Exercise as an intervention to reduce study-related fatigue among university students: a two-arm parallel randomized controlled trial. PLoS One 2016;11(3):1–21. https://doi.org/10.1371/journal.pone.0152137.
[40] Smith AP. Cognitive fatigue and the wellbeing and academic attainment of university students. J Educ Soc Behav Sci 2018;24(2):1–12. https://doi.org/10.9734/JESBS/2018/39529.
[41] Wang S. An analysis of the causes and countermeasures of students' English earning weariness in junior high school. Theory Pract Lang Stud 2015;5(8):1747–53. https://doi.org/10.2991/snce-17.2017.193.
[42] Barkan SE. Health, illness, and society: an introduction to medical sociology. Lanham, MD: Rowan and Littlefield; 2017.
[43] Laberge L, Ledoux É, Auclair J, Thuilier C, Gaudreault M, Gaudreault M, Perron M. Risk factors for work-related fatigue in students with school-year employment. J Adolesc Health 2011;48(3):289–94. https://doi.org/10.1016/j.jadohealth.2010.07.003.
[44] Shapiro TM, Williams KM. The causal effect of the school day schedule on adolescents' academic achievement. In: Paper presented at the SREE Spring 2015 Conference; 2015. Retrieved from: https://files.eric.ed.gov/fulltext/ED562479.pdf.
[45] Okano K, Kaczmarzyk JR, Dave N, Gabrieli JDE, Grossman JC. Sleep quality, duration, and consistency are associated with better academic performance in college students. npj Sci Learn 2019;4(16):1–5. https://doi.org/10.1038/s41539-019-0055-z.
[46] Mallon R. A field guide to social construction. Philos Compass 2006;2(1):93–108. https://doi.org/10.1111/j.1747-9991.2006.00051.x.
[47] Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. Health Serv Res 2007;42(4):1758–72. https://doi.org/10.1111/j.1475-6773.2006.00684.x.
[48] Leech NL, Onwuegbuzie AJ. Beyond constant comparison qualitative data analysis: using NVivo. Sch Psychol Q 2011;26(1):70–84. https://doi.org/10.1037/a0022711.
[49] Instructure, & Hanover Research. State of student success and engagement in higher education: 2020 global research study and trends. Utah: Retrieved from Salt Lake City; 2020. https://www.instructure.com/canvas/resources/insights/student-success-engagement-higher-education-research-study-trends#main-content.
[50] Berg BL. Qualitative research methods for the social sciences. 7th edition. Boston: Pearson; 2009.
[51] Fontana A, Frey JH. 'The interview: from neutral stance to political involvement. In: Lincoln NK, Denzin YS, editors. Collecting and interpreting qualitative materials. 3rd edition. London: Sage; 2008.
[52] Clarke V, Braun V. Teaching thematic analysis: overcoming challenges and developing strategies for effective learning. Psychologist 2013;26(2):120–3. Retrieved from, https://thepsychologist.bps.org.uk/volume-26/edition-2/methods-teaching-thematic-analysis.
[53] Maguire M, Delahunt B. Doing a thematic analysis: a practical, step-by-step guide for learning and teaching scholars. Ireland J Teach Learn High Educ 2017;9(3):3351. Retrieved from, https://ojs.aishe.org/index.php/aishe-j/article/view/335.
[54] Braun V, Clarke J. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77–101. https://doi.org/10.1191/1478088706qp063oa.
[55] Welsh E. Dealing with data: using nvivo in the qualitative data analysis process. Forum: Qual Soc Res 2002;3(2):26. https://doi.org/10.17169/fqs-3.2.865.
[56] Maher C, Hadfield M, Hutchings M, De Eyto A. Ensuring rigor in qualitative data analysis: a design research approach to coding combining nvivo with traditional material methods. Int J Qual Methods 2018;17:1–3. https://doi.org/10.1177/1609406918786362.
[57] Lackovic N, Kerry R, Lowe R, Lowe T. Being knowledge, power and profession subordinates: students' perceptions of Twitter for learning. The Internet and High Educ 2017;33:41–8. https://doi.org/10.1016/j.iheduc.2016.12.002.

Engineering Applications of Artificial Intelligence 117 (2023) 105560

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

An adaptive teacher–student learning algorithm with decomposed knowledge distillation for on-edge intelligence

Majid Sepahvand [a], Fardin Abdali-Mohammadi [a,*], Amir Taherkordi [b]

[a] Department of Computer Engineering and Information Technology, Razi University, Kermanshah, Iran
[b] Department of Informatics (IFI), University of Oslo, Oslo, Norway

ARTICLE INFO

Keywords:
Deep learning
Knowledge distillation
On-edge intelligence
Feature representation
Tensor decomposition

ABSTRACT

In case the spatial shape of the feature maps of the teacher in feature-based knowledge distillation (KD) is significantly greater than the student model, first, they cannot be compared directly. Second, the knowledge of these complex feature maps cannot be quite apprehensible for the student. This paper proposed a new KD, in which Tucker decomposition was used to decompose the large-dimension feature maps of a teacher to obtain core tensors from the feature maps of the teacher. The knowledge of these tensors can be easily understood by students due to their low complexity. Furthermore, in the proposed KD, an adaptor function is suggested, which balances the spatial shape of the core tensors of the teacher and student and helps compare them using a convolution regressor. Finally, a hybrid loss based on adaptor function is suggested to distill the knowledge of the core tensors of the teacher to the student. Both teacher and student models were implemented on smartphones used as edge devices, and the experiments were evaluated in terms of recognition rate and complexity. According to the results, the student model designed by ResNet-18 architecture has ~65.44 million fewer parameters, ~6.45 GFLOPs less computational complexity, ~1.12 G less GPU memory use, and ~265.67 times greater compression rate than its teacher model designed by ResNet-50 architecture. While the recognition rate of the student model merely dropped down to 1.5% in the benchmark dataset.

1. Introduction

In emerging technologies such as edge computing and the Internet of things (IoT), the addition of intelligence to the edge devices, a part of the computations of the central analysis servers are carried out on them (Zhu et al., 2021). These devices are connected directly to certain sensors (e.g., cameras, microphones, and medical sensors such as EEG and ECG) which provide a large amount of input data. Hence, they are considered appropriate targets for different artificial intelligence (AI) applications such as recognition and classification. The transfer of intelligence to edge devices resulted in specific advantages: (1) reducing response latency and cost of communication with a server, (2) decentralizing computation and increasing fault tolerance, and (3) developing smart edge devices and a centralized server with local and central decision-making features. Hence, large-scale intelligent tasks can be performed in a scalable manner (Teerapittayanon et al., 2017; Sepahvand and Abdali-Mohammadi, 2021).

Despite the above advantages in the transfer of intelligence to edge devices, they usually have low dimensions, low to moderate computing power, and limited storage capacity (Gupta et al., 2021). Thus, there are certain constraints on the execution of many deep learning models running on edge devices, such as CNN (Sohail et al., 2021), VGG (Karthiga and Narashimhan, 2021), AlexNet (Karthiga and Narashimhan, 2021), and ResNet (Reenadevi et al., 2021) with many deep layers. For instance, nearly 84.04 M parameters are created to train a VGG-16 model on a 256×256 image. If the 8-byte decimal floating point data type is employed to define parameters, the model size will be $84.04\ \text{M} \times 8\ \text{B} = 640$ MB, which is rather large (Fu et al., 2020). Thus, it is essential to employ methods to facilitate and reduce the computational complexity of deep learning models. In fact, researchers still face major unsolved challenges to the development of lightweight deep learning models with minimum computational complexity. Hence, it is necessary to propose devise appropriate solutions to the challenges introduced in the development of optimal learning models.

Many researchers have lately tried to develop lightweight deep learning models by proposing various methods such as model compression and acceleration (Cheng et al., 2018), parameter pruning and sharing (Wang et al., 2018), low-rank decomposition (Yu et al., 2017), transferred compact convolutional filters, and knowledge distillation (KD) (Gou et al., 2021). Certain models such as MobileNet (Gou et al., 2021) and MobileVGG (Sandler et al., 2018) have also been used. Among these methods, KD is a comprehensive method to supervise the training of a lightweight student using knowledge distilled from

* Corresponding author.
E-mail addresses: Sepahvand.majid@razi.ac.ir (M. Sepahvand), fardin.abdali@razi.ac.ir (F. Abdali-Mohammadi), amirhost@ifi.uio.no (A. Taherkordi).

https://doi.org/10.1016/j.engappai.2022.105560
Received 19 May 2022; Received in revised form 10 October 2022; Accepted 24 October 2022
Available online 7 November 2022
0952-1976/© 2022 Elsevier Ltd. All rights reserved.

the teacher. In fact, KD is adopted to bridge the architectural teacher–student gap so that a student model, despite its simplicity, can imitate a teacher's performance network (Hinton et al., 2015; Sepahvand and Abdali-Mohammadi, 2022; Pan et al., 2021). It is possible to obtain a kind of hint while distilling knowledge from an advanced teacher network to a student network in order to boost the student's performance. In recent years, various KD models based on feature maps of the layers of the teacher are introduced. The use of the middle layer of the teacher model was first suggested in a model called FitNets (Romero et al., 2015), to propose guides to train student networks. In this model, a guide represents the output of the middle layer of the teacher that supervises the student's learning. Different models were introduced inspired by this idea. In Shu et al. (2021), Changyong et al. acquired the attention map from the feature maps for knowledge expression. In Huang and Wang (2017), Huang et al. transferred the neuron selectivity to generalize the attention map. In Passalis and Tefas (2018), Passalis et al. distilling knowledge by adapting a probability distribution in the feature space. To simplify the process of knowledge distillation, they proposed factors in a more understandable form in the middle layers. Despite numerous efforts to propose feature-based KD, as stated in Gou et al. (2021) by Gou et al. in case the spatial shape of feature maps of teacher is significantly greater than the student, first, they cannot be compared directly. Second, the knowledge of these complex feature maps cannot be easily understood by the students. To resolve the stated challenge, Dai et al. (2021) and Ding et al. (2019), introduced a new method based on Tucker decomposition, which is the subcategory of the low-rank composition algorithm to decompose the teacher layers. However, error function of the response-based distillation in the Dai et al. (2021) and Ding et al. (2019) were used, which is merely related to the soft labels knowledge. Thus, the middle layers of teacher were not used to distill knowledge to students. Therefore, it is necessary to develop methods to decompose knowledge in teacher feature maps.

This paper proposed a new KD, in which Tucker decomposition was used to decompose the large-dimension feature maps of a teacher to obtain core tensors from the feature maps of the teacher. The knowledge of these tensors can be easily understood by students due to their low complexity. The research hypothesis pertains to students' better perception of the decomposed knowledge of an advanced teacher to resolve the feature-based KD challenges. Furthermore, in the proposed KD, an adaptor function is suggested, which balances the spatial shape of the core tensors of the teacher and student and helps compare them using a convolution regressor. Finally, a hybrid loss based on adaptor function is suggested to distill the knowledge of the core tensors of the teacher to the student. Both the teacher and student models were implemented on smartphones used as edge devices, and their performance outputs were evaluated in terms of recognition rate and complexity. According to the experimental results, the compression rate of the ResNet-18 based student model was nearly 265.67 times higher than that of its ResNet-50 based teacher model; however, its recognition rate declined by only $\sim 1.5\%$ on average in the benchmark dataset.

In brief, the main contributions of this paper are as follows.

- Introducing a novel method for the teacher–student KD through the idea of decomposing the tensors of feature maps in the inner layers of two models with the Tucker decomposition method.
- Providing a module of a flexible KD together with a regressor between the teacher and student models to resolve the incommensurability of feature maps in the inner layers of two models.
- Designing, implementing, and evaluating the lightweight deep learning method on a smartphone as used an edge device to add up to the intelligence features.

This paper consists of different sections. Section 2 reviews the literature on KD. The proposed KD-based method is then introduced in Sections 3 and 4. After that, Section 5 reports the experimental results. Finally, Section 6 draws a research conclusion.

2. Related work

There have been many attempts to compress, simplify, and lighten deep learning models for execution on devices with limited computing resources. Generally, these methods are classified as network pruning, low-rank factorization, parameter pruning and sharing, quantization, fuzzification (Versaci et al., 2022) and KD (Gou et al., 2021). KD is a comprehensive approach for monitoring the learning process of a simple student network through an advanced teacher network. In fact, KD aims to bridge the teacher–student architecture gap so that the student network can simply imitate a teacher's performance. When knowledge is distilled from an advanced teacher network to a student network, a type of hint can be acquired to improve the student network performance. In Hinton et al. (2015), Hinton et al. first proposed KD for a DNN, which defines the distribution of a teacher network features as knowledge in a label space.

Different approaches have been employed to enhance the KD performance. In Mirzadeh et al. (2020), an assistant adoption approach was employed to develop a teacher–student medium. This approach intends to better distillation an advanced teacher's knowledge through an assistant. Although this approach is interesting, it faces an excessive overhead due to the complexity caused by adding the assistant model. In a nearly similar approach in Bang et al. (2021), a professor model was used along with the teacher. The professor model was developed by a conditional adversarial autoencoder, and the KD strategy used the teacher's and the professor's knowledge simultaneously to distill knowledge to a student network. In fact, the professor model was employed to enhance the accuracy of soft responses given by the teacher model and to distill them to a student. The one-hot coded hard label can be utilized to only distill the corresponding information of classes to a student network, which cannot be given the other pieces of information such as relationships between different classes. Hence, soft labels are used in addition to hard labels in this kind of KD to train a student network. Many efforts have been made to improve the KD performance by using inner representations. In Romero et al. (2015), representations from the inner layers of a deep learning model were utilized to obtain hints and help train a simple deep student model. Afterward, a projection was adopted in the inner layers of a deep teacher model to obtain hints in Zhang et al. (2017), and a student model learned the projection at the same time. In Song et al. (2018), KD was integrated with an attention map for the distillation process. In Yang et al. (2018), another approach was introduced for distillation by using the knowledge existing in data instead of a distillation model. In Yang et al. (2018), the researchers reported that a tolerated teacher would be better than a high-accurate teacher for distillation. They also indicated that a teacher should have a smooth distribution on its predictions and that a high-accuracy teacher would not be appropriate with a spike distribution. An online KD approach was used in Lan et al. (2018) by allowing a teacher model to train a student model within the development process.

Another group of studies employed a method of decomposing the middle-layer information of a teacher model and obtaining the main information of those layers to use as hints while training a student. In Lee and Song (2020), an SVD was adopted to decompose the inner representation of a model, and a Euclidean distance criterion was utilized to shorten the gap between feature maps of a student model and those of its teacher model. For this purpose, 3D tensors of feature maps were converted into 2D tensors, to which an SVD was then applied. It is difficult and time-consuming to determine an accurate rank in the SVD algorithm, a process which is still performed approximately but not accurately, despite the introduction of least mean squares (LMS) algorithms. Recently, Dai et al. (2021) proposed a novel method based on the Tucker decomposition, which is the subset of low-rank decomposition algorithms, to improve the quality of KD and lower an advanced teacher's level. In fact, the Tucker decomposition can decompose N-order tensors in a teacher's middle layers into the product of a tensor

kernel with lower dimensions by a set of factor matrices of dimensions. This process intends to decompose knowledge and help a student better perceive the knowledge received from a decomposed teacher. Nevertheless, the method introduced in Dai et al. (2021) is response-based KD, which depends merely on the final-layer outputs, i.e., soft labels. Therefore, the inner-layer information of a teacher network was not used directly to distill knowledge to a student. Nonetheless, this information is very useful for representation learning by using a teacher network through the feature-based distillation, which is a more efficient method (Romero et al., 2015).

Despite many attempts in recent years to introduce various KD methods, Gou et al. (2021) reported in Gou et al. (2021) that the following challenges would still persist

- The main distinction of KD methods is to determine the knowledge that can show the information of a teacher network in the best way possible. Although KD methods are proposed constantly, most of the existing methods present the same direct features embedded in a teacher network. Thus, it is not true that they extract sufficient knowledge from a teacher network.
- In the teacher networks with much more sophisticated structures than that of a student network, the variance is high between spatial shapes and their middle-layer feature maps. Hence, it is difficult for a student to perceive a teacher's direct features in the form of a high-dimensional tensor.
- Knowledge depends on the spatial form of feature maps; hence there are limitations on the distillation of knowledge to student networks with different spatial shapes compared to the teacher network.

3. Knowledge distillation (KD)

There are two deep neural models called the teacher model and the student model in the standard KD. The former provides soft labels to train the latter (Hinton et al., 2015). From a mathematical point of view, the input matrix and its corresponding label are defined as $\mathbf{X}$ and $\mathbf{y}$, respectively. In fact, $\mathbf{y}$ denotes the hard label. If K is assumed to indicate the number of target classes, then $\mathbf{y}$ is a K-dimensional vector with the one-hot coding method. In addition, $\mathbf{q} = \{q_i\}_{i=1}^{K}$ is a vector that shows the soft labels acquired from the soft-max output layer in the teacher model, in which q_i refers to the probability of the ith class. Ultimately, z_i denotes the activation function output in the teacher's final layer for the ith class.

The relationship between q_i and z_i is described as below.

$$q_i = \frac{\exp\left(z_i/T\right)}{\sum_{i=1}^{K}\exp\left(z_i/T\right)} \tag{1}$$

Where T refers to the temperature. In usual classification applications, T is considered 1, whereas larger values are given to it in KD to obtain a softer probability distribution with further information than one-hot coding. As the temperature (T) increases, $\mathbf{q}$ approaches a normal distribution. The student model can be trained with both hard and soft labels (i.e., $\mathbf{y}$ and $\mathbf{q}$) by minimizing the following error function (Hinton et al., 2015).

$$L_{KD} = \frac{1}{R}\sum_{\mathbf{X},\mathbf{y}}\left((1-\gamma)\,L_{Hard}\left(\mathbf{X},\mathbf{y}\right) + \gamma L_{soft}\left(\mathbf{X},\mathbf{q}\right)\right) \tag{2}$$

Where γ denotes the weight coefficient employed to moderate the error based on soft and hard labels, whereas R indicates the number of input samples $\mathbf{X}$. The entropy error function used for hard and soft labels are defined respectively as below (Shen et al., 2020).

$$L_{Hard}\left(\mathbf{X},\mathbf{y}\right) = -\sum_{i=1}^{K} y_i \log p_i(\mathbf{X}) \tag{3}$$

and

$$L_{soft}\left(\mathbf{X},\mathbf{q}\right) = -\sum_{i=1}^{K} q_i \log p_i(\mathbf{X}) \tag{4}$$

Where $p_i\left(\mathbf{X}\right)$ represents the output probability of the student model for the ith class, and y_i denotes the value of the hard label ($\mathbf{y}$) corresponding to the ith class. Like Equation (1), $p_i(\mathbf{X})$ is calculated as below.

$$p_i = \frac{\exp\left(s_i/T\right)}{\sum_{i=1}^{K}\exp\left(s_i/T\right)} \tag{5}$$

Where s_i indicates the activation function output in the student's final layer for the ith class. Regarding the errors of hard labels, T was considered 1; however, it will be equal to the value of Eq. (1) for soft labels.

4. Proposed adaptive teacher–student learning via KD

In this paper, the KD model was employed along with adaptive joint learning to design a lightweight classifier. According to Fig. 1, pair of teacher–student models were utilized in ResNet architectures. The joint learning approach was employed in this paper to distill a teacher's knowledge in response layer (according to Eq. (2)) along with the feature maps of its middle layers to a student model. In the proposed method, the hybrid loss of proposed KD is defined as below.

$$L_{Hybrid} = \frac{1}{R}\sum_{\mathbf{X},\mathbf{y}}\left((1-\beta)\,L_{KD} + \beta L_{IF}\right) \tag{6}$$

In this equation, L_{KD} is calculated through Equation (2). Furthermore, the general error (L_{IF}) between the feature maps of middle layers in both teacher and student models are determined through the following equation.

$$L_{IF} = \mathrm{L}_2\left(\mathcal{F},\phi\left(\mathcal{H}\right)\right) = \left\|\mathcal{F} - \phi\left(\mathcal{H}\right)\right\|^2 \tag{7}$$

Where $\mathcal{F}$ and $\mathcal{H}$ represent the feature maps of convolution layers (according to Fig. 1) in teacher and student networks, respectively.

Considering its high complexity, the spatial shape of teacher feature maps, i.e., $\mathcal{F}$, had larger dimensions compared to the feature maps of the student, i.e., $\mathcal{H}$. Formal description is that if the feature maps of teacher are $\mathcal{F} \in \mathbb{R}^{I_1\times I_2\times I_3}$ and feature maps of student are $\mathcal{H} \in \mathbb{R}^{P_1\times P_2\times P_3}$. Then, the dimensions will demonstrate the length, width, and number of channels of feature maps, respectively. In the studies suggested for feature-based KD, it was assumed that $I_1 \times I_2 = P_1 \times P_2$ or $I_1 \times I_2 \cong P_1 \times P_2$. However, the challenge stated by Gou et al. in Gou et al. (2021) is that if $I_1 \times I_2 \gg P_1 \times P_2$, first, the direct comparison of $\mathcal{F}$ with $\mathcal{H}$ is impossible to distill another knowledge. Second, the knowledge of $\mathcal{F}$ feature maps (due to their great dimensions) cannot be easily understood by the student. Thus, in this paper, an adaptive function named ϕ is employed in Eq. (7), which adjusts the dimensions of $\mathcal{F}$ and $\mathcal{H}$ tensors decomposed by Tucker decomposition to achieve knowledge distillation. Section 4.1 first, elaborated on Tucker decomposition, then, Section 4.2 utilized it to describe the deep teacher and student models.

4.1. Tucker decomposition

A tensor is factorized into other tensors of smaller dimensions through Tucker decomposition. In fact, Tucker decomposition divides an N-dimensional tensor into a smaller matrix called the core tensor and a series of smaller tensors called the factor matrices corresponding to the division dimensions. Fig. 2 shows the decomposition result of a 3-order tensor ($\mathcal{X} \in \mathbb{R}^{I_1\times I_2\times I_3}$) in this method. Generally, there is $\mathcal{X} \in \mathbb{R}^{I_1\times I_2\times\cdots I_n\times\cdots I_N}$, in which $I_n\,(n = 1, 2\cdots N)$ refers to the nth order dimension, for an N-order tensor. Eq. (8) depicts the approximation of tensor $\mathcal{X}$ (Dai et al., 2021).

$$\mathcal{X} \cong \mathcal{E}\times_1\mathbf{U}^{(1)}\times_2\mathbf{U}^{(2)}\cdots\times_n\mathbf{U}^{(n)}\cdots\times_N\mathbf{U}^{(N)} = [\![\mathcal{E},\mathbf{U}^{(1)},\mathbf{U}^{(2)},\ldots,\mathbf{U}^{(n)},\ldots,\mathbf{U}^{(N)}]\!] \tag{8}$$

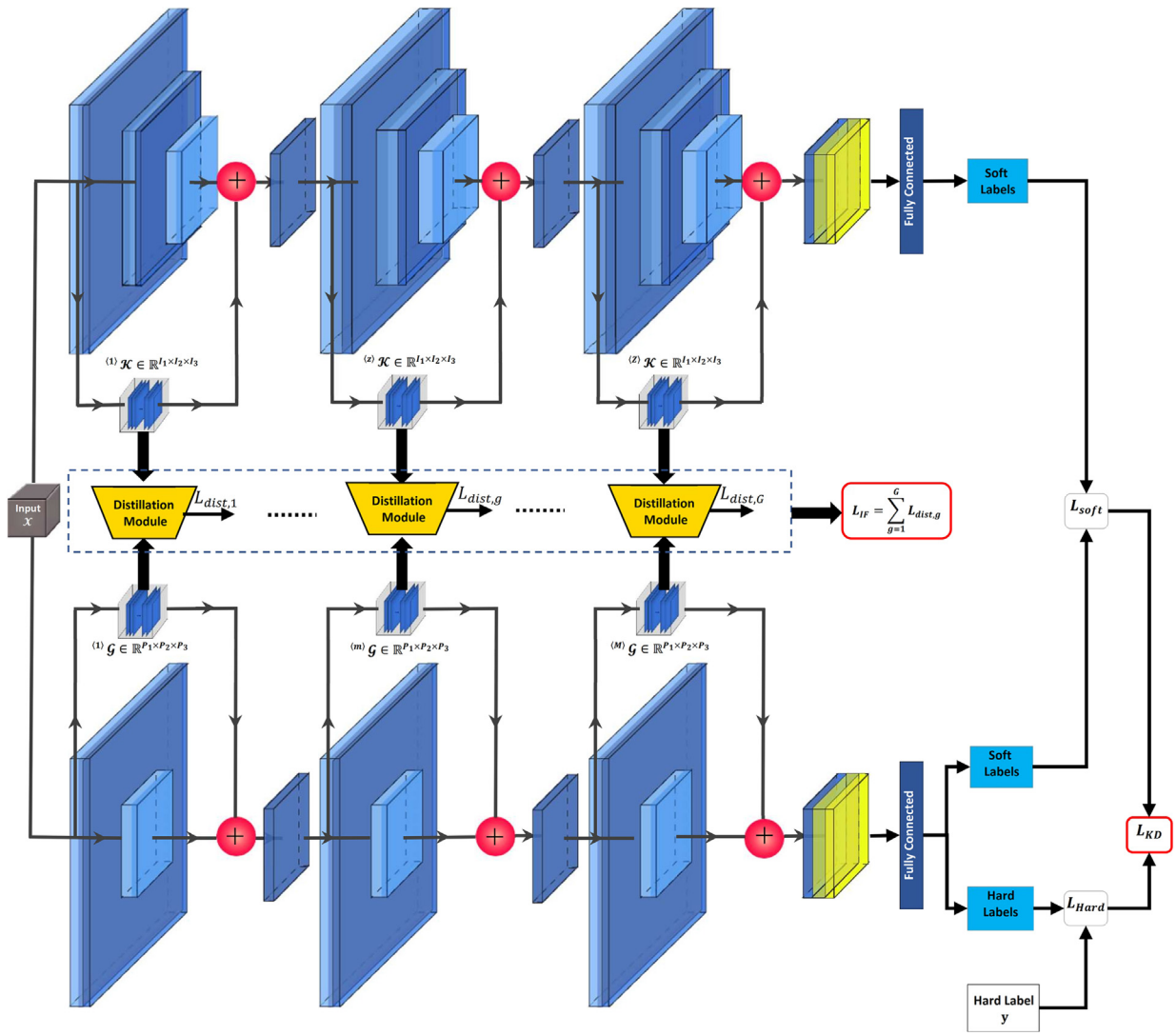


Fig. 1. The pair of teacher–student models in the proposed adaptive joint learning approach through KD. Teacher model has Z layers and student model has M layers, among which G modules are provided for the proposed knowledge distillation.

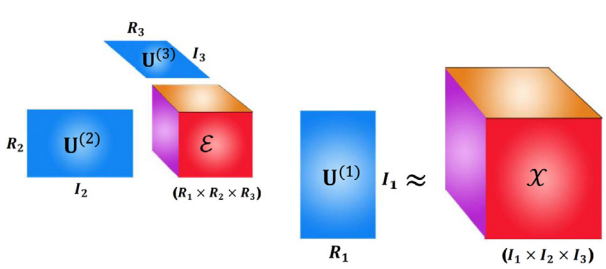


Fig. 2. Tucker decomposition for a 3D tensor. $\mathcal{X}$ input tensors with the dimensions of $I_1 \times I_2 \times I_3$ for decomposition, $\mathcal{E}$ core tensor with dimensions of $R_1 \times R_2 \times R_3$, $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$ factor matrices.

Where $\times_n (n = 1, 2 \cdots N)$ denotes the product of a core tensor ($\mathcal{E} \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_n \times \cdots I_N}$) and the corresponding nth order factor matrix ($\mathbf{U}^{(n)}$). This matrix is defined as $\mathbf{U}^{(n)} = \left[\mathbf{u}_1^{(n)}, \mathbf{u}_2^{(n)}, \ldots, \mathbf{u}_{R_n}^{(n)}\right] \in \mathbb{R}^{I_n \times R_n}$ $(n = 1, 2 \cdots N)$. Every item of Eq. (8) can evolve into Eq. (9).

$$\mathcal{X}_{i_1,\ldots,i_n} \cong \sum_{r_1=1}^{R_1} \cdots \sum_{r_n=1}^{R_N} \alpha_{r_1,\ldots,r_n} \mathbf{U}^{(1)}_{(i_1,r_1)} \cdots \mathbf{U}^{(N)}_{(i_N,r_N)} \tag{9}$$

4.2. Proposed KD module

Fig. 3 demonstrates the proposed distillation module. In this module, first, the tensors of teacher ${}^{\langle z \rangle}\mathcal{K}$ and student ${}^{\langle m \rangle}\mathcal{G}$ were decomposed using Tucker decomposition. Then, having employed the adaptor function, the dimensions of the core tensors obtained from the Tucker decomposition were equalized and their distances were calculated as the error between teacher and student. This module is explained herein below. $\mathcal{F}$ is defined for the teacher model in Eq. (7) and as follows for Eq. (10).

$$\mathcal{F} = {}^{\langle Z \rangle}f\left({}^{\langle Z \rangle}\mathcal{W}\,{}^{\langle Z-1 \rangle}f\left({}^{\langle Z-1 \rangle}\mathcal{W}\,{}^{\langle Z-2 \rangle}f\left({}^{\langle Z-2 \rangle}\mathcal{W}\left(\cdots\left({}^{\langle 1 \rangle}f\left({}^{\langle 1 \rangle}\mathcal{W}\mathbf{X} + {}^{\langle 1 \rangle}b\right)\right)\cdots\right) + {}^{\langle Z-2 \rangle}\right) + {}^{\langle Z-1 \rangle}b\right) + {}^{\langle Z \rangle}b\right) \tag{10}$$

In Eq. (10), ${}^{\langle z \rangle}\mathcal{W}$ and ${}^{\langle z \rangle}b$ represent the weight tensor and the bias of layer $z (z = 1, 2 \cdots Z)$, respectively. The Tucker decomposition can be employed to break down ${}^{\langle z \rangle}\mathcal{W}$ tensor to extract its core tensor and use it for KD operation. The following equation is the decomposed form of ${}^{\langle z \rangle}\mathcal{W}$ weight tensor through the Tucker decomposition.

$$\begin{aligned} {}^{\langle z \rangle}\mathcal{W} &\cong {}^{\langle z \rangle}\mathcal{Q} \times_1 {}^{\langle z \rangle}\mathbf{R}^{(1)} \times_2 {}^{\langle z \rangle}\mathbf{R}^{(2)} \cdots \times_l {}^{\langle z \rangle}\mathbf{R}^{(l)} \cdots \times_L {}^{\langle z \rangle}\mathbf{R}^{(L)} \\ &= [\![{}^{\langle z \rangle}\mathcal{Q}, {}^{\langle z \rangle}\mathbf{R}^{(1)}, {}^{\langle z \rangle}\mathbf{R}^{(2)}, \ldots, {}^{\langle z \rangle}\mathbf{R}^{(l)}, \ldots, {}^{\langle z \rangle}\mathbf{R}^{(L)}]\!] \end{aligned} \tag{11}$$

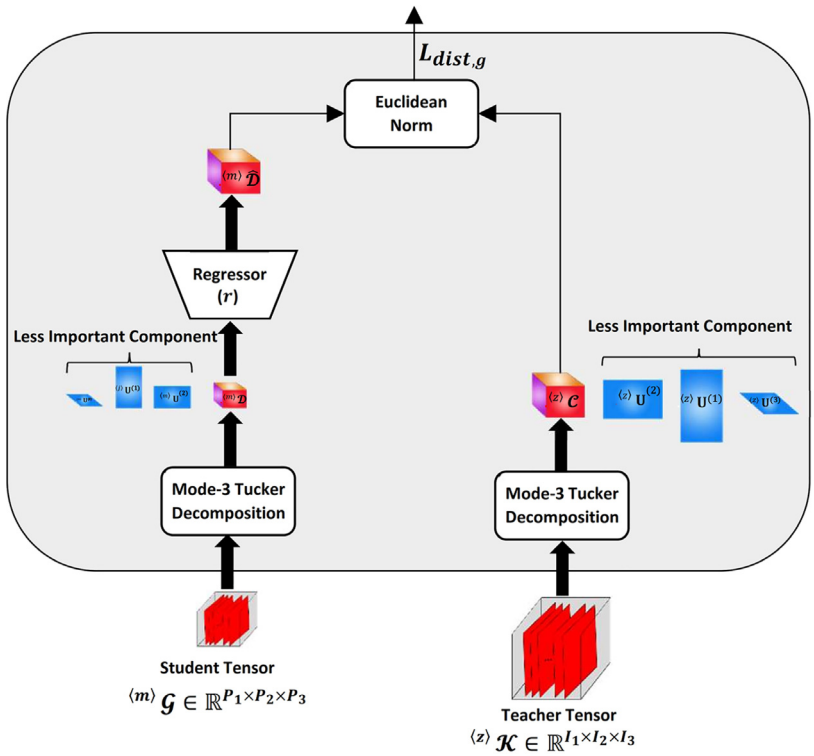


Fig. 3. The proposed KD module with mode-3 Tucker Decomposition. $^{\langle z\rangle}\mathcal{K}$ Layer z tensor of teacher with dimensions of $I_1 \times I_2 \times I_3$, $^{\langle m\rangle}\mathcal{G}$ Layer m tensor of student with dimension of $P_1 \times P_2 \times P_3$.

Each item (11) is represented as Eq. (12).

$$^{\langle z\rangle}\mathcal{W}_{i_1,\dots,i_L} = \sum_{r_1=1}^{R_1}\sum_{r_2=1}^{R_2}\cdots\sum_{r_L=1}^{R_L} {}^{\langle z\rangle}\alpha_{r_1,\dots,r_L}\, {}^{\langle z\rangle}\mathbf{R}^{(1)}_{(i_1,r_1)}\, {}^{\langle z\rangle}\mathbf{R}^{(2)}_{(i_2,r_2)} \cdots {}^{\langle z\rangle}\mathbf{R}^{(L)}_{(i_L,r_L)} \tag{12}$$

By decomposing $^{\langle z\rangle}\mathcal{W}$ tensor manifested in Eq. (11), the output of $^{\langle z\rangle}f$ can be demonstrated in the form of the decomposed tensors. If $^{\langle z\rangle}\mathcal{K}$ tensor with size of $I_1 \times I_2 \times I_3$ is the feature map obtained from $^{\langle z\rangle}f$, the resultant Tucker decomposition function will be in the following format.

$$\begin{aligned} {}^{\langle z\rangle}\mathcal{K} &\cong {}^{\langle z\rangle}\mathcal{C} \times_1 {}^{\langle z\rangle}\mathbf{V}^{(1)} \times_2 {}^{\langle z\rangle}\mathbf{V}^{(2)} \cdots \times_l {}^{\langle z\rangle}\mathbf{V}^{(l)} \cdots \times_L {}^{\langle z\rangle}\mathbf{V}^{(L)} \\ &= [\![{}^{\langle z\rangle}\mathcal{C}, {}^{\langle z\rangle}\mathbf{V}^{(1)}, {}^{\langle z\rangle}\mathbf{V}^{(2)} \dots, {}^{\langle z\rangle}\mathbf{V}^{(l)}, \dots, {}^{\langle z\rangle}\mathbf{V}^{(L)}]\!] \end{aligned} \tag{13}$$

Each item (13) is represented as Eq. (14).

$$^{\langle z\rangle}\mathcal{K}_{i_1,\dots,i_L} = \sum_{r_1=1}^{R_1}\sum_{r_2=1}^{R_2}\cdots\sum_{r_L=1}^{R_L} {}^{\langle z\rangle}\alpha_{r_1,\dots,r_L}\, {}^{\langle z\rangle}\mathbf{V}^{(1)}_{(i_1,r_1)}\, {}^{\langle z\rangle}\mathbf{V}^{(2)}_{(i_2,r_2)} \cdots {}^{\langle z\rangle}\mathbf{V}^{(L)}_{(i_L,r_L)} \tag{14}$$

Similarly, in Eq. (7), $\mathcal{H}$ for the student model will be represented as below.

$$\mathcal{H} = {}^{\langle M\rangle}f\left({}^{\langle M\rangle}\mathcal{W}\, {}^{\langle M-1\rangle}f\left({}^{\langle M-1\rangle}\mathcal{W}\, {}^{\langle M-2\rangle}f\left({}^{\langle M-2\rangle}\mathcal{W}\left(\cdots\left({}^{\langle 1\rangle}f\left({}^{\langle 1\rangle}\mathcal{W}\mathbf{X} + {}^{\langle 1\rangle}b \right)\right)\right)\cdots \right) + {}^{\langle M-2\rangle}b \right) + {}^{\langle M-1\rangle}b \right) + {}^{\langle M\rangle}b \right) \tag{15}$$

Through the decomposition of $^{\langle m\rangle}\mathcal{W}$ tensor, similar to Eq. (11), the tensor of feature maps of $^{\langle m\rangle}\mathcal{G}$ with size of $P_1 \times P_2 \times P_3$ obtained from $^{\langle m\rangle}f$ output at layer $m (m = 1, 2 \cdots M)$ will be as below:

$$\begin{aligned} {}^{\langle m\rangle}\mathcal{G} &\cong {}^{\langle m\rangle}\mathcal{D} \times_1 {}^{\langle m\rangle}\mathbf{Y}^{(1)} \times_2 {}^{\langle m\rangle}\mathbf{Y}^{(2)} \cdots \times_l {}^{\langle m\rangle}\mathbf{Y}^{(l)} \cdots \times_L {}^{\langle m\rangle}\mathbf{Y}^{(L)} \\ &= [\![{}^{\langle m\rangle}\mathcal{D}, {}^{\langle m\rangle}\mathbf{Y}^{(1)}, {}^{\langle m\rangle}\mathbf{Y}^{(2)} \dots, {}^{\langle m\rangle}\mathbf{Y}^{(l)}, \dots, {}^{\langle m\rangle}\mathbf{Y}^{(L)}]\!] \end{aligned} \tag{16}$$

$$^{\langle m\rangle}\mathcal{G}_{j_1,\dots,j_L} = \sum_{r_1=1}^{R_1}\sum_{r_2=1}^{R_2}\cdots\sum_{r_L=1}^{R_L} {}^{\langle m\rangle}b_{r_1,\dots,r_L}\, {}^{\langle m\rangle}\mathbf{Y}^{(1)}_{(j_1,r_1)}\, {}^{\langle m\rangle}\mathbf{Y}^{(2)}_{(j_2,r_2)} \cdots {}^{\langle m\rangle}\mathbf{Y}^{(L)}_{(j_L,r_L)} \tag{17}$$

Accordingly, when the decomposed tensors of feature maps are obtained from both teacher and student models, the next step is to distill knowledge from the teacher into a student network in a distillation module. According to Fig. 3, distilling knowledge in the proposed KD module is defined as imitating the teacher's behavior by the student at the inner convolution layers. Therefore, the distance between these two core tensors of $^{\langle m\rangle}\mathcal{C}$ belonging to the decomposed components of $^{\langle z\rangle}\mathcal{K}$ and the core tensor of $^{\langle m\rangle}\mathcal{D}$ belonging to the decomposed components of $^{\langle m\rangle}\mathcal{G}$ will be calculated at G module of knowledge distillation. However, due to the difference in the dimensions of $^{\langle m\rangle}\mathcal{C}$ and $^{\langle m\rangle}\mathcal{D}$ core tensors, ϕ function is used in this manuscript as an adaptor.

$$\phi(\mathcal{H}) = r\left({}^{\langle g=1\cdots G\rangle}\mathcal{D}, \mathcal{W}_r \right) \tag{18}$$

An r regressor bearing $\mathcal{W}_r$ parameters was used as the extra component to equalize the tensors of the feature maps of $^{\langle g\rangle}\mathcal{C}$ and $^{\langle g\rangle}\mathcal{D}$ in the distillation module. The r regressor will balance the $^{\langle g\rangle}\mathcal{C}$ tensor with $^{\langle g\rangle}\mathcal{D}$. The first candidate for r is a fully connected network; however, this network contains a large number of parameters that increase its time complexity, if $R_1 \times R_2$ and R_3 are the dimensions and number of channels of $^{\langle g\rangle}\mathcal{C}$ core tensor, respectively. Likewise, $H_1 \times H_2$ and H_3 denote the dimensions and number of channels of the $^{\langle g\rangle}\mathcal{D}$ core tensor, respectively. The number of parameters of the weight matrix of regressor that is fully connected to the r equal to $R_1 \times R_2 \times R_3 \times H_1 \times H_2 \times H_3$. In order to resolve this restriction and the overload of the fully connected regressor, a convolutional regressor was used in the proposed method. The convolutional regressor received the $^{\langle g\rangle}\mathcal{D}$

core tensor bearing $H_1 \times H_2$ dimensions as the input. Furthermore, it adjusted the shape of $K_1 \times K_2$ kernel so that its output would have the same measurements as the ${}^{\langle g \rangle}\mathcal{C}$ core tensor. The number of the parameters of convolutional regressor equals $K_1 \times K_2 \times R_3 \times H_3$. Evidently, $K_1 \times K_2$ is much smaller than $R_1 \times R_2 \times H_1 \times H_2$.

Finally, the feature based error (L_{IF}) between teacher and student will be obtained as follows in G KD modules.

$$L_{IF} = \frac{1}{G} \sum_{g=1}^{G} \left\| {}^{\langle g \rangle}\mathcal{C} - r\left({}^{\langle g \rangle}\mathcal{D}, \mathcal{W}_r\right) \right\|^2 \tag{19}$$

Algorithm 1 depicts the mechanism of training the student by the teacher model in the proposed KD whose complexity arises from $\theta\left(n^2\right)$ interval.

Algorithm 1 Training the Student Model Using Pretrained Teacher Model.

Input: $D = \{\mathbf{X}, \mathbf{y}\}$ // BreakHis Dataset
with:
t // Pretrained CNN Teacher Model
Train the Student Model s**:**
for each x_i in $\mathbf{X}$ **do**
  $p = t\left(x_i\right)$
  $q = s\left(x_i\right)$
  Use Eq. (2) to compute L_{KD} with p & q
  $\mathcal{K} = [{}^{\langle 1 \rangle}\mathcal{K}, {}^{\langle 2 \rangle}\mathcal{K}, \cdots, {}^{\langle Z \rangle}\mathcal{K}]$ // Feature Maps of t
  $\mathcal{G} = [{}^{\langle 1 \rangle}\mathcal{G}, {}^{\langle 2 \rangle}\mathcal{F}, \cdots, {}^{\langle M \rangle}\mathcal{G}]$ // Feature Maps of s
  for g=1 **to** G **do**
    ${}^{\langle g \rangle}\mathcal{K} \cong {}^{\langle g \rangle}\mathcal{C} \times_1 {}^{\langle g \rangle}\mathbf{V}^{(1)} \times_2 {}^{\langle g \rangle}\mathbf{V}^{(2)} \cdots \times_l {}^{\langle g \rangle}\mathbf{V}^{(l)} \cdots \times_L {}^{\langle g \rangle}\mathbf{V}^{(L)}$
    ${}^{\langle g \rangle}\mathcal{G} \cong {}^{\langle g \rangle}\mathcal{D} \times_1 {}^{\langle g \rangle}\mathbf{Y}^{(1)} \times_2 {}^{\langle g \rangle}\mathbf{Y}^{(2)} \cdots \times_l {}^{\langle g \rangle}\mathbf{Y}^{(l)} \cdots \times_L {}^{\langle g \rangle}\mathbf{Y}^{(L)}$
    $L_{dist,g} = \left\| {}^{\langle g \rangle}\mathcal{C} - r\left({}^{\langle g \rangle}\mathcal{D}, \mathcal{W}_r\right) \right\|^2$
  end for
  $L_{IF} = \frac{1}{G} \sum_{g=1}^{G} \mathcal{L}_{dist,g}$
  use Eq. (6) to compute overall Loss $\left(L_{Hybrid}\right)$ of the Student Model s
  Follow Backpropagation for s
end for

5. Experimental results

This section analyzes the outputs of the proposed teacher and student models based on ResNet and for further evaluation based on VGG. This manuscript used a ResNet-50 and a VGG-19 model consisting of 6 × 6 kernels, plus stride = 1, padding = 1 for the teacher network. The student network consisted of a ResNet-18 and a VGG-13 model including 3 × 3 kernels, plus stride = 1, padding = 1. The weights of teacher and student models were trained by the Adam optimizer, whereas T and α in Eq. (1) and Eq. (2) were considered 1.9 and 0.7, respectively. The proposed models were implemented on TensorFlow Lite for execution on smartphones simulated in Genymotion. Moreover, PyRapl was employed to determine the resource consumption rates of classification models. This experiment was conducted on Samsung smartphones with GPU power scaling shown in Fig. 4. The weakest Samsung smartphone selected was Galaxy S3 equipped with a dual-core Mali 400 MP GPU. In a dual-core processor, the computational complexity is estimated at up to 5.4 GFLOPs. However, a Galaxy S6 is equipped with an octa-core Vivante GC4000 GPU, the eight cores of which are used to increase the computational complexity up to 24.4 GFLOPs. Furthermore, a Galaxy S8 is equipped with a Mali T760 GPU, the eight cores of which are used simultaneously to increase its computational complexity up to 48.8 GFLOPs.

5.1. Dataset

Since the timely online medical diagnosis is an important task, a dataset of histopathological breast cancer images was selected as a case study in this paper to appraise the proposed lightweight classification model on smartphones in order to simulate the real-time diagnosis. Breast Cancer Histopathological Image Classification (BreakHis) is a dataset containing 9109 microscopic images of breast tumor tissues taken from 82 patients in different magnification factors (i.e., 40×, 100×, 200×, and 400×). The classification operations of this dataset were performed by pathologists. In fact, this dataset included 2480 malignant and 5429 benign samples (700 × 460 pixels, 3-channel RGB, 8-bit depth in each channel, and PNG format). Table 1 presents the sample details in each magnification factor (Spanhol et al., 2016). Fig. 5 shows different sample images magnification factors for eight classes in BreakHis dataset. The classification model performance is determined through the patient-level recognition rate as below.

Table 1
Image classification in BreakHis dataset.

Class	Sub-class	Magnification factors				Total	#Patients
		40×	100×	200×	400×		
Benign	adenosis	114	113	111	106	444	24
	fibroadenoma	253	260	264	237	1014	
	phyllodes_tumor	109	121	108	115	453	
	tubular_adenoma	149	150	140	130	569	
Malignant	ductal_carcinoma	864	903	896	788	3451	58
	lobular_carcinoma	156	170	163	137	626	
	mucinous_carcinoma	205	222	196	169	792	
	papillary_carcinoma	145	142	135	138	560	
Total		1995	2081	2013	1820	7909	82

$$Recognition\ rate = \frac{\sum Patient\ score}{Total\ number\ of\ patients} \tag{20}$$

$$Patient\ score = \frac{N_{Rec}}{N_P} \tag{21}$$

Where N_P denotes the number of images belonging to the patient P, and N_{Rec} indicates the number of images classified correctly.

5.2. Effect of Hyper-Parameter β on KD performance

In the proposed model, β parameter in Eq. (6) is a determinant that specifies the impact factor of the response-based KD (L_{KD}) and the decomposed feature-based KD (L_{IF}). Therefore, the diagram in Fig. 6 is employed to examine various amounts of β variable in Eq. (6). This diagram includes the performance of students developed based on VGG and ResNet architecture. According to the axis x of this diagram, β variable ranges between 0 to 1. Axis y demonstrates the recognition rate obtained by each student for all samples in various classes of the BreakHis dataset. The green line is the ResNet-based student and the red line is the VGG-based student. This diagram indicates that both student models with β variable amounting to 0.7 delivered their best performance. Thus, in the tests, the β was considered to be 0.7.

5.3. On-device evaluations

Tables 2 and 3 report the evaluation results of the proposed teacher and student models based on VGG and ResNet on average in terms of both recognition rate and model complexity on smartphones. Table 2 indicates the performance results of VGG Teacher, ResNet Teacher, VGG Student, and ResNet Student classification models for malignant samples as well as their complexity rates in the use of smartphone resources. According to the results, dimensions of the student models are much lower than those of the teacher models; however, their recognition rate results are not significantly different. The highest compression rate came from ResNet. In other words, the ResNet Student model was 262.18 times more compressed and had ∼ 11.92M parameters. The computational complexity of the ResNet Student model was ∼ 1.71 GFLOPs, a rate which is appropriate for all smartphones (Fig. 4). Furthermore, Table 3 reports the performance results of classification

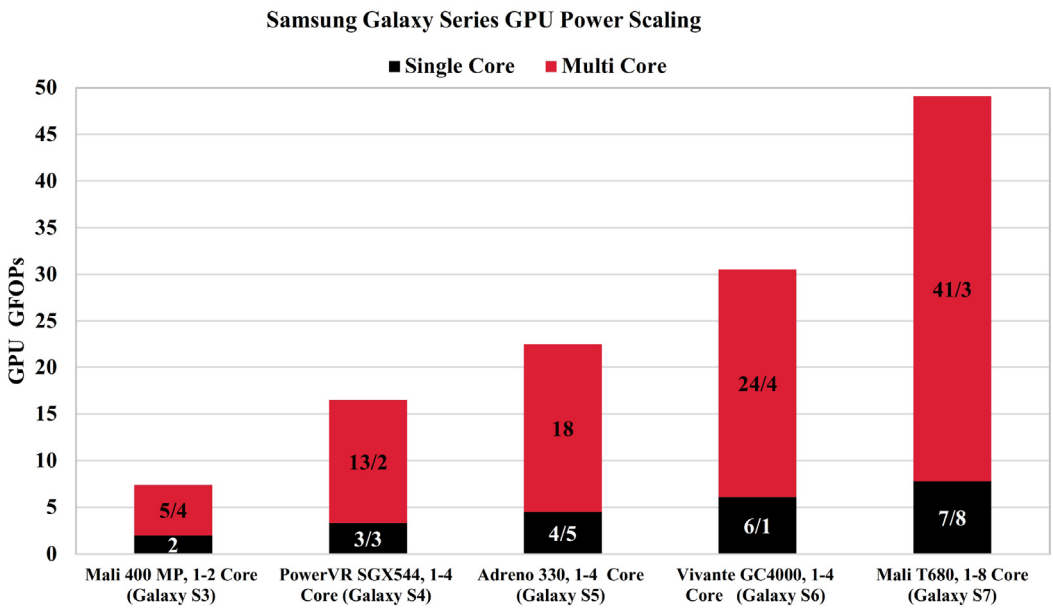


Fig. 4. The computational power of several Samsung smartphones based on GFLOPs value in GPU.

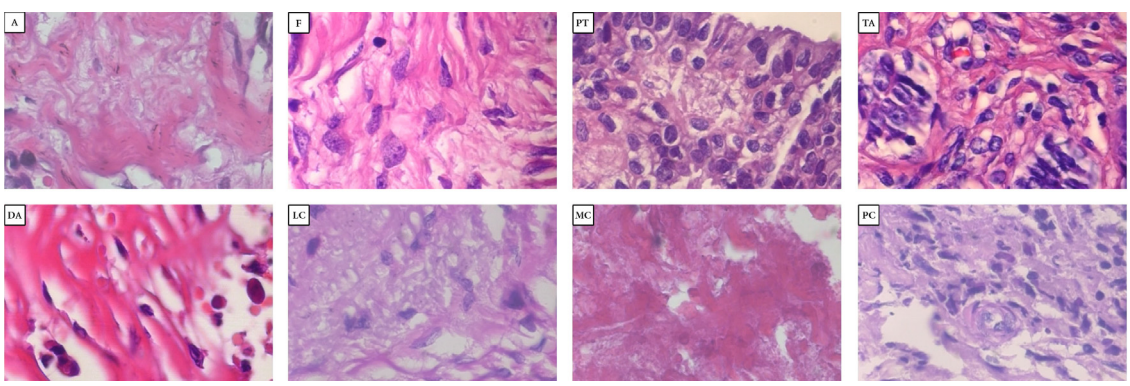


Fig. 5. Two Breast histopathology images from BreakHis dataset with four magnification factor.

Table 2
The performance results of the proposed teacher and student models for the classification of malignant samples on smartphones.

Models	No. of parameters	Computation (GFLOPs)	GPU memory	Compression rate	Recognition rate (%)
VGG Teacher	$\sim 102.18M$	~ 7.27	1.35 G	×5.51	96.88
VGG Student	$\sim 29.13M$	~ 2.82	0.69 G	×138.05	95.61
ResNet Teacher	$\sim 81.36M$	~ 8.11	1.92 G	×1.01	98.35
ResNet Student	$\sim 11.92M$	~ 1.71	0.81 G	×262.18	97.71

Table 3
The performance results of the proposed teacher and student models for the classification of benign samples on smartphones.

Models	No. of parameters	Computation (GFLOPs)	GPU memory	Compression rate	Recognition rate (%)
VGG Teacher	$\sim 104.12M$	~ 7.18	1.22 G	×5.43	96.65
VGG Student	$\sim 29.13M$	~ 2.82	0.71 G	×125.90	95.78
ResNet Teacher	$\sim 74.32M$	~ 8.42	1.96 G	×1.09	97.94
ResNet Student	$\sim 12.87M$	~ 1.92	0.82 G	×269.17	97.68

models on benign samples as well as the number of parameters in those models. The highest compression rate came from ResNet. In other words, the ResNet Student model was nearly 269.17 times more compressed and had $\sim 12.87M$ parameters. Both Table 2 and Table 3 indicate that VGG Student and ResNet Student models also used the GPU memory. For instance, Table 3 indicates that the ResNet Student model consumed nearly 1.14 G less GPU memory than its teacher. Regarding the benign samples, the computational complexity of the ResNet Student model was $\sim$ 1.92 GFLOPs, which is accepted as an appropriate rate.

5.4. The class-based evaluation of the proposed KD models

This section analyzes the classification results of malignant and benign samples in different magnification factors (i.e., 40×, 100×, 200×, and 400×). Tables 4 and 5 report the results of these tests. Accordingly, the student models indicate once again that they have no significant differences from their teachers in the details of results, for they employed the knowledge of their teachers in the training process. The results also indicate that the student models had higher recognition rates in 200× and 400× magnification factors. As Tables 4 and 5 demonstrate, the ResNet Student model was selected as a lightweight student with better outputs. It will then be analyzed in detail.

Fig. 7 depicts the confusion matrices of the ResNet Student model for the classification of binary and all malignant and benign samples in 8 different class of BreakHis. The main diagonal values in Fig. 7 indicate a high true positive; thus, the ResNet Student model can appropriately distinguish between malignant and benign samples. In fact, 5370 out of 5429 samples were recognized correctly for the malignant class, and

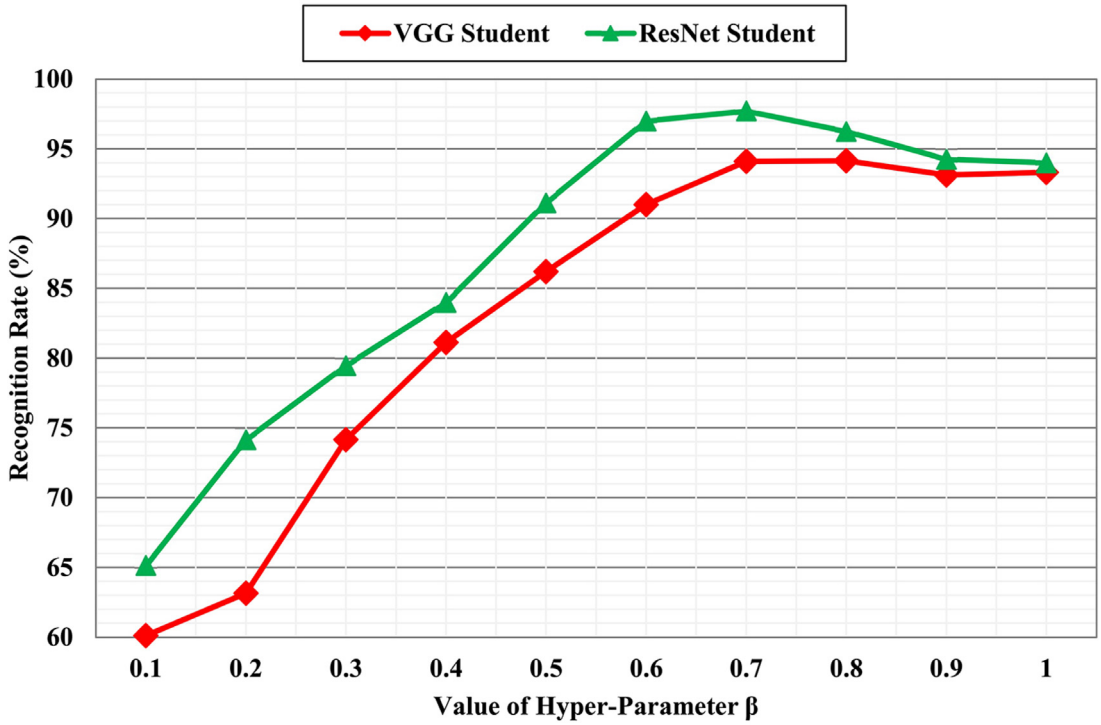


Fig. 6. The results of the Hyper-Parameter β on the proposed KD.

Table 4
The recognition rates of the teacher and student models in different magnification factors of malignant samples.

Models	Class	Recognition rate				
		40×	100×	200×	400×	Average
VGG Teacher	adenosis	99.30	95.90	97.35	96.95	**97.37**
	fibroadenoma	98.35	98.50	95.25	95.60	**97.17**
	phyllodes_tumor	99.25	98.10	94.20	94.70	**96.56**
	tubular_adenoma	99.10	97.90	95.10	94.60	**96.67**
VGG Student	adenosis	96.10	95.40	96.90	96.10	**96.12**
	fibroadenoma	96.00	96.20	95.10	96.20	**95.87**
	phyllodes_tumor	95.95	96.22	94.10	94.22	**95.12**
	tubular_adenoma	96.05	95.95	95.20	94.25	**95.53**
ResNet Teacher	adenosis	99.64	99.30	97.95	97.00	**98.47**
	fibroadenoma	99.10	99.65	97.33	97.00	**98.27**
	phyllodes_tumor	99.30	98.10	97.95	97.90	**98.31**
	tubular_adenoma	99.35	99.20	97.65	97.25	**98.36**
ResNet Student	adenosis	98.44	98.00	97.75	96.90	**97.77**
	fibroadenoma	98.36	98.10	97.25	97.00	**97.67**
	phyllodes_tumor	98.24	98.71	97.00	97.10	**97.76**
	tubular_adenoma	98.00	98.25	97.20	97.15	**97.65**

Table 5
The recognition rates of the teacher and student models in different magnification factors of benign samples.

Models	Class	Recognition rate				
		40×	100×	200×	400×	Average
VGG Teacher	ductal_carcinoma	99.10	98.80	96.40	96.00	**97.57**
	lobular_carcinoma	97.10	97.25	96.25	95.00	**96.40**
	mucinous_carcinoma	97.90	97.15	96.00	94.90	**96.48**
	papillary_carcinoma	98.05	97.05	95.10	94.45	**96.16**
VGG Student	ductal_carcinoma	97.00	96.30	96.00	95.00	**96.07**
	lobular_carcinoma	97.10	96.25	96.00	95.00	**96.08**
	mucinous_carcinoma	96.80	96.00	95.10	94.25	**95.53**
	papillary_carcinoma	96.66	96.10	95.00	94.01	**95.44**
ResNet Teacher	ductal_carcinoma	98.97	98.34	97.65	97.30	**98.06**
	lobular_carcinoma	98.65	98.84	97.10	97.05	**97.91**
	mucinous_carcinoma	99.00	98.95	97.00	96.98	**97.98**
	papillary_carcinoma	99.10	98.00	98.00	96.15	**97.81**
ResNet Student	ductal_carcinoma	98.46	98.10	97.20	97.10	**97.71**
	lobular_carcinoma	98.30	98.22	97.00	97.14	**97.66**
	mucinous_carcinoma	98.60	98.54	97.05	97.00	**97.79**
	papillary_carcinoma	98.25	98.03	97.00	96.98	**97.56**

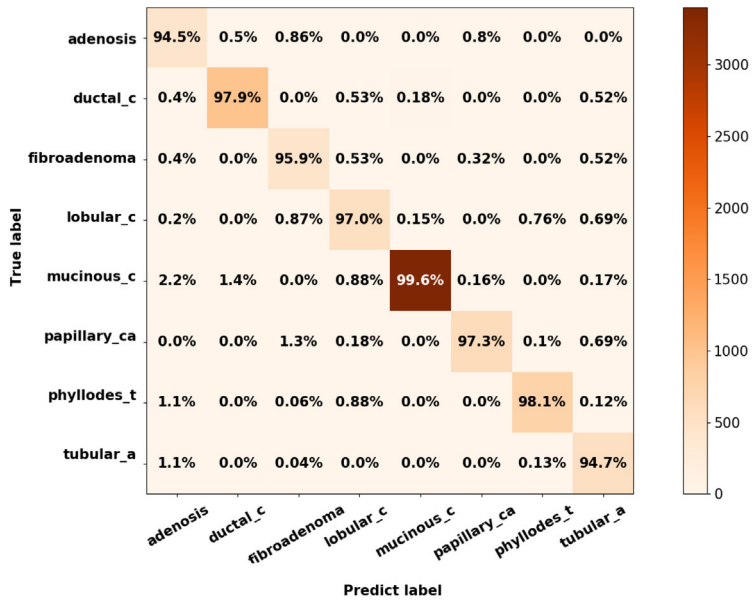
only 59 samples were recognized as the false positives of the benign class accounting for less than 1% of samples.

The results of these confusion matrices indicate the satisfactory performance of the ResNet Student model in terms of accuracy, precision, recall, and F-score. The receiver operating characteristic (ROC) curve was drawn by using the ratio of the true positive rate to the false positive rate in order to show the ResNet Student performance more accurately for the binary classification of all malignant and benign samples. The ROC curve is employed to determine the sensitivity and specificity of the ResNet Student model. Fig. 8 demonstrates the ROC curve of the ResNet Student model for the classification of all malignant and benign samples in different magnification factors (i.e., 40×, 100×, 200×, and 400×). This curve depicts the ratio of the true positive rate to the false positive rate suiting the ResNet Student model. The AUC was reported 97.80% in 40×.

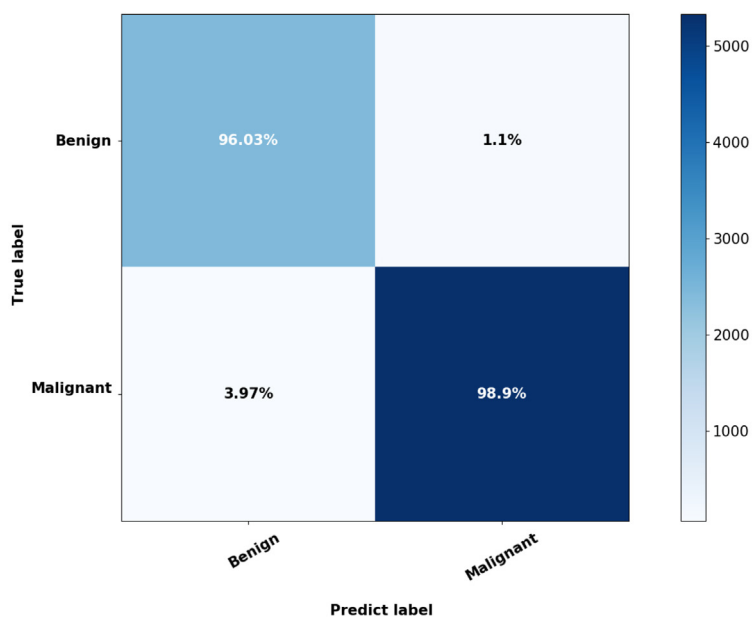
5.5. Comparing the proposed KD with state-of-the-art KD methods

This section entails evaluating the proposed KD method and comparing its outputs with those of state-of-the-art methods in KD. In recent years, various methods have been proposed for KD; thus, methods similar to the one proposed in this paper were selected in order to draw equitable comparisons. Hence, the following state-of-the-art methods were selected as the baselines.

- **O-KD** (Hinton et al., 2015): This was the first method that introduced KD. In this method, a teacher trains a student on the basis of response-based knowledge concerning the labels created by the teacher.
- **FitNet** (Romero et al., 2015): The idea to use the middle layer of the deep model was proposed for the first time by FitNets (Romero et al., 2015) to provide guidance for training the student



(a) eight-class classification



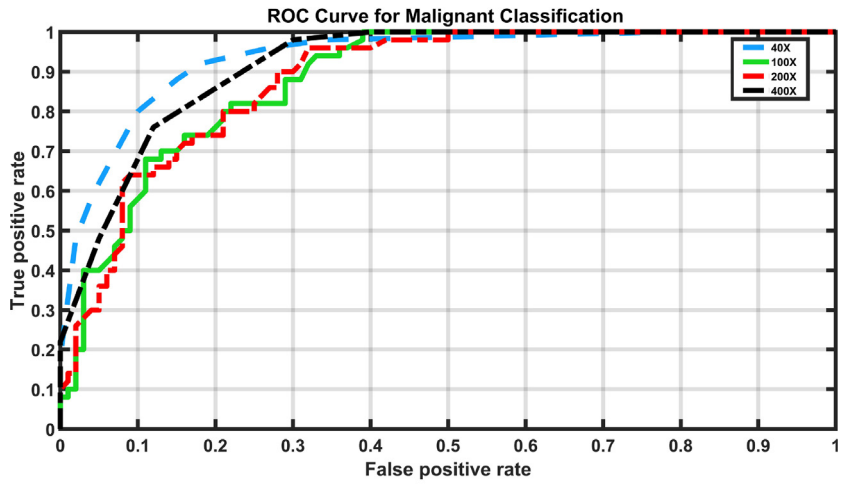
(b) binary classification

Fig. 7. The confusion matrices of the results for the ResNet Student classification: (a) eight-class classification, (b) binary classification.

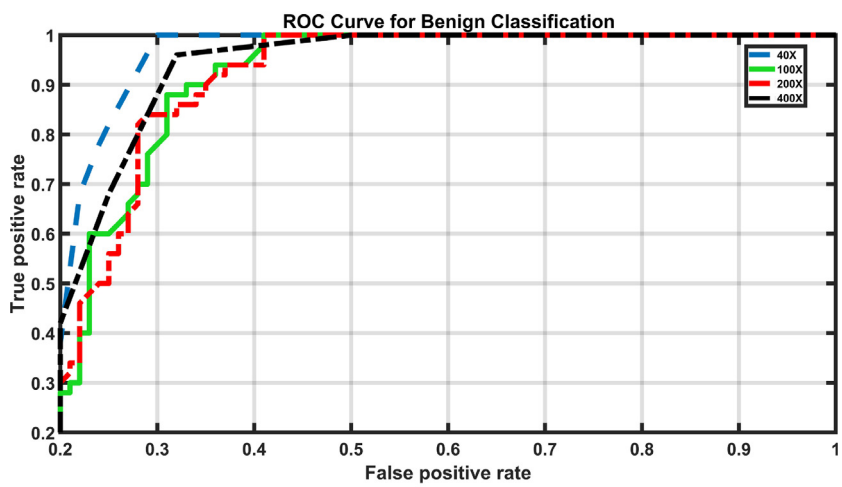
model. In this model, the guide represents the output of the hidden layer from a teacher that supervises students' learning processes. The principal idea is to directly adjust the distinct characteristics of teachers and students.

- **T-KD** (Dai et al., 2021): In this method, a KD model is proposed for learning students considering the Tucker decomposition as well as using the output responses of a teacher (response-based KD).
- **ET-KD** (Tan et al., 2021): This method uses the inter-class correlation regularization of an expressive teacher so that the teacher will be able to learn the student in case of using the beneficial knowledge.
- **AMTML-KD** (Liu et al., 2020): In this method, several teachers were uses to learn one student. The distilled knowledge is a combination of the internal layers from teachers as well as the responses provided by each teacher. In this method, each teacher randomly teaches a group of layers from a student.

Table 6 reports the results of evaluating and comparing the proposed KD method with the baseline methods. In this table, the proposed KD model was compared with other KD models in BreakHis for various classes with regard to the recognition rate criteria. Following the teacher and student comparisons designed per the proposed KD model and other KD models, they had a similar architecture. The results indicated that the proposed KD model had a recognition rate of 97.71% for malignant classes and a recognition rate of 97.68% for benign classes. For instance, the proposed KD cited by Tan et al. (2021) yielded a recognition rate of 91.74% for malignant classes and a recognition rate of 92.42% for benign classes. According to these results, the proposed KD had a 5.94% recognition rate for the malignant



(a) ROC curves for the malignant samples



(b) ROC curves for the benign samples

Fig. 8. The ROC curves for the ResNet Student: (a) malignant samples, (b) benign samples.

Table 6
A comparison between the proposed KD and some other state-of-the-art methods in performance.

Method	Malignant				Benign				Average
	40X	100X	200X	400X	40X	100X	200X	400X	
O-KD (Hinton et al., 2015)	80.01	81.14	83.32	81.08	81.11	80.75	82.55	80.07	**81.25**
FitNet (Romero et al., 2015)	84.07	83.74	82.25	81.55	80.25	81.19	82.11	82.96	**82.26**
T-KD (Dai et al., 2021)	91.25	93.75	92.68	91.42	93.08	92.75	91.42	90.09	**92.05**
ET-KD (Tan et al., 2021)	93.43	91.19	90.25	92.09	92.10	91.78	93.67	92.16	**92.08**
AMTML-KD (Mirzadeh et al., 2020)	95.25	94.43	95.25	96.10	94.63	97.47	96.75	96.24	**95.76**
Proposed KD	**97.77**	**97.67**	**97.76**	**97.65**	**97.71**	**97.66**	**97.79**	**97.56**	**97.68**

class and a 5.26% recognition rate for the benign class, which are higher than those reported by Tan et al. (2021). According to Table 6, it can be argued that the proposed KD outperformed the other previously proposed KD methods.

The paired t-test was conducted in order to examine whether the difference between the mean of the proposed KD method and other KD methods in Table 6 was statistically significant. If t statistic $(T_{obs}) < -t$ Critical(T_{crit}) in the paired t-test, the null hypothesis is then rejected. At the same time, if $p_value < \alpha = 0.05$, the null hypothesis is rejected again. It can be concluded that three differences between the mean scores are quite convenient, bearing a confidence level of 95%. Thus, it can be regarded as a significant difference. Table 7 indicates the results of calculating the paired t-test. Following Table 7, in all comparisons $T_{obs} < T_{crit}$, plus $p - value < 0.05$. Therefore, the null hypothesis is rejected, and it can be concluded that the recognition rates for the benign and malignant classes obtained for BreakHis in Table 6 between the proposed KD method and other KD methods are not due to chance.

6. Conclusion and future work

This paper designed a lightweight student model using feature-based KD. One of the main challenges of developing a student model in a feature-based KD is the difference in the spatial shape of the complex feature maps of teacher and simpler feature maps of student. It prevents their direct comparison to complete the process of distillation. Therefore, Tucker decomposition was used in the proposed KD to decompose the feature maps of teacher. Then, the core tensors of these feature maps were calculated as low dimensions tensors. Afterward, an

Table 7
The paired t-test results on the proposed KD and other KD methods.

Class	References	T_{obs}	T_{crit}	p_Value
Malignant	O-KD (Hinton et al., 2015)	−5.5890	3.4125	0.0041
	FitNet (Romero et al., 2015)	−4.0021	3.2458	0.0024
	T-KD (Dai et al., 2021)	−3.1116	2.1754	0.0241
	ET-KD (Tan et al., 2021)	−2.8724	2.0125	0.0210
	AMTML-KD (Mirzadeh et al., 2020)	−1.1093	1.8943	0.0417
Benign	O-KD (Hinton et al., 2015)	−5.5329	4.0254	0.0012
	FitNet (Romero et al., 2015)	−4.4672	3.6420	0.0021
	T-KD (Dai et al., 2021)	−2.9292	2.0001	0.0101
	ET-KD (Tan et al., 2021)	−2.0043	2.1254	0.0215
	AMTML-KD (Mirzadeh et al., 2020)	−1.3432	1.5705	0.0426

adaptor function was employed using convolution regressor to balance the dimension of core tensor of teacher with the tensors of student. Having suggested a hybrid loss function, the information of the core tensors in the decomposed layers of the teacher model were distilled to the student as the outcome of the previously-learned knowledge. Both teacher and student models were implemented on smartphones used as edge devices, and their performance outputs were evaluated in terms of recognition rate and complexity. According to the experimental results, the compression rate of the ResNet-18 based student model was nearly 265.67 times higher than that of its ResNet-50 based teacher; however, its recognition rate declined by only $\sim 1.5\%$ on average in the benchmark dataset.

Although the experimental results reported the appropriate performance of the proposed KD method for training a student, this method still faces a few constraints. In the proposed method, a pre-trained single teacher was employed to train a single student online through the teacher's middle knowledge. However, the teacher network was a statistical machine learning model in which diversity might not have been observed. In other words, this model might show high sensitivity to a specific segment of data, whereas it might be less sensitive to the other segments of data. A disadvantage of statistical learning algorithms is high sensitivity to a segment of input data and generation of different weights in each training process, something which results in different predictions. Generally, these models are called high-variance models. A solution to this problem is to employ multiple teachers simultaneously to train a student through ensemble learning. The ensemble of a variety of teachers can generally have positive effects on accuracy and reduce overfitting. Therefore, the student model will also avoid overfitting. Our plan for future work is to adopt an ensemble of teacher models to address this issue.

CRediT authorship contribution statement

Majid Sepahvand: Conceptualization, Software, Data curation, Writing – original draft, Visualization, Investigation, Validation. **Fardin Abdali-Mohammadi:** Conceptualization, Methodology, Project administration, Supervision, Writing – review & editing. **Amir Taherkordi:** Conceptualization, Methodology, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Bang, Duhyeon, Lee, Jongwuk, Shim, Hyunjung, 2021. Distilling from professors: Enhancing the knowledge distillation of teachers. Inform. Sci. 576, 743–755.

Cheng, Yu, Wang, Duo, Zhou, Pan, Zhang, Tao, 2018. Model compression and acceleration for deep neural networks: The principles, progress, and challenges. IEEE Signal Process. Mag. 35 (1), 126–136.

Dai, Cheng, Liu, Xingang, Li, Zhuolin, Chen, Mu-Yen, 2021. A tucker decomposition based knowledge distillation for intelligent edge applications. Appl. Soft Comput. 101, 107051.

Ding, Haisong, Chen, Kai, Huo, Qiang, 2019. Compressing CNN-DBLSTM models for OCR with teacher-student learning and tucker decomposition. Pattern Recognit. 96, 106957.

Fu, Shipeng, Li, Zhen, Liu, Kai, Din, Sadia, Imran, Muhammad, Yang, Xiaomin, 2020. Model compression for IoT applications in industry 4.0 via multiscale knowledge transfer. IEEE Trans. Ind. Inform. 16 (9), 6013–6022.

Gou, Jianping, Yu, Baosheng, Maybank, Stephen J, Tao, Dacheng, 2021. Knowledge distillation: A survey. Int. J. Comput. Vis. 129 (6), 1789–1819.

Gupta, Koyel Datta, Sharma, Deepak Kumar, Ahmed, Shakib, Gupta, Harsh, Gupta, Deepak, Hsu, Ching-Hsien, 2021. A novel lightweight deep learning-based histopathological image classification model for IoMT. Neural Process. Lett. 1–24.

Hinton, Geoffrey, Vinyals, Oriol, Dean, Jeff, 2015. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

Huang, Zehao, Wang, Naiyan, 2017. Like what you like: Knowledge distill via neuron selectivity transfer. arXiv preprint arXiv:1707.01219.

Karthiga, R., Narashimhan, K., 2021. Deep convolutional neural network for computer-aided detection of breast cancer using histopathology images. J. Phys. Conf. Ser. 1767 (1), 012042.

Lan, Xu, Zhu, Xiatian, Gong, Shaogang, 2018. Knowledge distillation by on-the-fly native ensemble. arXiv preprint arXiv:1806.04606.

Lee, Seunghyun, Song, Byung Cheol, 2020. Knowledge transfer via decomposing essential information in convolutional neural networks. IEEE Trans. Neural Netw. Learn. Syst. 1–12.

Liu, Yuang, Zhang, Wei, Wang, Jun, 2020. Adaptive multi-teacher multi-level knowledge distillation. Neurocomputing 415, 106–113.

Mirzadeh, Seyed Iman, Farajtabar, Mehrdad, Li, Ang, Levine, Nir, Matsukawa, Akihiro, Ghasemzadeh, Hassan, 2020. Improved knowledge distillation via teacher assistant. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34. (04), pp. 5191–5198.

Pan, Yiteng, He, Fazhi, Yan, Xiaohu, Li, Haoran, 2021. A synchronized heterogeneous autoencoder with feature-level and label-level knowledge distillation for the recommendation. Eng. Appl. Artif. Intell. 106, 104494.

Passalis, Nikolaos, Tefas, Anastasios, 2018. Learning deep representations with probabilistic knowledge transfer. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 268–284.

Reenadevi, R., Sathiya, T., Sathiyabhama, B., 2021. Breast cancer histopathological image classification using augmentation based on optimized deep ResNet-152 structure. Ann. Romanian Soc. Cell Biol. 25 (6), 5866–5874.

Romero, Adriana, Ballas, Nicolas, Kahou, Samira Ebrahimi, Chassang, Antoine, Gatta, Carlo, Bengio, Yoshua, 2015. Fitnets: Hints for thin deep nets. ICLR.

Sandler, Mark, Howard, Andrew, Zhu, Menglong, Zhmoginov, Andrey, Chen, Liang-Chieh, 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520.

Sepahvand, Majid, Abdali-Mohammadi, Fardin, 2021. A novel representation in genetic programming for ensemble classification of human motions based on inertial signals. Expert Syst. Appl. 185, 115624.

Sepahvand, Majid, Abdali-Mohammadi, Fardin, 2022. Overcoming limitation of dissociation between MD and MI classifications of breast cancer histopathological images through a novel decomposed feature-based knowledge distillation method. Comput. Bio. Medicine 145, 105413.

Shen, Peng, Lu, Xugang, Li, Sheng, Kawai, Hisashi, 2020. Knowledge distillation-based representation learning for short-utterance spoken language identification. IEEE/ACM Trans. Audio Speech Lang. Proc. 28, 2674–2683.

Shu, Changyong, Liu, Yifan, Gao, Jianfei, Yan, Zheng, Shen, Chunhua, 2021. Channel-wise knowledge distillation for dense prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5311–5320.

Sohail, Anabia, Khan, Asifullah, Wahab, Noorul, Zameer, Aneela, Khan, Saranjam, 2021. A multi-phase deep CNN based mitosis detection framework for breast cancer histopathological images. Sci. Rep. 11 (1), 1–18.

Song, Xuemeng, Feng, Fuli, Han, Xianjing, Yang, Xin, Liu, Wei, Nie, Liqiang, 2018. Neural compatibility modeling with attentive knowledge distillation. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. pp. 5–14.

Spanhol, Fabio A., Oliveira, Luiz S., Petitjean, Caroline, Heutte, Laurent, 2016. A dataset for breast cancer histopathological image classification. IEEE Trans. Biomed. Eng. 63 (7), 1455–1462.

Tan, Chao, Liu, Jie, Zhang, Xiang, 2021. Improving knowledge distillation via an expressive teacher. Knowl.-Based Syst. 218, 106837.

Teerapittayanon, Surat, McDanel, Bradley, Kung, H.T., 2017. Distributed deep neural networks over the cloud, the edge and end devices. In: 2017 IEEE 37th International Conference on Distributed Computing Systems. ICDCS, pp. 328–339.

Versaci, Mario, Angiulli, Giovanni, Crucitti, Paolo, De Carlo, Domenico, Laganà, Filippo, Pellicanò, Diego, Palumbo, Annunziata, 2022. A fuzzy similarity-based approach to classify numerically simulated and experimentally detected carbon fiber-reinforced polymer plate defects. Sensors 22 (11).

Wang, Yunhe, Xu, Chang, Xu, Chao, Tao, Dacheng, 2018. Packing convolutional neural networks in the frequency domain. IEEE Trans. Pattern Anal. Mach. Intell. 41 (10), 2495–2510.

Yang, Chenglin, Xie, Lingxi, Qiao, Siyuan, Yuille, Alan, 2018. Knowledge distillation in generations: More tolerant teachers educate better students. arXiv preprint arXiv:1805.05551.

Yu, Xiyu, Liu, Tongliang, Wang, Xinchao, Tao, Dacheng, 2017. On compressing deep models by low rank and sparse decomposition. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 67–76.

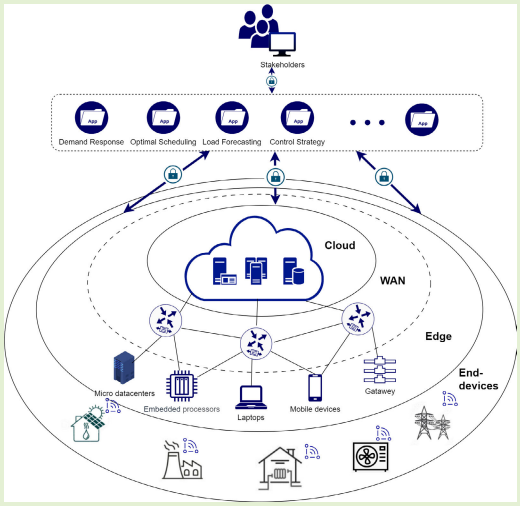
Zhang, Zhi, Ning, Guanghan, He, Zhihai, 2017. Knowledge projection for deep neural networks. arXiv preprint arXiv:1710.09505.

Zhu, Zhuangdi, Hong, Junyuan, Zhou, Jiayu, 2021. Data-free knowledge distillation for heterogeneous federated learning. arXiv preprint arXiv:2105.10056.

Edge-Cloud Architectures for Hybrid Energy Management Systems: A Comprehensive Review

Olha Boiko, Anton Komin, Reza Malekian, *Senior Member, IEEE*, and Paul Davidsson

***Abstract*—This article provides an overview of recent research on edge-cloud architectures in hybrid energy management systems (HEMSs). It delves into the typical structure of an IoT system, consisting of three key layers: the perception layer, the network layer, and the application layer. The edge-cloud architecture adds two more layers: the middleware layer and the business layer. This article also addresses challenges in the proposed architecture, including standardization, scalability, security, privacy, regulatory compliance, and infrastructure maintenance. Privacy concerns can hinder the adoption of HEMS. Therefore, we also provide an overview of these concerns and recent research on edge-cloud solutions for HEMS that addresses them. This article concludes by discussing the future trends of edge-cloud architectures for HEMS. These trends include increased use of artificial intelligence on an edge level to improve the performance and reliability of HEMS and the use of blockchain to improve the security and privacy of edge-cloud computing systems.**



***Index Terms*—Distributed computing, distributed energy, domestic energy consumption, edge computing, edge intelligence, hybrid renewable energy systems, Internet of Energy, power systems, renewable energy sources, residential energy consumption, sustainable development, systems architecture, trustworthiness, user data privacy.**

I. Introduction

ENERGY efficiency is a key factor in achieving climate neutrality and reducing energy dependence. Buildings, transport, and industry together are the main consumers of energy and sources of emissions. Buildings, in particular, account for around 40% of the total energy consumption in the European Union, contributing to 36% of the greenhouse gas emissions resulting from energy-related activities [1].

"Directive (EU) 2023/1791 of the European Parliament and of the Council" [1] sets the goal of ensuring more efficient primary energy consumption and increasing the share of renewable energy supplied to heating and cooling networks. At the same time, an efficient district heating and cooling system must use at least 50% renewable energy, 50% waste heat, 75% cogenerated heat, or 50% of a combination of such energy and heat by December 31, 2027, and from January 1, 2050, only renewable energy, only waste heat, or their combination.

Furthermore, the "Directive (EU) 2023/1791" [1] emphasizes that end-users should, upon request, receive access to information about their energy bills and data on historical consumption from their energy service provider in order to get a clear and comprehensible explanation of their actual energy consumption, its prices, and the energy sources used to satisfy their needs. It also requires information and estimates of energy costs to be provided to consumers on demand

Manuscript received 8 February 2024; accepted 21 March 2024. Date of publication 4 April 2024; date of current version 15 May 2024. This work was supported by the J. Gust. Richert stiftelse (Sweco Sverige) for the project Multi-Objective Optimization and Optimal Integration of Hybrid Energy Systems under Grant 01003438636, in part by the Kamprad Family Foundation (Familjen Kamprads Stiftelse) for the project Models of Distributed Information Processing in Smart Grid Systems, and in part by the Knowledge Foundation (Stiftelsen för kunskaps- och kompetensutveckling) for the project Intelligent and Trustworthy IoT Systems under Grant 20220087-H-01. The associate editor coordinating the review of this article and approving it for publication was Prof. Chen Yang. *(Corresponding author: Olha Boiko.)*

Olha Boiko is with the Department of Information Technologies, Sumy State University, 40007 Sumy, Ukraine, and also with the Department of Computer Science and Media Technology and the Internet of Things and People Research Centre, Malmö University, 205 06 Malmö, Sweden (e-mail: o.boiko@cs.sumdu.edu.ua).

Anton Komin is with the Department of Information Technologies, Sumy State University, 40007 Sumy, Ukraine (e-mail: anton.komin2@gmail.com).

Reza Malekian and Paul Davidsson are with the Department of Computer Science and Media Technology and the Internet of Things and People Research Centre, Malmö University, 205 06 Malmö, Sweden (e-mail: reza.malekian@mau.se; paul.davidsson@mau.se).

Digital Object Identifier 10.1109/JSEN.2024.3382390

1558-1748 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

on time and in an easily understandable format that allows consumers to compare deals on a like-for-like basis.

Achieving energy efficiency necessitates a holistic strategy that acknowledges the interconnectedness of components within a hybrid energy system (HES). The International Renewable Energy Agency (IRENA's) report emphasizes the importance of considering all aspects of the energy system, from production to consumption, to maximize efficiency [2]. This approach aligns with the findings of IEA, which underscores the importance of HESs in achieving energy security and resilience [3]. The recent Investors Dialogue on Energy as a platform by the European Commission highlighted the economic benefits of energy efficiency measures, emphasizing their potential to reduce energy costs and stimulate economic growth [4].

One way to ensure efficiency is to employ an HES based on renewable energy. HES is a system that consists of two or more energy sources that work together to provide the consumer with energy for various usage scenarios (heating, electricity, cooling, and water heating) [5]. Such a system can use different energy sources, such as renewable energy sources, conventional energy sources, and waste heat. However, energy efficiency requires a comprehensive approach that considers the interconnection of the HES, the security of supply, economic benefits, and climate neutrality throughout the entire value chain, from production to transportation and end-user consumption. This approach must also account for system performance and dynamic energy demand.

Management systems for HES are effective tools to solve the European Union's set goals. They automate the processes of monitoring and managing energy consumption, generating relevant reports to different stakeholders, and providing detailed feedback on energy consumption, which is critical for behavior change and supply and demand management [6]. Energy management in HES is important for several reasons. First, it can help to reduce the cost of energy by optimizing the use of renewable energy sources. Second, it can improve the reliability of HES by ensuring that there is enough energy to meet demand, even when renewable energy sources are not available. Third, it can help reduce the environmental impact of HES by minimizing greenhouse gas emissions. There are a number of existing hybrid energy management systems (HEMSs) on the market. HEMSs are typically installed in homes and businesses to help users manage their energy consumption [7].

HEMSs typically collect data on energy usage from smart meters and other sensors, and they collect and generate energy consumption data at different time intervals [8]. Users can see how much energy they are using in real time, as well as in daily, weekly, monthly, and yearly formats. This information helps them understand their energy habits and makes them more efficient [9]. For utility companies, these data can be used for accurate billing and customized services. Companies can use the data to determine how much energy each household uses and provide customized tariffs or services based on that [10]. For other new market players, these data can be used to create new products and services that help users manage their energy and use it more efficiently.

However, existing HEMS systems are often complex and difficult to use. An edge-cloud architecture can address the limitation in several ways. For instance, it can simplify the HEMS by offloading some processing and storage tasks to the cloud. This can make the HEMS easier to use and maintain. The cloud server can access and process data from various sources, including weather forecasts, smart meters, and user profiles. The edge server can perform some data processing tasks, even machine learning (ML) and artificial intelligence (AI). For example, the edge could predict energy demand or optimize energy consumption in real time. This can reduce the amount of data that need to be sent to the cloud server, which can improve performance and reduce costs.

Edge-cloud architectures for HEMS have gained significant traction in recent years, driven by the need to process and analyze critical energy data in real time. This trend is fueled by the growth of IoT devices, big data, 5G, and users [11]. Edge-cloud architectures combine the benefits of edge computing and cloud computing to provide a scalable and efficient solution for HEMS. This article serves this purpose by offering readers an in-depth understanding of state-of-the-art edge-cloud architectures, providing a comprehensive landscape view. The key contributions of this work are given as follows.

1) This article provides a detailed analysis of a five-layer edge-cloud architecture designed specifically for HEMS. Delving into the intricacies of each layer, architectural nuances, challenges, and potential solutions are analyzed to provide a comprehensive understanding.
2) It discusses in detail the key performance indicators for evaluating the efficiency and effectiveness of the proposed architecture.
3) In addition, it provides an overview of support methods that address user challenges in HEMS implementation, emphasizing a user-centered approach for successful integration.
4) It also provides an in-depth analysis of privacy issues within HEMS and reviews recent research aimed at mitigating these issues, balancing technological advances, and protecting privacy.
5) To anticipate the evolution of the HEMS landscape, this article outlines key emerging trends.
6) The multifaceted contribution of this work aims to assist stakeholders, researchers, and practitioners in understanding the intricacies of HEMS, fostering innovation, and addressing emerging challenges in energy management.

A. Organization

The organization of the rest of this article is given as follows. Section II examines edge-cloud architecture studies for IoT-enabled HEMS, emphasizing benefits and identifying shortcomings in existing reviews. Section III discusses the research methodology applied in this review. Section IV presents the common approach to IoT systems architecture, and Section V discusses a five-layered edge-cloud architecture for HEMS, as well as IoT cloud services for the five-layered architecture of HEMS, followed by an overview of edge-cloud

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

architecture's performance metrics in Section VI and an overview of support methods for users' concerns regarding HEMS adoption in Section VII. Sections VIII and IX introduce the edge-cloud architectures and edge intelligence for HEMS and blockchain, respectively. Finally, Section X highlights the challenges faced by edge-cloud architecture and future trends, and Section XI gives a conclusion.

Fig. 1 provides a taxonomy outlining this article's context on edge-cloud architectures for HEMS

II. Background

There has been a growing scholarly interest in edge-cloud architecture.

Feng et al. [12] provide a comprehensive review of the integration of edge computing and smart grids. Their review highlights the potential benefits of incorporating edge computing into smart grids, emphasizing improved latency, reduced data traffic, and the paramount importance of security. However, the discussion around privacy protection schemes (including data breaches, encryption, and anonymization) seems somewhat generalized. The authors refrain from delving into specific examples, challenges, and considerations regarding the implementation of these schemes in the energy sector. In addition, while the paper argues that edge computing is more resilient than cloud computing in specific scenarios, it lacks a comparative analysis with alternative solutions or architectures. It is noteworthy that the review does not consider the latest architectural models of peripheral cloud structures developed for smart networks, focusing exclusively on the three-tier model. Given the potential performance and scalability optimizations offered by these newer architectures, their exclusion is a notable gap in the discussion.

Sittón-Candaned et al. [11] explore the potential of integrating edge computing, the Internet of Things (IoT), and social computing to improve energy efficiency in smart energy scenarios. They argue that this integration can lead to a number of benefits, including reduced data transfer from the IoT to the cloud, faster and more real-time responses to energy events, improved security and privacy, and increased user engagement and participation in energy efficiency initiatives. The paper provides a number of examples of how these technologies can be used in smart energy scenarios. The authors briefly mention the need to protect energy-consumption data from unauthorized access and use, but they do not discuss specific security measures or how these measures can be implemented. The authors mention that the edge layer can be used to preprocess data and reduce the amount of data that need to be transferred to the cloud; however, they do not discuss how this can be done in a way that ensures that the edge layer is not overloaded.

Furthermore, Alavikia and Shabro [13] present a comprehensive layered approach for implementing an IoT-enabled smart grid. They define the IoT and the smart grid and then discuss the benefits of their integration. A five-layered architecture for the IoT-enabled smart grid is also presented. The authors then discuss the challenges and opportunities associated with implementing each layer of the architecture. They also provide a review of the state-of-the-art technologies and solutions for each layer. However, the paper does not discuss the security and privacy challenges associated with the IoT-enabled smart grid in much detail. It does not provide a roadmap for the future development of the IoT-enabled smart grid.

Hassija et al. [14] provide a comprehensive overview of the security challenges and solutions for the IoT. It identifies the key security threats to IoT, including data confidentiality, integrity, and availability. The paper also discusses the various solution architectures that have been proposed to address these threats, including blockchain, fog computing, edge computing, and ML. However, it does not discuss specific services that are well-suited for HEMS applications.

Table I summarizes the main contributions of previous comprehensive studies on edge cloud architectures for HEMS. Although many works are devoted to this field, they focus on specific aspects of architecture. Thus, more research and development are needed to address the gaps in the literature and to make technology more widely available.

III. Research Methodology

This review was conducted and completed according to well-established guidelines [19], [20]. The process was divided into three main phases: establishing the research goal, formulating the research questions, and filtering the search results.

A. Research Goal

This article aimed to identify the trends, evolution, and key research areas of edge-cloud architectures for HEMSs. It provides a comprehensive overview of the current state of research in this area.

B. Formulation of the Research Questions

To guide the literature review and ensure that relevant and up-to-date information was gathered, the following research questions were formulated before conducting the review.

RQ1: What are the state-of-the-art edge-cloud architectures for supporting the management of HESs?

Objective: Identify key features and capabilities of these architectures that are essential for effective management of HES within edge-cloud architectures.

RQ2: How can edge-cloud architecture be used to address the security and privacy concerns of HESs?

Objective: Identify the best practices for ensuring data security and user privacy within the context of HEMS.

RQ3: What are the future trends in the use of edge-cloud architectures for HESs?

Objective: Identify emerging technologies and research trends that have the potential to significantly impact the future of edge-cloud architectures for HEMS.

C. Filtering Process

We constructed the search string and defined some of the parameters for our data extraction form. In addition, we manually searched for white papers in the gray literature

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

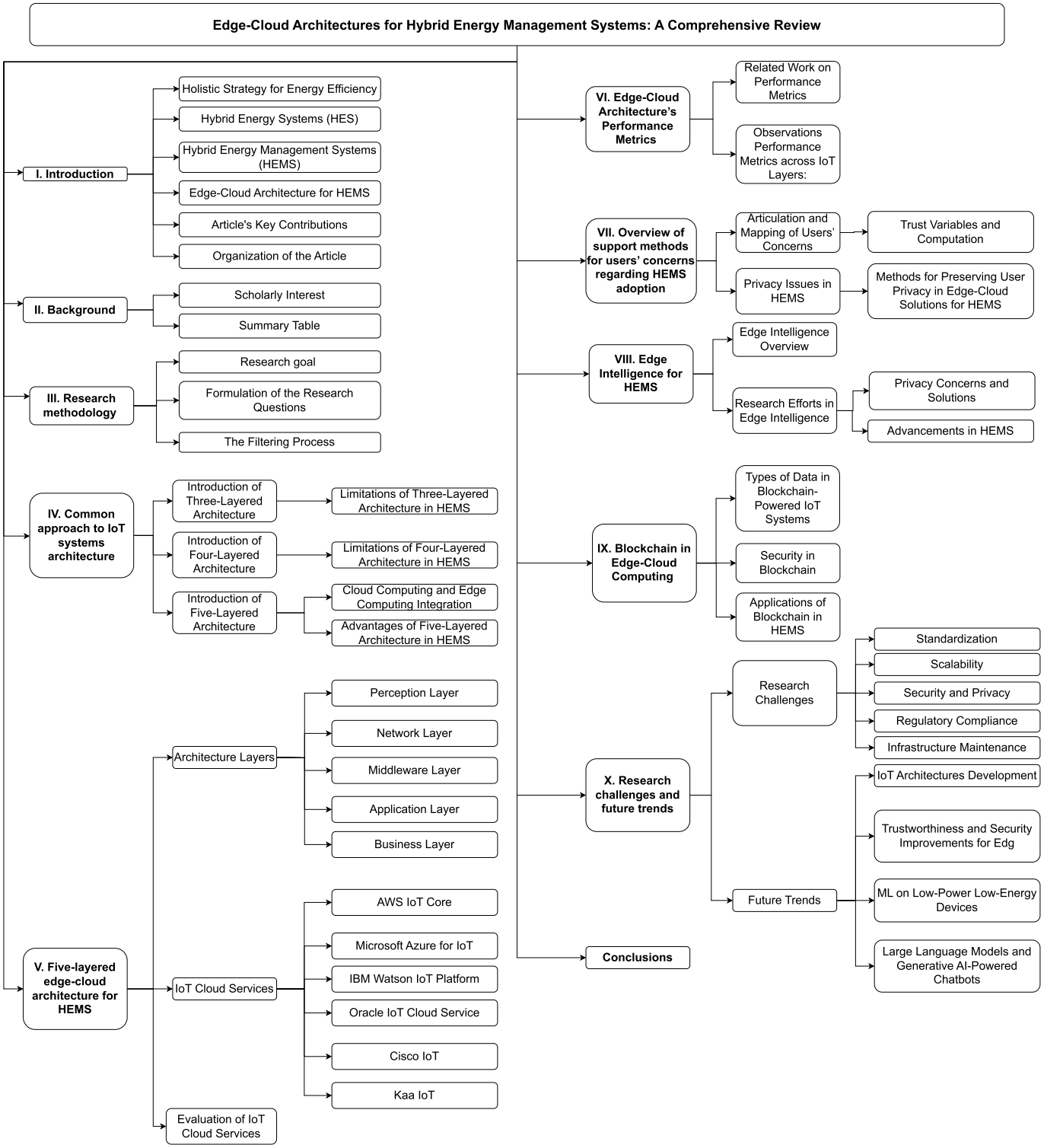


Fig. 1. Taxonomy of article's content.

(e.g., government reports and internal documents from businesses) on Google to identify any relevant information for our study. We also found it helpful to review the papers cited in the selected papers' references. The search strategy was designed to identify all relevant primary studies on the topic of edge-cloud architecture for HESs with user preferences. This was done by searching through a variety of databases, such as ACM, IEEE Xplore, and ScienceDirect. The search strategy included a combination of relevant keywords, such as "edge-cloud architecture," "hybrid energy systems," "user preferences," and "energy management" (see Table II).

Papers Screened: To screen the search results, we read the title and abstract of each paper to determine whether it was relevant to our research questions. If a paper was relevant,

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

TABLE I
COMPARISON OF SURVEYS IN THE EDGE-CLOUD ARCHITECTURES FOR HEMS DOMAIN

Paper	Contribution	Limitation
Feng et al. [12]	A comprehensive review of existing interdisciplinary research in the integration of edge computing and smart grids	Lacks a comparative analysis with alternative solutions or architectures, especially those involving more layers. While identifying security challenges, the paper lacks specific recommendations or solutions to address the outlined security concerns in the context of Smart Grids.
Sittón-Candaned et al. [11]	Explores the potential of integrating edge computing, IoT, and social computing to improve energy efficiency. Introduces a three-layered Edge Computing architecture, offering a clear depiction of each layer's role in data processing, routing, computing operations, and cloud services.	The paper lacks a comprehensive comparison with existing frameworks or architectures. Does not discuss specific security measures or how these measures can be implemented.
Alavikia and Shabro [13]	Presents a comprehensive layered approach for implementing an IoT-enabled smart grid. The paper identifies and discusses specific use cases within the energy system, illustrating how IoT technologies are applied in real-world scenarios.	Does not discuss specific services that are well-suited for five-layerd architecture. The absence of exploration of performance metrics related to Cloud-Edge Architecture specifically tailored for IoT cloud services within a five-layered architecture.
Hassija et al. [14]	The paper identifies and discusses security threats at different layers of IoT applications. The paper addresses application-specific security threats, recognizing that each IoT domain has unique challenges.	The paper briefly touches on the security challenges of energy systems, which is a vast and evolving domain. Does not discuss specific services that are well-suited for HEMS applications.
Talebkhah et al. [15]	A review of the latest advances in IoT, Big Data, and AI technologies. A discussion of the application of these technologies in smart cities. An analysis of the challenges and opportunities associated with the use of these technologies in smart cities.	The paper focuses primarily on data visualization techniques and tools, overlooking other important aspects of IoT and big data, such as data collection, processing, and analysis. The privacy and security considerations are briefly touched upon
Hafizah et al. [16]	The paper introduces a three-layered IoT architecture designed for integration into the smart power grid. It highlights key IoT features, including scalability, heterogeneity, energy efficiency, dynamic nature, and distributed cyber-attack detection through edge computing. Emphasis is placed on implementing security measures across all layers of the IoT architecture, from the bottom to the top layer.	The paper briefly mentions blockchain technology in the context of potential future research trends. However, it falls short of providing an in-depth exploration of how blockchain or edge, fog computing can address security concerns in IoT.
Guerron et al. [17]	Introduces a taxonomy to categorize quality metrics for cloud services. Aligns the taxonomy with established standards (ISO/IEC 25010, NIST SP 800–145, NIST SP 800–146) to enhance its applicability and relevance.	The paper lacks in-depth exploration of emerging technologies like edge computing or blockchain. It refrains from discussing these characteristics within the context of a five-layered architecture.
Barros et al. [18]	This study identifies commonly used IoT IoT platforms and delves into their core components, offering valuable insights for software developers. A heat map effectively summarizes the prevalence of architectural elements across these platforms.	The authors do not mention safety and quality characteristics of identified IoT platforms.

we marked it with "include." If a paper was not relevant, we marked it with "exclude." We also marked papers with "maybe" if we were unsure whether or not they were relevant. In this case, we performed full-text skimming to determine whether a paper was relevant.

Papers Assessed for Eligibility: Once we had screened all of the abstracts, we retrieved the full text of all of the papers marked with "include." Then, we applied the inclusion and exclusion criteria. If a paper met all of our inclusion criteria, we marked it with "include." If it did not meet one or more of our inclusion criteria, we marked it with "exclude."

We used the following *exclusion criteria* for papers to be considered in the review:

1) papers that are not peer-reviewed;
2) papers that are not published in English;
3) papers that do not focus on edge-cloud architecture for HESs;
4) videos and webinars;
5) one-page abstracts, editorials, and poster papers;
6) books and Ph.D. theses.

We applied the following *inclusion criteria* for papers to be considered in the review:

1) papers published between 2019 and October 2023;
2) papers written in English;

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

TABLE II
SEARCH QUERIES USED IN OUR RESEARCH

Repository	Search String	Search results
IEEE	Edge AND Cloud AND "energy management" AND ("hybrid energy system" OR "energy system" OR "smart grid" OR microgrid) AND ("energy consumption" OR "demand side management" OR "demand response" OR "intelligence" OR "load forecasting" OR "consumption forecasting" OR "consumption prediction" or "privacy" or "security") NOT (smart park OR electric vehicle) NOT ("resource allocation" OR "offloading policy")	275
ScienceDirect	Edge AND Cloud AND "energy management" AND ("hybrid energy system" OR "smart grid") AND ("consumption" OR "demand" OR "intelligence" OR "security")	669
ACM	Edge AND Cloud AND "energy management" AND ("hybrid energy system" OR "energy system" OR "smart grid" OR microgrid) AND ("energy consumption" OR "demand side management" OR "demand response" OR "intelligence" OR "load forecasting" OR "consumption forecasting" OR "consumption prediction" or "privacy" or "security")	97

3) papers that are peer-reviewed (i.e., published in a peer-reviewed journal, workshop, or conference);
4) papers available as full text;
5) papers that present original research on edge-cloud architecture for HESs;
6) papers that describe novel architectures, algorithms, or methods for edge-cloud energy management;
7) papers that evaluate the performance of their proposed solutions through simulations or experiments;
8) papers that discuss the challenges and limitations of their proposed solutions.

Papers Included in the Review: Finally, we conducted a detailed analysis of the extracted data and wrote a report of our comprehensive review.

The filtering process of the papers selected for the literature review is illustrated in Fig. 2.

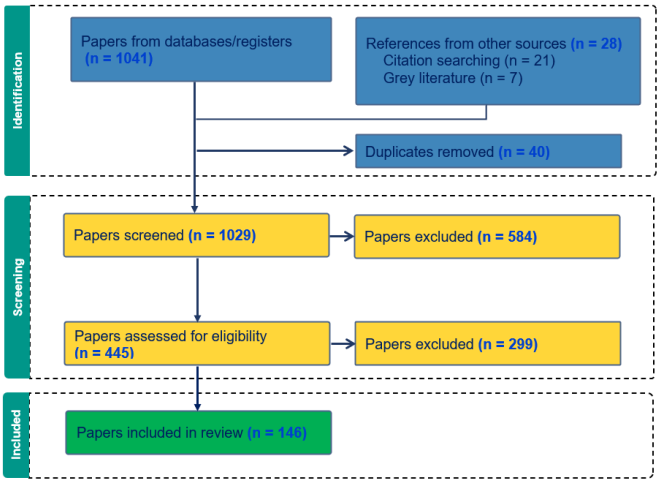


Fig. 2. Filtering process of the studies.

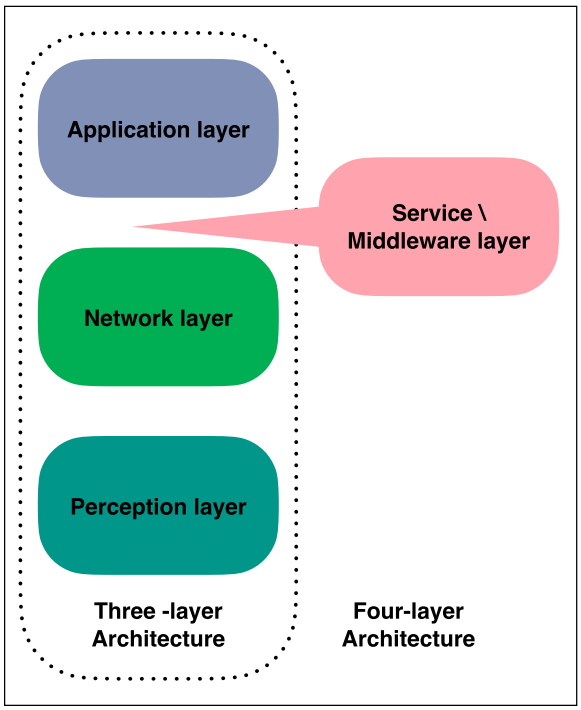


Fig. 3. Three- and four-layered architectures.

IV. COMMON APPROACH TO IOT SYSTEMS ARCHITECTURE

HEMS have strict requirements, including low latency, high scalability, reliability, interoperability, resilience, data security, and data privacy. Since the evolution and development of IoT devices have made them an inalienable part of every modern energy grid, HEMS would benefit from interconnecting with such devices, processing the huge amounts of data that they provide, and making effective operational and management decisions based on them.

The effective interconnection and data processing of millions of heterogeneous IoT devices are a nontrivial task and has been a hot topic for research for the last decade. Researchers have proposed various architectures, but the typical approach is to collect data from smart devices and transfer them to the processing and visualization level. This approach forms a three-layered architecture that consists of sensing, network, and application layers (see Fig. 3). Some researchers have expanded this basic architecture with additional layers such as the service (i.e., middleware) and the business layers to form four- and five-layered architectures [15], [21].

The three-tier architecture can face limitations. Generic three-layered architecture is a simplified approach to IoT systems and a design where raw data from the perception layer are transferred to the application layer though the network layer. This architecture is suitable for simple systems design but lacks details and abstraction levels for more complex systems. For example, if this architecture is used in HEMS

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

development, the perception layer will generate a massive amount of real-time raw data, which will cause high loads on the networking layer. To process these data, the application layer needs a powerful dedicated server or cloud services. In the case of cloud usage, the latency may not be suitable for some tasks that require real-time data processing, such as operational mode selection and network failure handling. Extensive data flows and continuous data processing can have a significant impact on both the networking and computational costs of cloud services. To overcome these challenges, this architecture needs a further description of how to distribute network and computing loads and how to prepare data for efficient usage on the application level. In addition, HEMS must adapt to changes in energy sources and loads. The rigid structure of the three-tier architecture can limit the adaptability of the system.

On the other hand, the four-layered architecture adds a middleware layer that acts as an intermediary between the network layer and the application layer and makes the architecture service-oriented. This layer provides various services that add an abstraction level to simplify application development. The middleware layer allows the architecture to effectively organize data flows and data processing using microservices, virtualization, and edge and fog computing technologies. Each service receives and shares only the data necessary for its functioning, which increases the security level and protects personal information. This concept works well in HEMS as it allows them to provide separate functions for energy consumers and energy providers, protecting their data and supporting a high level of scalability and flexibility.

In an HES, the four-tier architecture can face limitations, especially when integrating various energy sources. As the system expands with more solar panels, energy storage, and other components, the data management and processing load can significantly increase, potentially leading to inefficiencies or errors due to the architecture's strained capabilities [22]. The rigid structure of the four-tier architecture also limits its flexibility, making it challenging to adapt to new devices or services as technology advances and the building's needs change. These limitations underscore the need for more advanced architectural designs, such as five-tier architectures, which offer improved scalability and flexibility [23], [24].

In a five-layer architecture (5LA), an extra layer is added on top of the application layer—a business layer [25]. This layer is focused on high-level business logic, workflow management, and business intelligence. Because HEMS not only focuses on their users' data management but also provides necessary analytics to stakeholders, it is important to include this layer in their architecture.

Cloud computing provides virtually limitless scalability, while edge computing dramatically reduces data processing latency by offloading redundant cloud networking and computation to IoT devices. This approach is effective for all discussed architectures, but four- and five-layered architectures offer a more explicit description of the layers in which it can be applied. In these architectures, the different layers have well-defined responsibilities and interfaces, which allows for clear communication and coordination between the cloud and edge components. Architectures that combine cloud and edge technologies are known as edge-cloud architectures.

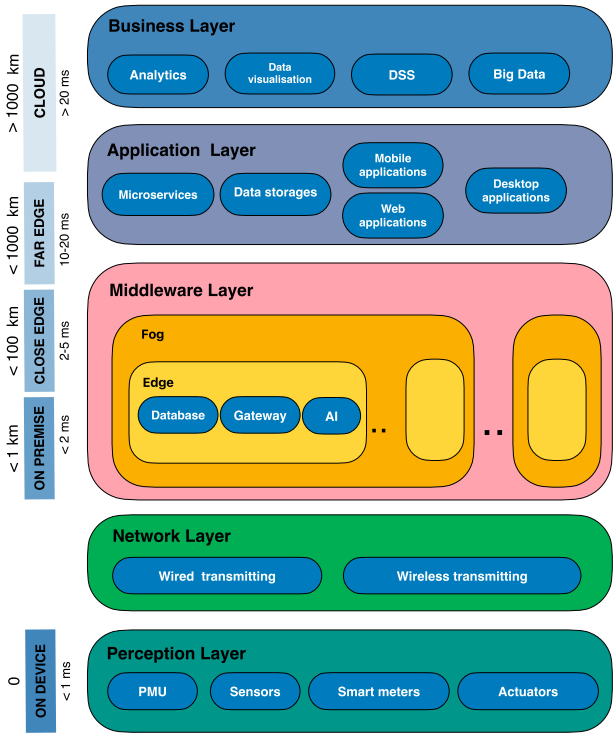


Fig. 4. Five-layered edge-cloud architecture of HEMS.

This article focuses on the five-layered edge-cloud architecture as it better describes HEMS components and their functions. Edge computing, through local data processing, diminishes data transmission delays and lowers the chances of information loss. It supports distributed and decentralized architectures, enhancing network and cloud resource utilization. The collaboration between edge and cloud computing enables real-time management of energy systems, responsiveness to end-users, and improved security and privacy in smart grid operations. This synergy also fosters energy consumption optimization, relieving computational burdens on portable devices and resulting in overall energy savings. In summary, the 5LA effectively encapsulates the essence of HESs by enabling efficient IoT applications, minimizing latency, optimizing energy consumption, and ensuring real-time response and security.

V. Five-Layered Edge-Cloud Architecture for HEMS

The five-layered edge-cloud architecture for HEMS consists of the perception, network, middleware, application, and business layers (see Fig. 4).

In terms of HEMS, the perception layer is the closest layer to the energy generators and consumers and has the least latency. The networking layer is the connection layer between all other layers, so the distance and latency to this layer should fit the requirements of the layers that it connects. The middleware includes computation and data processing nodes that can be located on premises or on the close edge, where the close edge is the infrastructure that is physically close to the

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

perception level devices. The application layer usually requires significant scaling and computation capabilities and uses the fog or cloud part of the infrastructure, but customer-oriented applications can be deployed on premise and close-edge infrastructure for better user data privacy. In turn, the business layer mainly uses cloud infrastructure to be able to save and process consolidated data; provide a holistic picture of the system, its actors, and its components; and help with business strategy planning.

A. Architecture Layers

1) Perception/Sensing Layer: It is the closest layer to the energy generation and consumption parts of the system. It consists of physical devices such as smart meters, sensors, phasor measurement units [26], actuators, and monitoring terminals that collect data about the system's environment and equipment in real time. These devices send and receive their data using local wired or wireless networks [27]. This layer generates a huge amount of raw and heterogeneous data that will be sent to further layers for processing.

2) Network/Transport Layer: It is an intermediate layer for data transfer between other layers. This layer includes both wireless and wired communication technologies [28], depending on the distance to the upper layer nodes. Energy efficiency, latency, bandwidth, and data protection are key parameters for the transport layer. Short-range communication usually includes battery-dependent devices that are more demanding to energy efficiency than the bandwidth, and protocols such as Bluetooth [29] can be used for communication in personal access networks. Ultrawide can offer much higher bandwidth with very low energy consumption levels [30]. Wi-Fi can be used for high-bandwidth personal access networks. For energy-efficient wide-range communication, low-power, wide-area networks, such as ZigBee, LoRa [31], SigFox [32], and NB-IoT [33], can be used. Moreover, 5G [34] networks offer high bandwidth and low latency in wireless networks with a high density of connected devices, which can dramatically improve edge infrastructure performance [35].

3) Middleware Layer: It is responsible for a wide range of functions. Its main function is to unload a cloud part of the infrastructure and bring a part of the computations and networking load on the edge side, closer to the data sources. The edge is a critical component of the IoT because it is responsible for collecting the raw data that are used to power all other layers of the IoT [36]. It allows HEMS to reduce network latency [37], reduce cloud network traffic costs, and keep users' sensitive data close to them. Because of the heterogeneous nature of the data from the perception layer, the middleware layer is responsible for data format standardization, filtering, and aggregation. Only processed data are offloaded to the next layers. The middleware layer controls interoperability on technical, syntax, and semantic levels of the system [15] to support a variety of perception-layer devices. Middleware edge nodes can run data analysis algorithms and AI models to make decisions on operational modes of the system on the edge side to improve overall system resilience and security and to support decentralization [38]. For example, an edge node can control a household's hybrid energy grid, forecast energy consumption levels [39], decide when to use energy, predict energy usage spikes, and balance network load. With the advance of ML technologies for mobile devices, ML models for forecasting energy generation and consumption [40] can be run on a mobile phone, which now can play an edge node role for a household's HEMS. Edge nodes can be grouped using fog nodes to process and aggregate bigger amounts of data [41] for larger groups of hybrid energy producers and customers, but they still keep their sensitive data on their region level to fit legal and data privacy requirements. Edge and fog nodes allow HEMS to scale [42] with less impact on the cloud network and computation costs.

4) Application Layer: It provides services and applications that allow HEMS to implement its main functions. As the middleware layer returns filtered, standardized, and preprocessed data, applications can be designed as thin clients that can be used on devices with limited computational power (e.g., smartphones, smart wearables, and single-board computers). In terms of HEMS, the application layer provides functionality for both energy providers and consumers; namely, it provides energy production- and consumption-level tracking, generated energy quality tracking, consumption spikes prediction, automatic load balancing, grid maintenance scheduling, and others.

Different commercial platforms provide instruments for IoT applications, including Amazon's AWS, Microsoft Azure, Google Cloud IoT [43], Xively, Mozilla WebThings Gateway [21], Cisco IoT Cloud Connect, and Salesforce IoT Cloud.

5) Business Layer: It contains a consolidated data aggregation and storage mechanism for business-level analytics and DMS. This layer includes complex reports, monitoring, and data visualization that can help stakeholders decide on business strategy, monetary policy, and so on. In the case of management systems for hybrid energy grids on the business level, stakeholders can regulate prices on a regional level, analyze network loads to decide on network growth strategies, and evaluate the return on investment, among other things.

B. IoT Cloud Services for Five-Layered Architecture HEMS

Building HEMS from scratch is a complex task, from both infrastructure and software perspectives. As cloud and edge computing with a combination of IoT are highly demanded technologies, there are many types of solutions (e.g., platform as a service and software as a service) offered by both technological giants (e.g., Amazon, Google, and IBM) and smaller companies (e.g., Kaa). Such solutions can help to reduce development costs and time, which are crucial in such a quickly advancing domain as HEMS. A platform decision can break or make the effectiveness of an HEMS. In this section, we present a quick review of popular edge-cloud platforms and their compatibility with the 5LA that was reviewed earlier in this section. Despite the fact that Google IoT Core is one of the most popular platforms for IoT, it will not be included in this review as Google decided to discontinue it starting from August 16, 2023 [44].

1) Amazon Web Services IoT Core (AWS IoT Core): It is a scalable cloud and edge computation platform that allows systems to manage and collect data from IoT devices [45]. Using

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

FreeRTOS as a real-time operation system for embedded systems and IoT devices allows users to build applications with real-time constraints on microcontrollers and microprocessors and use them on IoT devices of the perception level. AWS IoT Core allows systems such as FreeRTOS to register IoT devices on the platform, create their digital twins, and manage devices remotely. Connected devices interconnect with upper layers using MQTT and HTTPS protocols, which are industry standards for organizing IoT perception and networking layers' communication. To fit the middleware requirements, AWS IoT Core includes different message brokers and event handlers. AWS IoT Greengrass allows organizing edge computing, and AWS IoT ExpressLink allows a high-speed secure network connection for edge and on-premises nodes. For the application layer, AWS IoT provides SiteWise to store, process, organize, and visualize collected IoT data. IoT application programming interface (API) help to integrate applications with a variety of available AWS services. The business layer can be covered by AWS IoT Analytics, which helps to visualize aggregated data from cloud storage and provide trends and insights for the decision-making process.

In sum, AWS IoT Core provides all the necessary services and functions to implement HEMS based on a five-layered architecture.

2) Microsoft Azure for IoT: It is a set of cloud services and solutions from Microsoft specifically designed to build applications and systems based on IoT [46]. Azure IoT Hub service suits the perception and networking layers, facilitates the management of IoT devices, and securely connects them to edge or cloud services using different protocols, including MQTT, HTTPS, and specific protocols, such as AMQP for advanced asynchronous communication. Azure IoT Edge allows performing data processing, analysis, and storage locally for real-time and low-latency scenarios. Edge modules can be developed using different programming languages and allow the deployment of containerized modules to edge devices. These services perfectly fit middleware layer needs. For the application layer, Azure IoT provides various services such as Azure App Service, which simplifies web application development, deployment, and scaling. In addition, Azure Kinesis services enable real-time data streaming and analysis. Various database management system (DBMS), such as Azure Cosmos DB and Azure SQL DB, cover data storage functionality. The Time Series Insight service helps users to visualize time-series data and identify trends, anomalies, and patterns. The Notifications Hub provides a flexible messaging service for different front-end platforms. Microsoft PowerBI is an advanced data analysis instrument that allows the building of complex reports for both application and business layers. Azure Data Lake can help users to aggregate and analyze data for the business layer.

All five layers of the architecture can be covered by Azure IoT services, which are well-suited for HEMS development.

3) IBM Watson IoT Platform: It offers cloud and edge computing solutions [47]. It provides industry-specific solutions for automotive, manufacturing, and healthcare. There are no end-to-end solutions for energy management, but IBM offers the IBM Energy and Utilities IoT solution to address some of the specific tasks in energy management that can be used as a part of HEMS. For the perception layer and its communication with the networking layer, the IoT Connection Service of the platform allows the interconnection of IoT devices using MQTT and HTTPS protocols and makes them available for other platform services. The IBM Edge Application Manager covers the middleware requirements of the architecture, facilitates edge computing and edge device management, and allows the secure deployment, management, and monitoring of edge applications on edge devices. For the application layer, IBM Watson offers a wide range of services, such as IBM Watson's Machine Learning, IBM API, and a variety of DBMSs. The IBM Watson IoT platform has powerful analytics services for the business. For example, the IBM DB2 Warehouse analyzes aggregated data from the IBM Cloud, and the IBM Maximo service allows managing and maintaining different types of assets (e.g., facilities and equipment), which can be useful in HEMS.

Thus, IBM Watson is useful for all layers of the five-layered architecture.

4) Oracle IoT Cloud Service: It is a cloud-based platform to organize, collect, and analyze data from IoT devices and sensors [48]. For the perception and network layers, Oracle IoT offers the services of connectors, devices, and digital twins; these services help to build communication between IoT devices and cloud services using MQTT and HTTPS protocols. In addition, Oracle ComputeCloud@Customer and Roving Edge Infrastructure allow users to build secure edge and fog nodes using Oracle hardware and software, which might be not very flexible but is a reliable way to build a middleware layer for HEMS. For the application layer, Oracle IoT Cloud provides a wide range of services: the data flow service allows users to organize real-time data streaming solutions for timely insights and responses. In addition to different data storage services, the Oracle IoT Cloud Service offers data science and ML solutions for deep data analysis and processing. For the business layer data storage and analysis, the platform offers Oracle Business Intelligence, Analytics Cloud, Oracle Fusion Analytics Warehouse, and Oracle BigData Cloud services.

Considering the above information, the Oracle IoT Cloud Service is well-suited for HEMS implementation based on a five-layered architecture.

5) Cisco IoT: It is a platform that offers a variety of services for IoT devices' connectivity, data collection, management, and analysis [49]. As a prominent player in the IoT industry, Cisco provides cloud services, IoT gateways, routers, network equipment, and other products that help to build complex systems based on IoT devices. For the perception layer, Cisco's IoT Gateways allow IoT devices to be organized and interconnected using MQTT, OPC-UA, Modbus, and HTTPS protocols. Moreover, Cisco IoT offers a variety of network hardware and software solutions for building a secure and fast networking layer. For instance, Cisco Edge Cloud and Cisco+ Hybrid Cloud provide both hardware and software for building edge computing infrastructure for the middleware layer. In addition, the Cisco Edge Intelligence orchestration software connects assets at the edge to multicloud application destinations and performs data extraction, transformation,

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

governance, and delivery. For the business layer, the platform offers different products for data aggregation and analytics, such as Cisco Intersight Infrastructure Service and Microsoft Endpoint Analytics. The platform also provides the possibility to store and analyze data using other platforms, such as Microsoft Azure and Amazon AWS.

Some Cisco IoT services and solutions require the use of specific Cisco hardware, but, overall, the platform provides all the necessary services for HEMS implementation based on 5LA.

6) Kaa IoT: It is an open-source platform for IoT device management, data collection, and analysis [50]. It provides services, software, and hardware for IoT systems development. Every feature of Kaa is developed as a customizable microservice collection that can be integrated with third-party systems via open API. Each component can be used as is or can be redesigned and developed from scratch, which gives much higher flexibility than closed commercial platforms. In a five-layered architecture, the perception and network layers can be covered by Kaa device management and Kaa protocol communication service collections. Kaa supports MQTT and HTTPS, and it allows the use of other protocols as custom implementations. It also offers its own Kaa IoT Gateway hardware to connect IoT devices to other layers or systems. For the middleware layer, the data collection adapters can be used to collect, save, and process data from connected IoT devices, and the Configuration Management Extension Service can help to organize edge and fog computing nodes. For the application layer, Kaa provides Web Dashboard and Endpoint Configuration Repository services for data visualization, REST Command Invocation services for controlling IoT devices remotely, and Endpoint Time Series for representing historical data as a time series. For the business layer, Kaa IoT provides cloud storage to aggregate historical data and Open Distro to build solutions for data search and analysis. Although Kaa does not offer as many out-of-the-box services and applications as closed commercial platforms, the flexible architecture of its microservices' collection and provided solutions allow users to cover every requirement for five-layered architecture HEMS.

A comparison of the IoT cloud services available for five-layered architecture HEMS is presented in Table III.

Although compatibility is crucial, it is also necessary to consider a fairly large number of criteria, such as security and privacy capabilities, interoperability, quality of experience (QoE), costs, and others. The evaluation of these criteria is beyond the scope of this article; however, it is a popular topic for studies. Barros et al. [18] analyze the anatomy of IoT platforms, suggest main capabilities for evaluation, and, based on them, define the most important platforms including AWS IoT, Azure IoT, and Watson IoT. Ande et al. [51] provide a comparison of different IoT platforms with a focus on IoT security at all levels of 5LA. Mijuskovic et al. [52] suggest an evaluation framework based on functional and nonfunctional requirements and evaluate five IoT platforms, including AWS IoT, Azure IoT, and Kaa IoT, which were reviewed in this article.

In terms of this article, we were not able to find studies that fully implement a prototype of an HEMS or a similar system using all layers of the 5LA, creating a foundation for future research. Table IV provides a list of studies that developed prototypes for energy management with partial coverage of the 5LA layers using IoT platforms described in this article. Unfortunately, there are no such prototypes implemented using Oracle IoT or Cisco IoT yet or we could not find them.

While the aforementioned studies refrain from delving into the specific limitations or challenges encountered by authors in their utilization of IoT platforms, it is pertinent to acknowledge that overarching limitations may be noted. The lack of standardized protocols across IoT platforms is a recognized challenge in the field. The absence of uniform standards for networking protocols, authentication, and quality of service (QoS) levels poses interoperability issues among a wide variety of connected devices [58]. For example, it can be difficult to integrate devices from different vendors into a single system. However, each of the platforms that we mentioned has its own set of protocols and standards (as detailed in Table III at the network layer). Moreover, IoT platforms encounter obstacles in effectively managing substantial amounts of real-time data, ensuring secure communication from device to cloud, scaling to support an increasing number of devices, and offering user-friendly interfaces. Regarding availability, most leading platforms claim a guaranteed uptime of 95%, while Kaa Technologies assures flexible, no-downtime scalability. Oracle provides 99% uptime monthly for its cloud services and resources, but it is uncertain whether this includes its IoT platform [48]. Most providers make it easier to use their platforms and develop applications by offering documentation and/or coding examples. However, Oracle does not provide public documentation and code, possibly because they prefer to offer application creation as a service to their customers. Similarly, Kaa Technologies does not have publicly available documentation. AWS IoT Core, Microsoft Azure for IoT, IBM Watson IoT, Oracle IoT Cloud Service, and Cisco IoT Cloud Connect offer free tiers with varying capabilities, all transitioning to usage-based pricing after the initial period. These pricing models typically charge for message exchange, data processing, and device connections, requiring careful consideration of individual needs and budget when choosing the most suitable platform. Notably, AWS offers the most generous free tier with 250 000 messages/month, while Microsoft grants 8000 messages/day, and the others provide lower initial quotas. Regarding data storage, Oracle provides the most generous free tier with 10 GB, followed by Cisco's 250 MB. Device connection limits vary, with Oracle and Cisco offering ten devices initially, while others provide fewer. However, the choice of platform would ultimately depend on individual needs and budget, considering the varying capabilities and pricing models of these platforms.

While the existing landscape of IoT cloud services provides options for five-layered architecture HEMS, a critical gap remains in understanding their performance within the context of edge-cloud architecture. This necessitates further investigation into performance metrics specific to this distributed computing model.

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

TABLE III
IoT Cloud Services for Five-Layered Architecture HEMS

Platform	Perception	Network	Middleware	Application	Business	5LA Compatible
AWS IoT Core	AWS IoT Device Management, FreeRTOS	MQTT, HTTPS	AWS IoT Greengrass, ExpressLink, AWS IoT Events	AWS IoT API, AWS Kinesis, AWS Data Stream, Rules, AWS Lambda, AWS IoT SiteWise	AWS IoT Analytics,	Yes
Microsoft Azure for IoT	Azure IoT Hub,	MQTT, AMQP, HTTPS	Azure IoT Hub, Azure IoT Edge,	Azure App Service, Azure Cosmo DB, Azure SQL DB, Rules, Notification Hub, PowerBI	Azure Data Lake, Time series insights, PowerBI	Yes
IBM Watson IoT	IoT Platform Connect	MQTT, HTTPS	IBM Edge Application manager,	IBM Watson Machine Learning, IBM API Connect, Cloudant NoSQL DB, MongoDB	IBM Cloud, IBM DB2 Warehouse, IBM Maximo Monitor	Yes
Oracle IoT Cloud Service	Oracle IoT Connectors, Devices, and Digital Twins	MQTT, HTTPS	Compute Cloud@Customer, Roving Edge Infrastructure	Data Flow, Events, Database Tools, Data Science, ML	Oracle Business Intelligence, Analytics Cloud, Oracle Fusion Analytics Warehouse, Oracle BigData Cloud	Yes
Cisco IoT Cloud Connect	IoT Gateway	MQTT, HTTPS, Modbus, OPC-UA	Cisco Edge Cloud, Cisco+ Hybrid	Cisco Edge Intelligence	Cisco Intersight Infrastructure Service, Microsoft Endpoint Analytics, Cloud integrations	Yes
Kaa IoT	Kaa device management, Kaa protocol communication,	MQTT, HTTPS	Data Collection Adapters, Configuration Management Extension Service	Web Dashboard, REST Command Invocation, Endpoint Time Series, Microsoft Endpoint configuration, Rule engine	Open Distro for Elasticsearch	Yes

TABLE IV
IoT Cloud Services for Five-Layered Architecture HEMS

Paper	Platform	Domain	Covered layers
[53]	AWS IoT	Smart Home Energy Management System	Perception, Network, Middleware, Application
[54]	AWS IoT	Wind Energy Generation Forecast for Microgrids	Perception, Network, Middleware, Application
[55]	Azure IoT	Edge Computing System for A Smart Home	Perception, Network, Middleware, Application
[56]	Watson IoT	Energy Metering System for Effective Pre-Paid Energy Usage	Perception, Network, Middleware, Application
[57]	Kaa	Demand-Side Energy Management in A Smart Home and Smart Homes Community	Perception, Network, Middleware, Application

VI. Edge-Cloud Architecture's Performance Metrics

Performance metrics for edge-cloud computing are important to monitor and improve the overall performance of edge-cloud-based applications and services. Deng et al. [59] state that the quality of application within IoT architecture depends on multiple criteria, including performance, cost, privacy (security), efficiency, and reliability. However, the degree to which each performance metric is considered varies depending on the computing layer. For instance, Aslanpour et al. [60] present a comprehensive review of performance evaluation metrics for cloud, fog, and edge computing and propose a taxonomy of metrics that group them into four categories: latency, throughput, reliability, and

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

cost. The document also discusses the importance of certain metrics for different architecture models. The opinion that the metrics and their features depend on the architectural level is also confirmed by Saovapakhiran et al. [61]. They claim that performance metrics in the context of IoT applications can include network-related metrics such as packet loss, throughput, delay, and reliability. In addition, system-related performance metrics can encompass energy consumption, service access rate, resource usage, processing time, hardware outage, and storage cost. Bashir et al. [62] underscore the significance of energy efficiency within cloud environments, highlighting that the provision of cloud computing services comes at the cost of substantial energy consumption. This heightened energy usage contributes to elevated expenditure budgets, heightened levels of greenhouse gases, and increased CO_2 emissions. Slama [63] reviews the use of intelligent edge computing for prosumer-based smart grids; he identifies key performance metrics for evaluating the performance of edge-cloud computing systems in smart grids for different management tasks. Li et al. [64] consider the tradeoff between different metrics such as latency, energy consumption, and reliability to optimize resource allocation and offloading.

Based on the review, we distinguish the next commonly used performance metrics: latency, throughput, reliability, jitter, packet loss, cost, energy efficiency, scalability, and security. Table V provides a list of performance metrics for edge-cloud architecture, highlighting how metrics can be measured and how their features can be observed across the different IoT layers. For instance, the perception layer involves the collection and transmission of data, where metrics such as latency, reliability, and energy efficiency are crucial [65]. In the network layer, metrics such as throughput, packet loss, and scalability are significant for data transmission and network performance [61]. The middleware layer, responsible for data processing and analysis, requires metrics such as reliability, cost, and security to ensure efficient data management [15]. The application layer, which provides specific services, demands metrics such as latency, security, and scalability to ensure optimal user experiences [18]. Finally, the business layer, which focuses on the strategic aspects, requires metrics such as cost and scalability to ensure efficient resource allocation and business operations [17]. These metrics play a vital role in evaluating and optimizing IoT systems across different layers to ensure their effectiveness and reliability.

VII. Overview of Support Methods for Users' Concerns Regarding HEMS Adoption

Hargreaves et al. [73] claim that the implementation and use of HEMSs often do not meet the expectations of society. This is explained by the fact that society does not always participate in the development and implementation of these systems.

Public concerns regarding the implementation of HEMS include trust, control, privacy, data security [74], profit generation [75], technology control and autonomy [76], [77], health impacts [78], environmental solutions, sustainability [79], equity and equality [80], and inclusiveness in the adoption process solutions [81]. Importantly, the implementation of measures to ensure energy efficiency should contribute to cyber security and ensure that the trust, confidentiality, and protection of end-user data are in accordance with the current EU legislation [1]. Recent research shows that energy data security and privacy issues can hinder energy efficiency efforts, especially in residential energy consumption [82], [83].

When users have concerns about a system, they are less likely to trust it. Thus, user trust is a critical factor for the success of any edge-cloud architecture for HEMS. Users need to trust that the system will protect their privacy and security and it will provide accurate and reliable results [84]. If users have reason to believe that the system does not ensure any of these things, they are less likely to use it. However, trust is an intricate concept that is influenced by a variety of user characteristics and features depending on the context. In the context of IoT, trust can be defined as the confidence that users and stakeholders have in the system's ability to operate as expected and protect their interests [85]. Therefore, trustworthiness encompasses not only performance considerations but also security metrics [86]. Trust in a system can be computed using two main variables: QoS trust and social trust [87]. QoS trust refers to the system's ability to meet the users' expectations in terms of performance, reliability, and security. This can be measured by factors such as packet delivery ratio, load balancing, energy efficiency, latency, bandwidth, and uptime. Social trust plays a pivotal role in promoting user willingness to share energy data with HEMS [10]. This trust is predicated on the system's robust privacy and security measures, which can be assessed based on several factors. Transparency, data minimization, encryption, secure data storage, third-party audits, user education, and transparency in data usage are paramount in fostering user trust and enabling the widespread adoption of HEMS [88].

Articulating and mapping users' concerns are vital because it will help developers take these concerns into account when designing an appropriate architectural solution for an energy management system, thereby instilling trust and developing systems that address users' diverse motivations and their social and environmental concerns without infringing on their privacy. Among the abovementioned general concerns, we singled out those that should be considered in the construction of an edge-cloud solution for HEMS (see Fig. 5). We also highlighted which of these concerns affect different trust variables.

A. Privacy Issues in HEMS and Recent Research

Privacy issues can be a major challenge for HEMS adoption. Users must be confident that their personal and energy data are protected before they share their data and use these systems [63]. The general data protection regulation (GDPR) (Art. 4) [89] defines personal data as any information relating to an identified or identifiable natural person ("data subject"). Energy consumption data are not personal data under this definition because they do not contain direct information about an identified or identifiable natural person. However, GDPR recommends that data on energy consumption and some other types of energy-related data can be treated as personal data. This recommendation concerns data protection in smart grids that are expected to use smart meters to monitor energy consumption.

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

TABLE V
OVERVIEW OF EDGE-CLOUD ARCHITECTURE'S METRICS

Metric	Measurement	Perception Layer	Network Layer	Middleware Layer	Application Layer	Business Layer
Latency [61], [66]	Average latency, 95th percentile latency, 99th percentile latency in milliseconds (ms)	Latency in the perception layer refers to the time it takes for sensor data to be collected and transmitted.	Latency can be measured at the network layer to evaluate the time it takes for data to travel across the network.	Latency in the middleware layer can be assessed in terms of the time it takes for data to be processed and analyzed.	Latency in the application layer refers to the time it takes for applications to respond to user requests.	Latency in the Business Layer can be evaluated in terms of the time it takes for business processes to be executed.
Throughput [14], [64]	Average throughput, peak throughput	Perception layer throughput is assessed by measuring sensor data collection and transmission rates, influenced by application characteristics and data types, including text, images, audio, video, and others.	To measure throughput in the network layer, you would evaluate the rate at which data is transmitted across the network.	Throughput in the middleware layer can be measured to assess the rate at which data is processed and analyzed.	Throughput in the application layer refers to the rate at which applications process and respond to data.	The throughput in the Business Layer can be evaluated in terms of the rate at which business processes are executed.
Reliability [15], [67]	Mean time between failures, mean time to repair	Perception layer reliability can be impacted by node capturing attacks and malicious code injection, which can affect the reliability of data collection and transmission.	Network layer reliability ensures consistent and dependable data transmission across the IoT network, safeguarding data integrity and controlling signal delivery. Attacks on this layer can disrupt data transmission and affect the overall reliability of the IoT system.	Middleware layer reliability hinges on consistent data processing and analysis, ensuring the integrity of information exchanged between the network and application layers. Its vulnerability to attacks underscores the criticality of fortifying this layer to maintain the overall reliability and robustness of IoT applications.	Application layer reliability assesses the consistency of application responses to user requests and interactions. This layer delivers direct services to end users and is susceptible to security vulnerabilities that can compromise user data and privacy, ultimately impacting the reliability of the provided services.	Business layer reliability of IoT systems is crucial for effective operations. It relies on reliable data processing and analysis for decision-making, making its reliability essential for ensuring the overall effectiveness of IoT applications.
Jitter [68], [69]	Average jitter, peak jitter	Jitter at the perception layer can impact the accuracy and timeliness of data collection, potentially affecting the reliability of the entire IoT system.	Jitter at the network layer can lead to delays in data transmission, affecting the overall responsiveness of the IoT system.	Jitter in the middleware layer can impact the processing and analysis of data, potentially affecting the timeliness of feedback decisions and data relay to the cloud.	Jitter at the application layer can affect the responsiveness and reliability of IoT applications, potentially impacting user experience and the effectiveness of IoT services.	Jitter in the Business Layer can impact the reliability and robustness of business services, potentially affecting the overall performance of IoT applications.
Packet loss [61], [70]	Average packet loss, peak packet loss	N/A	Packet loss is not typically measured at all layers except the network layer. It is usually measured to evaluate the rate of packet loss across the network.	N/A	N/A	N/A
Cost [60], [71]	Total cost of ownership, cost per unit of work	The cost may vary based on the variety and quantity of devices and sensors utilized, along with their specifications, and the context of their deployment.	Costs can be influenced by the choice of communication technology (such as Wi-Fi, LoRa, Zigbee), the necessary range and bandwidth, and the intricacy of the network structure.	The cost might be influenced by the IoT platform or middleware solution employed.	Costs could be contingent upon the intricacy of the application, the desired functionalities, and the development procedures involved.	Cost can be measured at the business layer to assess the overall expenses incurred in executing business processes.

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

TABLE V
(Continued.) OVERVIEW OF EDGE-CLOUD ARCHITECTURE'S METRICS

Metric	Measurement	Perception Layer	Network Layer	Middleware Layer	Application Layer	Business Layer
Energy efficiency [61], [64]	Power usage effectiveness, joules per bit	Energy efficiency in the perception layer refers to the evaluation of the energy consumption of sensors and devices used for data collection.	Energy efficiency in the network layer can be measured to assess the energy consumption of network components. This involves evaluating the energy consumption of the networking infrastructure and devices used for data transmission across the network.	Energy efficiency in the middleware layer can be assessed in terms of the energy consumption of data processing and analysis tasks. This involves evaluating the energy efficiency of the computing and storage capabilities provided by the middleware layer.	Energy efficiency in the Application Layer can be measured to determine the energy consumption of applications. This involves assessing the energy consumption of the specific IoT services provided to end users within the Application Layer.	Energy efficiency is not specifically measured at the Business Layer.
Scalability [14], [60], [72]	Horizontal scalability, vertical scalability	Scalability in the perception layer involves the ability to accommodate a large number of sensors and actuators, ensuring that the system can handle the increasing volume of data generated by these devices.	Scalability in the network layer involves the ability to handle increasing data traffic, accommodate a growing number of connected devices, and ensure efficient data transmission across the network.	Scalability in the middleware layer refers to the ability of the middleware layer in a software system to handle an increasing amount of work or requests as the system grows.	Scalability in the application layer refers to the ability to handle an increasing workload by adding resources. Horizontal Scalability – adding more application instances. Vertical Scalability – increasing the capabilities of existing application instances.	Scalability in the business layer refers to the ability to add more resources or nodes to handle increased business data and processing requirements. As well as the ability to enhance the processing power or resources of individual components within the business layer.
Security [14], [64]	Security posture, vulnerability management, compliance	The security in the perception layer includes preventing vulnerability to node capturing attacks, eavesdropping attacks, sleep deprivation attacks, and attacks during the boot process. These challenges highlight the need for robust security measures to ensure the confidentiality, integrity, and availability of data and services in the perception layer.	The security in the network layer includes preventing susceptibility to phishing site attacks, access attacks, DDoS/DoS attacks, and advanced persistent threat (APT) attacks. These security challenges highlight the need for robust security measures to ensure the confidentiality, integrity, and availability of data and services in the network layer.	The security in the middleware layer includes preventing susceptibility to man-in-the-middle attacks, database security challenges, and cloud security challenges. These challenges emphasize the need for robust security measures to ensure the confidentiality, integrity, and availability of data and services in the middleware layer.	The security in the application layer includes preventing issues such as data theft, privacy concerns, and the need for end-to-end encryption to prevent attacks.	The security in the business layer involves addressing challenges in IoT device service management, data integration for business intelligence, and ensuring robust measures for data and service confidentiality, integrity, and availability.

There is debate as to whether the measurement of energy consumption itself increases the likelihood of privacy infringement [90], [91]. For example, there are concerns about smart meters [92], [93] in relation to technical security issues in measuring and storing energy consumption data. Some questions have been raised about communication system vulnerabilities that allow third parties to read data in the current system, vulnerabilities in ICT products and services [94], [95], and vulnerabilities in the data repositories where measurement data are stored [65]. Thus, technical security issues can lead to energy consumption data becoming available to unauthorized persons. This may lead to the following privacy violations that can damage users' trust:

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

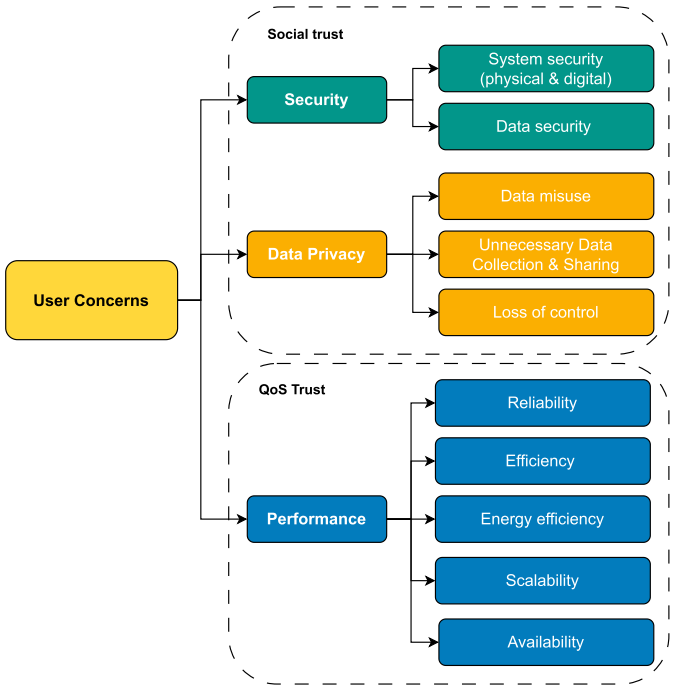


Fig. 5. User concerns about HEMS.

1) *Data Leakage:* Unauthorized persons can gain access to energy consumption data and use it for illegal purposes.
2) *Data Misuse:* Unauthorized persons can use energy consumption data for fraud or other illegal activities.
3) *Data Corruption:* Unauthorized persons can damage or destroy energy consumption data, which can lead to financial losses or other problems for consumers.

More personal information can be obtained from information on energy consumption over time, especially through the use of ML and big data. Many studies have shown that information about personal behavior and characteristics can be obtained from energy data with a certain degree of accuracy. For example, information that can be obtained from data on household electricity consumption includes household characteristics such as number of residents, social class based on main income [96], [97], types of energy appliances in the household [98], and habits of household residents [99]. Therefore, HEMS providers must take steps to address privacy concerns. This can be done by implementing transparent privacy policies, protecting user data, and giving users control over their data.

Due to the relevance of privacy concerns, we investigated how these concerns are addressed in recent research on edge-cloud solutions for HEMS. The analysis revealed two groups of methods for preserving user privacy. The first group of techniques focuses on protecting the privacy of data in transit and at rest. For instance, blockchain encryption or compression methods can be used to protect data from unauthorized access and tampering. Access control can be used to restrict who can access data. Stochastic strategy, generative adversarial network (GAN) Q-learning, encryption, anonymization, and distributed systems can be used to store data on multiple servers in a way that protects privacy.

The second group of techniques involves protecting the privacy of individual data points. Differential privacy adds noise to data to make it less sensitive while still allowing for an accurate analysis. Secure aggregation allows multiple parties to aggregate their data without revealing any individual data points. Moreover, federated learning (FL) allows multiple parties to train an ML model without sharing their data with each other. Finally, secure multiparty computation allows multiple parties to perform a joint analysis of data without revealing their data to each other.

A vast majority of reviewed studies proposed the use of edge computing to implement an HEMS with AI energy demand forecasting. A brief overview of the solutions to users' privacy concerns is presented in Table VI.

VIII. Edge Intelligence for HEMS

In general, recent research on edge computing has highlighted two aspects that demonstrate the mutual benefit between AI and edge computing. First, AI can solve edge computing problems, for example, by optimizing the use of resources, predicting delays, and ensuring data privacy. Second, edge computing can improve the efficiency of AI. Traditional cloud computing can provide powerful computing resources, but it can also lead to significant delays and privacy issues. Edge computing can significantly improve AI performance by allowing AI tasks to be performed closer to users, which reduces latency and increases privacy. This article examines different scenarios for AI at the edge level for HEMS.

Edge intelligence involves implementing computationally intensive intelligent models (such as machine and deep learning models) in edge clouds as opposed to the central cloud [109].

Research efforts in AI on edge have been divided into three categories [59]: framework design, model adaptation, and processor acceleration. The paper highlights the importance of framework design to create the underlying infrastructure for deploying AI models on peripheral devices. Effective infrastructure design is critical to optimizing AI models in HEMS, addressing energy consumption, predictive maintenance, and anomaly detection. Model adaptation, including techniques such as FL, overcomes the limitations of peripheral devices, enabling personalized power management without compromising data privacy. CPU acceleration improves computing performance for real-time decision-making in HEMS by using hardware accelerators such as GPUs and specialized AI chips. Despite the potential benefits, power consumption, cost, and temperature constraints must be carefully considered when integrating accelerators into peripherals. The combined approach of framework development, model adaptation, and processor acceleration has the potential to revolutionize AI-based HEMS by overcoming resource constraints, data privacy issues, and device heterogeneity. These research efforts contribute to sustainable and intelligent energy management.

Han et al. [110] introduce an edge intelligence framework composed of two tiers: the energy management tier and the energy consumption prediction tier. The energy management tier is responsible for the management of current

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

TABLE VI
OVERVIEW OF SUPPORT METHODS FOR USERS' CONCERNS

Ref	Year	Users' Privacy Preserving Techniques	Users' Engagement	Sensitive Data Stored and Processed
[100]	2023	Zero-knowledge proofs and privacy-preserving algorithms. Blockchain encryption. Only authorized parties with decryption keys can access and use the encrypted data.	Customers interact with the system via a user interface and notification system to learn about demand response incentives and make informed energy consumption choices. The system alerts customers to incentives via push notifications, email, SMS, and sound notifications, which customers can personalize.	Data are stored and processed only locally on the blockchain or in non-blockchain storage.
[101]	2023	The article outlines a security and privacy layer within the IoT architecture to safeguard user information. This layer utilizes Wi-Fi and MQTT protocols for secure communication and authentication, and employs name/password protection on web-based interfaces to enhance privacy protection.	The system provides users with real-time monitoring and control of the building environment. The user is provided with a device usage plan and offered an incentive scheme, such as energy discounts.	Data are stored and processed only locally on the user's device and on the fog nodes.
[102]	2020	Implementing data unification, validating metadata, and mapping for clear data source context. Privacy enforcement agents are utilized to mask, deidentify, and encrypt data.	Users can specify their preferences for forecasting accuracy and privacy and can also access and manage their data through a user portal.	Sensors in the smart building collect data. The data is sent to the edge server, which is a computer located within the smart building. Sensitive data undergoes secure storage and processing through a privacy-preserving smart building framework, employing a big data computing architecture that integrates edge and cloud computing.
[103]	2019	Use encryption-sharing algorithms to encrypt data before transmission. Implement web application firewalls to prevent common attacks. Utilize SSLVPN technology to certify unique device and user identities.	Users are not directly involved in the management process.	Sensitive data is stored and processed on a private virtual cloud (VPC) and transmitted over a secure virtual private network (VPN).
[104]	2019	Wi-Fi and MQTT protocols to safeguard communication and data exchange. Name/password protection for web-based interfaces.	The system provides users with real-time monitoring and controlling of building ambience. Users have the option of selecting automatic or manual control for each actuator in the system. Users have the option of selecting automatic or manual control for each actuator in the system.	The data is collected by sensors and sent to the fog computing layer. The fog computing layer aggregates and preprocesses the data before sending it to the cloud computing layer.
[105]	2020	Utilize blockchain-based secure frameworks. Implement secure authenticated key exchange schemes. Federated learning to train machine learning models.	The system provides users with access to their data, so consumers can be informed on possible decisions, such as planning how much power they need to buy and how they can sell power to other users at maximum profit.	Sensitive data is stored and processed locally in fog computing nodes.
[106]	2021	The article does not discuss privacy-preserving techniques but mentions that sensitive data can be anonymized before being used for machine learning.	The system proposes choices to be determined by individuals such as building managers, city authorities, and industry-level managers. These choices encompass decisions about energy regulation and decisions regarding investments in retrofitting endeavors.	The article suggests using Big Data technology to store and analyze sensitive data, such as energy consumption and CO2 emission data. The data is stored in a centralized database with a web application accessible by all managers of public buildings.
[107]	2022	Utilized on-site data storage is disconnected from the Internet and exclusively accessible to end-users within the building's "Intranet."	The system delivers personalized suggestions to end-users to promote responsible energy usage, as well as providing the capability to monitor household appliances. It also supplies users with their energy consumption statistics via a mobile application and presents visualizations of unusual consumption patterns.	Implemented using a local, offline solution for protection from external threats. The system utilizes a No-SQL CouchDB server for storing micro-moment data, user preferences, and power-saving recommendations.
[108]	2019	Data encryption, differential privacy, and secure multi-party computation	The platform analyzes user behavior and makes predictions for future energy demand and healthcare applications. End users can oversee and operate their intelligent residences and devices through various channels, such as web and mobile applications. They can also engage with external vendors through these interfaces.	Sensitive data are stored in three layers: (1) the edge layer, which collects and processes data from IoT devices in real-time and provides a secure gateway to the cloud; (2) the fog layer, which pre-processes and filters data before sending it to the cloud, providing caching and storage to reduce load; and (3) the cloud layer, which stores and analyzes large volumes of data.

energy consumption. The energy consumption prediction tier has a central role in equipping consumers with a resource-constrained device for future energy prediction. The researchers propose a multilayered gated recurrent unit model

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

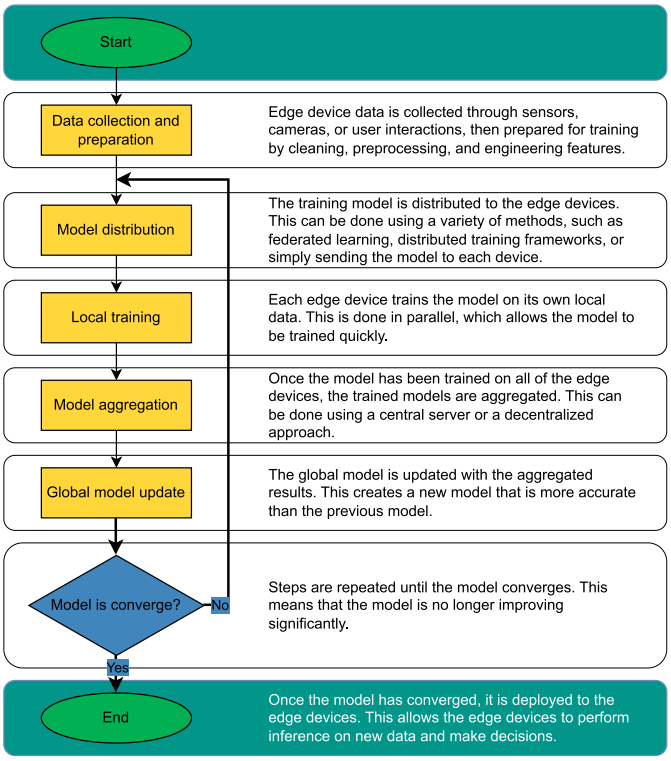


Fig. 6. Overview of how to train and deploy an AI model on edge devices.

for energy forecasting to be employed on edge devices, such as Raspberry Pi. In addition, the authors highlight the challenges and prospects of implementing algorithms at edge nodes of IoT environments to optimize energy use. The main challenges in the field of edge intelligence for energy management in IoT networks involve increasing the accuracy of forecasting, ensuring efficient communication between interconnected devices by executing the algorithm at the edge nodes, and minimizing the time complexity of energy forecasting methods, especially for short-term load forecasting. The paper notes the underutilization of edge-cloud computing paradigms in energy forecasting literature, despite their reliability for effective big data analysis and real-time decision-making.

Training and deployment of an AI model on the edge device process involve several steps. The AI solution for edge mostly comes from iterative learning-based approaches [59]. Fig. 6 provides a general overview of how to train and deploy an AI model on edge devices.

What is more, training edge intelligence models using data stored on end-user devices raises certain privacy concerns. In addition, the distribution of training data among end users is often not independent identically distributed, making it difficult to build reliable models due to the unbalanced nature of the data. In non-independent identically distributed, the samples of data are not independent of each other and do not follow the same probability distribution [111]. This can be a problem for ML because many ML algorithms assume that the training data are independent and identically distributed. For instance, if the data are collected from different users, each user's data may have a different distribution pattern; that is, if we are collecting data on energy consumption from different households, the energy consumption of each household will vary depending on the size of the household, the number of appliances, and other factors.

FL and split learning are promising solutions for distributed learning and can address these concerns [112]. FL allows end users to train a shared model collaboratively without sharing their data with each other or with a central server. Split learning allows end users to split their data into two parts: one part for training a local model and the other part for training a global model. The local models are then aggregated to form the global model, which is then shared with all end users.

Accurate energy consumption forecasts are essential for real-time energy management solutions. However, the complexity and variety of energy data, combined with the dynamic nature of IoT data, pose challenges for building effective forecasting systems. Addressing this challenge, Hudson et al. [109] developed an energy management framework aimed at reducing energy consumption in urban areas. In this framework, the authors put forward two approaches to energy planning based on deep reinforcement learning: edge deep reinforcement learning, in which model training and reasoning tasks are performed at the edge, and cooperative deep reinforcement learning, where the deep reinforcement learning agent is pretrained on the energy cloud server. The pretrained deep neural network weights are then sent to the energy edge server, which operates the deep Q-learning process. Nonetheless, the paper does not discuss the privacy challenges of the proposed framework. Luo et al. [113] experimentally proved that collaboration at the edge of the cloud provides the best results in terms of energy consumption forecasting. The approach of exclusively deploying AI algorithms at the edge showed the second-best outcome, and the method of deploying AI algorithms exclusively in the cloud is a last resort. Another study [114] describes a set of experimental scenarios to assess the performance of energy prediction using FL with edge computing. The scenarios differ in the number and placement of edge nodes (centralized server, one edge-computing node per household, one edge-computing node per neighborhood, and one edge-computing node per building), as well as the amount of data that are shared between the edge nodes and the central server. The study found that the best scenario for FL is having one edge-computing node per neighborhood because this enables the on-site model to gain a deeper understanding of energy consumption patterns. However, the centralized solution results in better predictions. As the level of privacy increases, the accuracy decreases, which, in turn, reduces the amount of data available to train the model. Since the loss of accuracy is negligible in the collaborative approach, its implementation remains viable. While the FL solution does increase network overhead, the total amount of data transferred over the network remains minimal and can be accommodated by various cost-effective wide-area network technologies.

In addition, Han et al. [115] proposed an FL-based IoT scheme for private energy data sharing in smart grids with edge-cloud collaboration. The proposed framework consists of three main components: edge devices, edge servers, and a cloud server. Edge devices are responsible for collecting and

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

preprocessing local energy data. Edge servers are responsible for performing local model training and aggregating the local model updates from edge devices. The cloud server is responsible for aggregating the global model updates from edge servers and generating the global model. To address the non-independent identically distributed data problem, the authors design a local data evaluation mechanism in FL. The authors also propose a two-layer incentive algorithm based on deep reinforcement learning to promote edge devices' participation.

Furthermore, Powroźnik et al. [116] proposed an elastic energy management algorithm using IoT technology for devices with smart appliance functionality. The algorithm works by collecting real-time data from smart appliances using IoT devices. The data are then used to predict future energy consumption and generate an optimal energy management schedule. The schedule considers the user's preferences, the availability of renewable energy sources, and real-time energy prices. The algorithm is elastic in the sense that it can be adapted to changing conditions. For example, if the user changes their preferences or if the availability of renewable energy sources changes, the algorithm can automatically update the energy management schedule. The algorithm was implemented in an edge-intelligent manner by deploying the algorithm on the Raspberry Pi 2 platform.

All the above studies have privacy as a major concern and propose FL-based solutions for energy prediction such as the edge intelligence solution. However, Zhang et al. [117] claim that privacy is also a challenge of FL. They point out that ML algorithms can be implemented at the edge only if the edge nodes have sufficient computational capabilities and storage capacity. They consider cases when the initial gradient for the training task is transmitted from the cloud server to the edge nodes. Edge nodes perform model training using their localized data. After training is complete, these edge nodes transmit the gradients back to the cloud server. However, in such cases, the participants have concerns about the possible leakage of privacy. Although the current functional encryption algorithm in FL offers some computational reduction and protects the privacy of the data provided by each party, it misses information that can be revealed through the participant model weights. The authors suggest the use of multiblockchains as a solution. Each organization would have its own blockchain, and the data would be packaged and sent to the cloud server from the blockchains. This would prevent the cloud server from knowing where the data came from, protecting the privacy of the participants.

In summary, the discussion highlights the significance of edge intelligence in the context of HEMS, showcasing advancements in AI, model training, and energy management. Privacy considerations, addressed through FL and blockchain integration, demonstrate the ongoing efforts to make edge intelligence for HEMS both efficient and secure.

IX. Blockchain in Edge-Cloud Computing

Blockchain [118] is a decentralized way of storing and transmitting information without the need for a central authority. It can be used to store and transfer all kinds of data, including financial records, energy consumption, and more. It is essentially a distributed database where every user on the network has a copy of the data. Any changes to the data must be agreed upon by all users on the network, which makes tampering with the data difficult. Specifically, blockchain works by grouping data into blocks that are then linked into a chain. Each block contains a hash of the previous block, making it difficult to change data without detection. The blocks are also protected by cryptography, hindering attempts to hack into the system.

Blockchain-powered IoT systems generate two types of data: data about the blockchain itself and data from the IoT devices [119]. Blockchain-related data in an energy grid could include the following:

1) account information for each participant in the HES, such as energy producers, consumers, and distributors;
2) balance information for each account, which tracks how much energy has been produced, consumed, and traded;
3) transaction fee information for each transaction that takes place on the HES.

These data can be verified using the same methods as other blockchain applications, such as checking that the energy production from renewable energy sources is greater than or equal to the energy consumption of consumers. This would help to ensure that the HES is sustainable and we are not using more energy than we produce.

The other data type in a blockchain-powered HES, IoT-related data, includes the following:

1) energy production data from solar panels, wind turbines, and various other sources of renewable energy;
2) energy consumption data from smart meters in homes and businesses;
3) HES status data from sensors that monitor the health of the HES.

Each IoT device could sign the data that it sends to the blockchain with its own private key. This would allow anyone to verify that the data have not been tampered with since only the device with the corresponding public key could have generated the signature.

In addition, oracles can be used to verify the correctness of the data from the IoT devices [120], for example, by checking it against data from weather stations or other reliable sources. Blockchain oracles are trusted third-party services that provide data from the outside world to smart contracts on blockchains [121]. Smart contracts are self-executing contracts that can be used to automate a wide range of transactions and agreements. However, smart contracts cannot access data from the outside world on their own, so they rely on oracles to provide them with the necessary data. They act as bridges between the on-chain and off-chain worlds, allowing smart contracts to interact with real-world data [122].

Since smart contracts rely on the accuracy and reliability of the data feeds provided by oracles, the trustworthiness of blockchain oracles is a critical issue [121]. Breiki et al. [123] propose a system for IoT data using blockchain and trusted oracles. The system consists of the following components: smart contracts that are used to store access control policies

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

and to manage access requests; trusted oracles to verify the identity of users and to provide additional data that may be needed to make access control decisions; and a management hub that serves as a bridge connecting IoT devices with the blockchain network. This interface accepts requests from IoT devices, verifies access permissions using smart contracts, and responds to the IoT device by transmitting the requested data (after confirming the permission).

Patil et al. [124] argue that grid-based edge devices suffer from real-time monitoring, management, and security deficiencies. They propose real-time edge computing that processes data from smart grid devices and executes smart contracts with blockchain to facilitate peer-to-peer energy trading between consumers and producers, implement demand response programs, and integrate renewable energy sources more efficiently. Furthermore, they propose a decentralized cooperative energy management framework for a community of smart buildings using blockchain technology. The proposed framework consists of two main components: 1) a decentralized algorithm for day-ahead community power profile optimization and 2) a blockchain-based implementation of the algorithm for secure and transparent communication and billing among the participants. The algorithm works by iteratively coordinating the energy consumption and production of the participating smart buildings to minimize the overall cost of energy and/or maximize renewable energy consumption. The blockchain is used to store the community power profile and facilitate secure and transparent communication and billing among the participants.

Importantly, resource-constrained IoT devices require high computing power (which is typical for HES) [118]. Cryptographic methods in blockchain consume more energy, which is a big problem for blockchain technology. As exemplified by Abed et al. [125], the incorporation of asymmetric encryption methods has been suggested to address security and anonymity concerns within blockchain. However, it is imperative to acknowledge that the utilization of asymmetric encryption methods may exacerbate power consumption and computational requirements, thereby intensifying the overall energy demand. This issue becomes particularly noteworthy in the context of blockchain applications involving IoT devices, which often operate under constrained power resources.

Therefore, Fu et al. [126] argue that achieving energy efficiency requires careful optimization of the allocation of cache memory, computation, and communication resources. They advocate the use of a geometric programming model as an efficient solution to an efficient optimization problem. In addition, the authors argue that strategic tradeoffs between cache capacity, computing power, and communication resources can be strategically executed to maximize the overall energy efficiency of the system.

Despite these challenges, blockchain is a promising technology that can make energy management systems more efficient, reliable, and secure. Its potential contribution is consistent with the continuous evolution and optimization of sustainable energy practices. As researchers continue to improve blockchain's energy efficiency and performance, its role in future energy management will grow significantly.

X. Research Challenges and Future Trends

Within the domain of HEMS, the convergence of edge and cloud computing introduces a number of challenges. This section intricately explores the research challenges intertwined with edge-cloud architecture in the context of HEMS. In addition, we delineate forthcoming trends aimed at addressing specific challenges inherent in the management of these systems.

A. Research Challenges

1) Standardization: The increasing popularity and availability of Do-It-Yourself (DIY) smart home solutions create a large number of possible users that can be connected to the HEMS. DIY smart home devices are often controlled by separate apps, which can be cumbersome and inefficient. However, DIY smart home systems give users complete control over their devices, from configuration to interaction [127]. The perception layer of HEMS may include millions of devices from different manufacturers. These devices have different hardware and software configurations and provide data with different frequencies, precisions, and formats. This adds significant complexity to the middleware layer implementation, which should be able to process and aggregate such data with a high level of compatibility. In particular, IoT devices for HEMS are usually expensive and have a long lifecycle, so the system should provide backward compatibility to different generations of IoT devices. Several organizations—such as the European Telecommunications Standards Institute [128], the Institute of Electrical and Electronics Engineers [129], the Industrial Internet Consortium [130], and the OpenFog Consortium [131]—are working on IoT and edge computing standardization, and while the overall state of IoT standardization is mature, the evolution of edge computing is driven by the standardization and commoditization of hardware and platforms, paralleling the transformative shifts witnessed in the realms of cloud computing and data centers [132].

2) Scalability: The density of IoT devices in smart grids is growing fast, together with the amount of data that these devices can generate. This sets serious requirements for the network bandwidth, latency, and computational power of the edge and fog nodes of HEMS. Edge nodes are limited in all these parameters; therefore, the strategies for data offloading [134], computation task scheduling [123], and resource sharing [135], [136] require advanced algorithms for efficient work.

With 5G networks, HEMS based on edge-cloud architecture could get significant performance improvements [137], reducing communication bottlenecks between edge-cloud layers.

3) Security and Privacy: Edge-cloud architecture has challenges both in physical security and cybersecurity groups [138]. Because of edge computing's nature, it is much harder to control the physical access to and the protection level of edge nodes: edge nodes can be located anywhere and can be exposed to physical intrusion, hardware modification, or damage. As edge-cloud architecture commonly uses wireless communication channels, it is exposed to eavesdropping, data tampering, radio frequency jamming, and other types of

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

attacks. Every layer of the edge-cloud architecture introduces its own possible security issues [139], which should be taken into consideration during HEMS development.

In terms of HEMS, the privacy of users is crucial as information from the system can expose users' personal data, habits, location, and other sensitive data leading to trustworthiness issues. The security and privacy of edge-cloud architecture can be improved by using secure data-sharing schemes with encryption [140], cyberattacks' detection mechanisms using AI models [141], and other tools. This area has been subject to many promising research endeavors.

4) Regulatory Compliance: Government and legal restrictions may have a significant impact on how to transfer, process, and save HEMS users' data in edge-cloud architecture. Data privacy and security requirements could have a significant impact on edge nodes' topology and data storage's physical location. The violation of mandatory regulations may lead to significant penalties. For example, the violation of regulations described in Article 83 of the GDPR leads to a penalty of up to €20 million or up to 4% of the total global annual turnover of the previous financial year, whichever is higher [142]. Restrictions on available radio frequencies can make some wireless communication channels, such as 5G and ultrawideband, unavailable. These complexities may have a negative impact on HEMS scalability, performance, and support costs.

5) Infrastructure Maintenance: HEMSs based on edge-cloud architecture that includes edge and fog nodes require additional efforts in terms of hardware and software. These nodes can be located on the infrastructure side of different owners and have limited physical access. In the case of maintenance outages of edge and fog nodes, HEMS should be able to automatically balance computational loads of connected devices from the perception layer so that users can proceed using the system without issues.

B. Future Trends

HESs play a crucial role in sustainable and optimal energy production, distribution, and consumption. To address the challenges involved in the management of these systems, it is essential to be aware of emerging trends in this field. This section discusses some of these trends.

1) IoT Architectures' Development: As the architecture defines both how HEMSs work and how they are structured with the advancement of IoT technologies, the development of future architectures is unavoidable. There is no set-in-stone standard for edge-cloud IoT architectures yet, and researchers are proposing new visions regarding architecture components, their relations, and their structure. For instance, the seven-layered architecture proposed in [143] has a slightly different layer structure and roles compared to the five-layered architecture reviewed in this article. The proposed seven-level architecture for decentralized energy systems stands out for its comprehensive coverage of functional components, offering a detailed and informative framework, including layers such as user, network, system operator, market, distributed ledger, communication, and regulation, which together relate to various functions and operations in energy transaction mechanisms.

2) Trustworthiness and Security Improvements for Edge: It is another trending topic. Because of the nature of edge computing, it is crucial to build mechanisms for ensuring trust in edge nodes connected to HEMS. The trustworthiness properties of edge computing systems are interrelated, and one way to address the relationships between trustworthiness properties is to use a layered approach that integrates trustworthiness attributes into each layer of the system design [144].

The holistic approach involves considering all of the trustworthiness properties at the same time and making tradeoffs between them when necessary. For example, it may be necessary to sacrifice some performance aspects in order to achieve a higher level of security or safety. This approach can be challenging, as it requires a deep understanding of all of the trustworthiness properties and how they interact with each other. Technologies such as blockchain [85] can solve several trust challenges in IoT-based systems.

The layered approach involves designing and implementing different layers of protection for each trustworthiness property. For example, the security layer might include authentication, authorization, and encryption mechanisms. The safety layer might include fault detection and recovery mechanisms, and the predictability layer might include monitoring and diagnostic mechanisms. This approach is easier to implement than a holistic approach, as it can be broken down into smaller, more manageable pieces. However, it is important to ensure that the different layers are well-integrated and they work together to provide the desired level of trustworthiness.

3) ML on Low-Power Low-Energy Devices: Edge computing can dramatically improve HEMS latency and personal data privacy. With the development of ML models and ML computation hardware, ML models' inference has become possible even on small IoT devices [145], such as smart meters, IoT gateways, and sensors. This opens new opportunities for on-premises or on-edge automatic decision-making on operational mode selection of minigrids, early fault, and threat detections. Moreover, modern smartphones can run even more complex models using CoreML, Tensorflow Lite, and PyTorch Mobile and provide forecasts, trends, and insights for energy grids without offloading data to cloud services.

4) Large Language Models and Generative AI-Powered Chatbots: It can dramatically improve the way HEMSs are developed and used. Such models can help to build proof-of-concept solutions quicker to test new technologies and ideas and reduce time on boilerplate steps implementation. In addition, HEMS can benefit from the analytical abilities of large language models in the business layer to find new IoT data trends and insights [135].

XI. Conclusion

Edge-cloud architecture holds the promise of transforming how we manage and consume energy. By combining the strengths of edge computing and cloud computing, these architectures can provide a more efficient, secure, and privacy-preserving way to manage HESs.

This article provides a comprehensive overview of edge-cloud architectures for HEMSs. It investigates and answers three research questions. It addresses state-of-the-art

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

architectures, their security and privacy concerns, and their future trends. By addressing these questions, this article fills a gap in the literature and provides valuable insights for researchers and practitioners working in the field of HEMS.

Regarding RQ1, this article provides a review of three- and four-layered edge-cloud architectures and a comprehensive review of the five-layered architecture from an HEMS development perspective. The architecture layers are described with consideration of actual technologies and approaches. Six of the most popular IoT cloud platforms are reviewed, and a quick analysis of their proposed services and technologies from the perspective of building an HEMS based on 5LA is provided. All reviewed platforms are able to fit the architecture requirements and can be considered for further in-depth research.

This article discusses how edge-cloud architecture can be used to address the security and privacy concerns of HESs (RQ2). Privacy and trust issues can be a major challenge for HEMS adoption. Users must be confident that sensitive data are protected and their privacy will be respected before they share their data and use these systems.

This article also provides a brief overview of how recent research addresses user concerns regarding privacy in edge-cloud HEMS solutions. We identify two groups of privacy-preserving techniques: 1) techniques focused on protecting the privacy of data in transit and at rest and 2) techniques focused on protecting the privacy of individual data points. We also discuss the potential of FL and blockchain as promising techniques for privacy-preserving edge intelligence in HEMS.

Regarding RQ3, this article identifies several future trends in edge-cloud architectures for HEMS, including the increasing adoption of edge computing and ML on low-power devices, the development of more intelligent and autonomous HEMS, and the use of LLMs and chatbots to develop personalized energy management assistants.

References

[1] *Directive (EU) 2023/1791 of the European Parliament and of the Council of 13 September 2023 on Energy Efficiency and Amending Regulation (EU) 2023/955 (Recast)*. Accessed: Nov. 1, 2023. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AJOL_2023_231_R_0001&qid=1695186598766

[2] *Global Energy Transformation: A Roadmap to 2050*, IRENA, Abu Dhabi, UAE, 2019.

[3] *World Energy Outlook 2022, IEA Paris*, IEA, Paris, France, 2022.

[4] *European Commission, Directorate-General for Energy, Financial Instruments and Models for Energy Production—Investors Dialogue on Energy*, Eur. Commission, Belgium, 2023.

[5] M. Kamran, "Microgrid and hybrid energy systems," in *Fundamentals of Smart Grid Systems*. Amsterdam, The Netherlands: Elsevier, 2023, doi: 10.1016/b978-0-323-99560-3.00006-5.

[6] G. Ulpiani, N. Vetters, D. Shtjefni, G. Kakoulaki, and N. Taylor, "Let's hear it from the cities: On the role of renewable energy in reaching climate neutrality in urban Europe," *Renew. Sustain. Energy Rev.*, vol. 183, Sep. 2023, Art. no. 113444, doi: 10.1016/j.rser.2023.113444.

[7] D. Mariano-Hernández, L. Hernández-Callejo, A. Zorita-Lamadrid, O. Duque-Pérez, and F. S. García, "A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis," *J. Building Eng.*, vol. 33, Jan. 2021, Art. no. 101692, doi: 10.1016/j.jobe.2020.101692.

[8] J. Leitão, P. Gil, B. Ribeiro, and A. Cardoso, "A survey on home energy management," *IEEE Access*, vol. 8, pp. 5699–5722, 2020, doi: 10.1109/ACCESS.2019.2963502.

[9] J. Fernandez, L. del Portillo, and I. Flores, "A novel residential heating consumption characterisation approach at city level from available public data: Description and case study," *Energy Buildings*, vol. 221, Aug. 2020, Art. no. 110082, doi: 10.1016/j.enbuild.2020.110082.

[10] J. D. Godoy, K. Otrel-Cass, and K. H. Toft, "Transformations of trust in society: A systematic review of how access to big data in energy systems challenges Scandinavian culture," *Energy AI*, vol. 5, Sep. 2021, Art. no. 100079, doi: 10.1016/j.egyai.2021.100079.

[11] I. Sittón-Candanedo, R. S. Alonso, Ó. García, L. Muñoz, and S. Rodríguez-González, "Edge computing, IoT and social computing in smart energy scenarios," *Sensors*, vol. 19, no. 15, p. 3353, Jul. 2019, doi: 10.3390/s19153353.

[12] C. Feng, Y. Wang, Q. Chen, Y. Ding, G. Strbac, and C. Kang, "Smart grid encounters edge computing: Opportunities and applications," *Adv. Appl. Energy*, vol. 1, Feb. 2021, Art. no. 100006, doi: 10.1016/j.adapen.2020.100006.

[13] Z. Alavikia and M. Shabro, "A comprehensive layered approach for implementing Internet of Things-enabled smart grid: A survey," *Digit. Commun. Netw.*, vol. 8, no. 3, pp. 388–410, Jun. 2022, doi: 10.1016/j.dcan.2022.01.002.

[14] V. Hassija, V. Chamola, V. Saxena, D. Jain, P. Goyal, and B. Sikdar, "A survey on IoT security: Application areas, security threats, and solution architectures," *IEEE Access*, vol. 7, pp. 82721–82743, 2019, doi: 10.1109/ACCESS.2019.2924045.

[15] M. Talebkhah, A. Sali, M. Gordan, S. J. Hashim, and F. Z. Rokhani, "Comprehensive review on development of smart cities using industry 4.0 technologies," *IEEE Access*, vol. 11, pp. 91981–92030, 2023, doi: 10.1109/ACCESS.2023.3302262.

[16] A. H. Mohd Aman, E. Yadegaridehkordi, Z. S. Attarbashi, R. Hassan, and Y.-J. Park, "A survey on trend and classification of Internet of Things reviews," *IEEE Access*, vol. 8, pp. 111763–111782, 2020, doi: 10.1109/ACCESS.2020.3002932.

[17] X. Guerron, S. Abrahão, E. Insfran, M. Fernández-Diego, and F. González-Ladrón-De-Guevara, "A taxonomy of quality metrics for cloud services," *IEEE Access*, vol. 8, pp. 131461–131498, 2020, doi: 10.1109/ACCESS.2020.3009079.

[18] T. G. F. Barros, E. F. Da Silva Neto, J. A. D. S. Neto, A. G. M. De Souza, V. B. Aquino, and E. S. Teixeira, "The anatomy of IoT platforms—A systematic multivocal mapping study," *IEEE Access*, vol. 10, pp. 72758–72772, 2022, doi: 10.1109/ACCESS.2022.3189660.

[19] G. Pare and S. Kitsiou, "Chapter 9. Methods for literature reviews," in *Handbook of eHealth Evaluation: An Evidence-Based Approach*. Victoria, BC, Canada: Univ. of Victoria, 2016.

[20] M. Templier and G. Paré, "A framework for guiding and evaluating literature reviews," *Commun. Assoc. Inf. Syst.*, vol. 37, no. 1, p. 6, 2015, doi: 10.17705/1cais.03706.

[21] M. Lombardi, F. Pascale, and D. Santaniello, "Internet of Things: A general overview between architectures, protocols and applications," *Information*, vol. 12, no. 2, p. 87, Feb. 2021, doi: 10.3390/info12020087.

[22] J. Zhang and H. Wei, "A review on configuration optimization of hybrid energy system based on renewable energy," *Frontiers Energy Res.*, vol. 10, Aug. 2022, Art. no. 977925, doi: 10.3389/fenrg.2022.977925.

[23] J. K. Verma, D. K. Saxena, V. G. P. Diaz, and V. Shendryk, *Cloud IoT: Concepts, Paradigms, and Applications*. New York, NY, USA: Chapman & Hall, 2022, doi: 10.1201/9781003155577.

[24] B. Mahapatra and A. Nayyar, "Home energy management system (HEMS): Concept, architecture, infrastructure, challenges and energy management schemes," *Energy Syst.*, vol. 13, no. 3, pp. 643–669, Aug. 2022, doi: 10.1007/s12667-019-00364-w.

[25] S. C. Mukhopadhyay, S. K. S. Tyagi, N. K. Suryadevara, V. Piuri, F. Scotti, and S. Zeadally, "Artificial intelligence-based sensors for next generation IoT applications: A review," *IEEE Sensors J.*, vol. 21, no. 22, pp. 24920–24932, Nov. 2021, doi: 10.1109/JSEN.2021.3055618.

[26] M. Hojabri, U. Dersch, A. Papaemmanouil, and P. Bosshart, "A comprehensive survey on phasor measurement unit applications in distribution systems," *Energies*, vol. 12, no. 23, p. 4552, Nov. 2019, doi: 10.3390/en12234552.

[27] C. C. Sobin, "A survey on architecture, protocols and challenges in IoT," *Wireless Pers. Commun.*, vol. 112, no. 3, pp. 1383–1429, Jun. 2020, doi: 10.1007/s11277-020-07108-5.

[28] F. E. Abrahamsen, Y. Ai, and M. Cheffena, "Communication technologies for smart grid: A comprehensive survey," *Sensors*, vol. 21, no. 23, p. 8087, Dec. 2021, doi: 10.3390/s21238087.

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

[29] G. Kalanandhini, A. R. Aravind, G. Vijayalakshmi, J. Gayathri, and K. K. Senthilkumar, "Bluetooth technology on IoT using the architecture of piconet and scatternet," in *Proc. AIP Conf.*, 2022, Art. no. 020121, doi: 10.1063/5.0074188.

[30] A. Bonci, E. Caizer, M. C. Giannini, F. Giuggioloni, and M. Prist, "Ultra wide band communication for condition-based monitoring, a bridge between edge and cloud computing," *Proc. Comput. Sci.*, vol. 217, pp. 1670–1677, Jan. 2023, doi: 10.1016/j.procs.2022.12.367.

[31] M. Shaban, I. Ben Dhaou, M. F. Alsharekh, and M. Abdel-Akher, "Design of a partially grid-connected photovoltaic microgrid using IoT technology," *Appl. Sci.*, vol. 11, no. 24, p. 11651, Dec. 2021, doi: 10.3390/app112411651.

[32] A. Lavric, A. I. Petrariu, and V. Popa, "SigFox communication protocol: The new era of IoT?" in *Proc. Int. Conf. Sens. Instrum. IoT Era (ISSI)*, Aug. 2019, pp. 1–4, doi: 10.1109/ISSI47111.2019.9043727.

[33] M. Kanj, V. Savaux, and M. Le Guen, "A tutorial on NB-IoT physical layer design," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2408–2446, 4th Quart., 2020, doi: 10.1109/COMST.2020.3022751.

[34] G. F. Huseien and K. W. Shah, "A review on 5G technology for smart energy management and smart buildings in Singapore," *Energy AI*, vol. 7, Jan. 2022, Art. no. 100116, doi: 10.1016/j.egyai.2021.100116.

[35] N. Hassan, K. A. Yau, and C. Wu, "Edge computing in 5G: A review," *IEEE Access*, vol. 7, pp. 127276–127289, 2019, doi: 10.1109/ACCESS.2019.2938534.

[36] J. Portilla, G. Mujica, J.-S. Lee, and T. Riesgo, "The extreme edge at the bottom of the Internet of Things: A review," *IEEE Sensors J.*, vol. 19, no. 9, pp. 3179–3190, May 2019, doi: 10.1109/JSEN.2019.2891911.

[37] L. Corneo et al., "(How much) can edge computing change network latency?" in *Proc. IFIP Netw. Conf. (IFIP Networking)*, Jun. 2021, pp. 1–9, doi: 10.23919/IFIPNetworking52078.2021.9472847.

[38] S. S. Ali and B. J. Choi, "State-of-the-art artificial intelligence techniques for distributed smart grids: A review," *Electronics*, vol. 9, no. 6, p. 1030, Jun. 2020, doi: 10.3390/electronics9061030.

[39] A. Lekidis and E. I. Papageorgiou, "Edge-based short-term energy demand prediction," *Energies*, vol. 16, no. 14, p. 5435, Jul. 2023, doi: 10.3390/en16145435.

[40] S. Ahmed and M. Bons, "Edge computed NILM: A phone-based implementation using MobileNet compressed by tensorflow lite," in *Proc. 5th Int. Workshop Non-Intrusive Load Monitor.*, Nov. 2020, pp. 44–48, doi: 10.1145/3427771.3427852.

[41] J. Li, C. Gu, Y. Xiang, and F. Li, "Edge-cloud computing systems for smart grid: State-of-the-art, architecture, and applications," *J. Modern Power Syst. Clean Energy*, vol. 10, no. 4, pp. 805–817, Jul. 2022, doi: 10.35833/MPCE.2021.000161.

[42] Q. N. Minh, V.-H. Nguyen, V. K. Quy, L. A. Ngoc, A. Chehri, and G. Jeon, "Edge computing for IoT-enabled smart grid: The future of energy," *Energies*, vol. 15, no. 17, p. 6140, Aug. 2022, doi: 10.3390/en15176140.

[43] P. Pierleoni, R. Concetti, A. Belli, and L. Palma, "Amazon, Google and Microsoft solutions for IoT: Architectures and a performance comparison," *IEEE Access*, vol. 8, pp. 5455–5470, 2020, doi: 10.1109/ACCESS.2019.2961511.

[44] *Google IoT Core*. Accessed: Nov. 5, 2023. [Online]. Available: https://cloud.google.com/iot-core

[45] *Amazon IoT Core*. Accessed: Nov. 5, 2023. [Online]. Available: https://aws.amazon.com/iot-core/

[46] *Microsoft Azure IoT*. Accessed: Nov. 5, 2023. [Online]. Available: https://azure.microsoft.com/en-us/solutions/iot

[47] *IBM Watsons IoT*. Accessed: Nov. 5, 2023. [Online]. Available: https://internetofthings.ibmcloud.com

[48] *Oracle IoT Cloud Service*. Accessed: Nov. 5, 2023. [Online]. Available: https://docs.oracle.com/en/cloud/paas/iot-cloud/index.html

[49] *Cisco IoT*. Accessed: Nov. 5, 2023. [Online]. Available: https://www.cisco.com/c/en/us/solutions/internet-of-things/overview.html

[50] *Kaa IoT*. Accessed: Nov. 5, 2023. [Online]. Available: https://www.kaaiot.com/products/overview

[51] R. Ande, B. Adebisi, M. Hammoudeh, and J. Saleem, "Internet of Things: Evolution and technologies from a security perspective," *Sustain. Cities Soc.*, vol. 54, Mar. 2020, Art. no. 101728, doi: 10.1016/j.scs.2019.101728.

[52] A. Mijuskovic, I. Ullah, R. Bemthuis, N. Meratnia, and P. Havinga, "Comparing apples and oranges in IoT context: A deep dive into methods for comparing IoT platforms," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1797–1816, Feb. 2021, doi: 10.1109/JIOT.2020.3016921.

[53] F. Condon, J. M. Martínez, A. M. Eltamaly, Y.-C. Kim, and M. A. Ahmed, "Design and implementation of a cloud-IoT-based home energy management system," *Sensors*, vol. 23, no. 1, p. 176, Dec. 2022, doi: 10.3390/s23010176.

[54] C. A. Marino, F. Chinelato, and M. Marufuzzaman, "AWS IoT analytics platform for microgrid operation management," *Comput. Ind. Eng.*, vol. 170, Aug. 2022, Art. no. 108331, doi: 10.1016/j.cie.2022.108331.

[55] A. H. Shehab and S. T. Faraj Al-Janabi, "Microsoft Azure IoT-based edge computing for smart homes," in *Proc. Int. Conf. Decis. Aid Sci. Appl. (DASA)*, Nov. 2020, pp. 315–319, doi: 10.1109/DASA51403.2020.9317274.

[56] S. Goswami, A. Sarkar, and S. Sengupta, "Design and implementation of a prepaid energy meter using IBM IoT," in *Proc. Michael Faraday IET Int. Summit (MFIIS)*, Oct. 2020, pp. 219–223, doi: 10.1049/icp.2021.1154.

[57] M. Mehr Nezhad, M. H. Y. Moghaddam, and M. Asadi, "Joint peak clipping and load scheduling based on user behavior monitoring in an IoT platform," *IEEE Syst. J.*, vol. 15, no. 1, pp. 1202–1213, Mar. 2021, doi: 10.1109/JSYST.2020.3009699.

[58] M. Goudarzi, S. Ilager, and R. Buyya, "Cloud computing and Internet of Things: Recent trends and directions," in *New Frontiers in Cloud Computing and Internet of Things*. Cham, Switzerland: Springer, 2022, doi: 10.1007/978-3-031-05528-7_1.

[59] S. Deng, H. Zhao, W. Fang, J. Yin, S. Dustdar, and A. Y. Zomaya, "Edge intelligence: The confluence of edge computing and artificial intelligence," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7457–7469, Aug. 2020, doi: 10.1109/JIOT.2020.2984887.

[60] M. S. Aslanpour, S. S. Gill, and A. N. Toosi, "Performance evaluation metrics for cloud, fog and edge computing: A review, taxonomy, benchmarks and standards for future research," *Internet Things*, vol. 12, Dec. 2020, Art. no. 100273, doi: 10.1016/j.iot.2020.100273.

[61] B. Saovapakhiran, W. Naruephiphat, C. Charnsripinyo, S. Baydere, and S. Özdemir, "QoE-driven IoT architecture: A comprehensive review on system and resource management," *IEEE Access*, vol. 10, pp. 84579–84621, 2022, doi: 10.1109/ACCESS.2022.3197585.

[62] S. Bashir, S. Mustafa, R. W. Ahmad, J. Shuja, T. Maqsood, and A. Alourani, "Multi-factor nature inspired SLA-aware energy efficient resource management for cloud environments," *Cluster Comput.*, vol. 26, no. 2, pp. 1643–1658, Apr. 2023, doi: 10.1007/s10586-022-03690-4.

[63] S. Ben Slama, "Prosumer in smart grids based on intelligent edge computing: A review on artificial intelligence scheduling techniques," *Ain Shams Eng. J.*, vol. 13, no. 1, Jan. 2022, Art. no. 101504, doi: 10.1016/j.asej.2021.05.018.

[64] Q. Li, Y. Deng, W. Sun, and W. Li, "Communication and computation resource allocation and offloading for edge intelligence enabled fault detection system in smart grid," in *Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm)*, Nov. 2020, pp. 1–7, doi: 10.1109/SmartGridComm47815.2020.9302930.

[65] K. D. Kumar, M. A. Jawale, M. Sujith, and D. B. Pardeshi, "Cybersecurity threats, detection methods, and prevention strategies in smart grid: Review," in *Proc. 3rd Int. Conf. Artif. Intell. Smart Energy (ICAIS)*, Feb. 2023, pp. 1609–1614, doi: 10.1109/ICAIS56108.2023.10073843.

[66] X. Chen et al., "Achieving low tail-latency and high scalability for serializable transactions in edge computing," in *Proc. 16th Eur. Conf. Comput. Syst.*, Apr. 2021, pp. 210–227, doi: 10.1145/3447786.3456238.

[67] R. Liu, R. Yang, Z. Wang, and X. Sun, "Application of edge computing in smart grid," in *Proc. 3rd Int. Conf. Big Data, Artif. Intell. Internet Things Eng. (ICBAIE)*, Jul. 2022, pp. 62–65, doi: 10.1109/ICBAIE56435.2022.9985874.

[68] Y. Mansouri and M. A. Babar, "A review of edge computing: Features and resource virtualization," *J. Parallel Distrib. Comput.*, vol. 150, pp. 155–183, Apr. 2021, doi: 10.1016/j.jpdc.2020.12.015.

[69] A. F. R. Trajano, A. A. M. de Sousa, E. B. Rodrigues, J. N. de Souza, A. D. C. Callado, and E. F. Coutinho, "Leveraging mobile edge computing on smart grids using LTE cellular networks," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jun. 2019, pp. 1–7, doi: 10.1109/ISCC47284.2019.8969784.

[70] M. Chen, Y. Miao, H. Gharavi, L. Hu, and I. Humar, "Intelligent traffic adaptive resource allocation for edge computing-based 5G networks," *IEEE Trans. Cognit. Commun. Netw.*, vol. 6, no. 2, pp. 499–508, Jun. 2020, doi: 10.1109/TCCN.2019.2953061.

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

[71] S. H. Al-Awami, M. Mahfud Al-Aty, and M. F. Al-Najar, "Comparison of IoT architectures based on the seven essential characteristics," in *Proc. IEEE 3rd Int. Maghreb Meeting Conf. Sci. Techn. Autom. Control Comput. Eng. (MI-STA)*, May 2023, pp. 305–310, doi: 10.1109/MI-STA57575.2023.10169289.

[72] S. Sotiriadis, N. Bessis, C. Amza, and R. Buyya, "Elastic load balancing for dynamic virtual machine reconfiguration based on vertical and horizontal scaling," *IEEE Trans. Services Comput.*, vol. 12, no. 2, pp. 319–334, Mar. 2019, doi: 10.1109/TSC.2016.2634024.

[73] N. Hargreaves, T. Hargreaves, and J. Chilvers, "Socially smart grids? A multi-criteria mapping of diverse stakeholder perspectives on smart energy futures in the united kingdom," *Energy Res. Social Sci.*, vol. 90, Aug. 2022, Art. no. 102610, doi: 10.1016/j.erss.2022.102610.

[74] M. Talebkhah, A. Sali, M. Marjani, M. Gordan, S. J. Hashim, and F. Z. Rokhani, "IoT and big data applications in smart cities: Recent advances, challenges, and critical issues," *IEEE Access*, vol. 9, pp. 55465–55484, 2021, doi: 10.1109/ACCESS.2021.3070905.

[75] P. Rodriguez-Garcia, Y. Li, D. Lopez-Lopez, and A. A. Juan, "Strategic decision making in smart home ecosystems: A review on the use of artificial intelligence and Internet of Things," *Internet Things*, vol. 22, Jul. 2023, Art. no. 100772, doi: 10.1016/j.iot.2023.100772.

[76] M. A. Hassanain, M. O. Sanni-Anibire, and A. S. Mahmoud, "An assessment of users' satisfaction with a smart building on university campus through post-occupancy evaluation," *J. Eng., Des. Technol.*, vol. 3, pp. 1–17, May 2022, doi: 10.1108/jedt-12-2021-0714.

[77] Y. Strengers, M. Hazas, L. Nicholls, J. Kjeldskov, and M. B. Skov, "Pursuing pleasance: Interrogating energy-intensive visions for the smart home," *Int. J. Hum.-Comput. Stud.*, vol. 136, Apr. 2020, Art. no. 102379, doi: 10.1016/j.ijhcs.2019.102379.

[78] H. Sequeiros, T. Oliveira, and M. A. Thomas, "The impact of IoT smart home services on psychological well-being," *Inf. Syst. Frontiers*, vol. 24, no. 3, pp. 1009–1026, Jun. 2022, doi: 10.1007/s10796-021-10118-8.

[79] F. Shirani, K. O'Sullivan, R. Hale, N. Pidgeon, and K. Henwood, "Transformational innovation in home energy: How developers imagine and engage with future residents of low carbon homes in the united kingdom," *Energy Res. Social Sci.*, vol. 91, Sep. 2022, Art. no. 102743, doi: 10.1016/j.erss.2022.102743.

[80] J. Beneicke, A. A. Juan, F. Xhafa, D. Lopez-Lopez, and A. Freixes, "Empowering Citizens' cognition and decision making in smart sustainable cities," *IEEE Consum. Electron. Mag.*, vol. 9, no. 1, pp. 102–108, Jan. 2020, doi: 10.1109/MCE.2019.2941457.

[81] Y. Yang, J. Hao, S. Wang, Z. Ge, J. Sun, and K. Wu, "Configuration and performance analysis of clean heating systems based on distributed energy conversion technology," in *Proc. 4th Int. Conf. Smart Power Internet Energy Syst. (SPIES)*, Dec. 2022, pp. 415–419, doi: 10.1109/SPIES55999.2022.10082125.

[82] K. M. Ramokapane, C. Bird, A. Rashid, and R. Chitchyan, "Privacy design strategies for home energy management systems (HEMS)," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Apr. 2022, pp. 1–15, doi: 10.1145/3491102.3517515.

[83] E. T. Tchao et al., "On cloud-based systems and distributed platforms for smart grid integration: Challenges and prospects for Ghana's grid network," *Sci. Afr.*, vol. 12, Jul. 2021, Art. no. e00796, doi: 10.1016/j.sciaf.2021.e00796.

[84] A. Llaria, J. Dos Santos, G. Terrasson, Z. Boussaada, C. Merlo, and O. Curea, "Intelligent buildings in smart grids: A survey on security and privacy issues related to energy management," *Energies*, vol. 14, no. 9, p. 2733, May 2021, doi: 10.3390/en14092733.

[85] L. Fotia, F. Delicato, and G. Fortino, "Trust in edge-based Internet of Things architectures: State of the art and research challenges," *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–34, Sep. 2023, doi: 10.1145/3558779.

[86] V. T. Le, N. El Ioini, C. Pahl, H. R. Barzegar, and C. Ardagna, "A distributed trust layer for edge infrastructure," in *Proc. 6th Int. Conf. Fog Mobile Edge Comput. (FMEC)*, Dec. 2021, pp. 1–8, doi: 10.1109/FMEC54266.2021.9732606.

[87] J. Al Muhtadi, R. A. Alamri, F. A. Khan, and K. Saleem, "Subjective logic-based trust model for fog computing," *Comput. Commun.*, vol. 178, pp. 221–233, Oct. 2021, doi: 10.1016/j.comcom.2021.05.016.

[88] M. H. Panahi Rizi and S. A. Hosseini Seno, "A systematic review of technologies and solutions to improve security and privacy protection of citizens in the smart city," *Internet Things*, vol. 20, Nov. 2022, Art. no. 100584, doi: 10.1016/j.iot.2022.100584.

[89] *The Regulation (EU) 2016/679 (General Data Protection Regulation)*. Accessed: Nov. 10, 2023. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

[90] T. Jakobi, S. Patil, D. Randall, G. Stevens, and V. Wulf, "It is about what they could do with the data: A user perspective on privacy in smart metering," *ACM Trans. Comput.-Hum. Interact.*, vol. 26, no. 1, pp. 1–44, Feb. 2019, doi: 10.1145/3281444.

[91] V. von Loessl, "Smart meter-related data privacy concerns and dynamic electricity tariffs: Evidence from a stated choice experiment," *Energy Policy*, vol. 180, Sep. 2023, Art. no. 113645, doi: 10.1016/j.enpol.2023.113645.

[92] D. Bacnar, L. Leytner, R. Prenc, V. Jardas, and J. Lerga, "On security and privacy in smart metering systems," in *Proc. 7th Int. Conf. Smart Sustain. Technol. (SpliTech)*, Jul. 2022, pp. 1–6, doi: 10.23919/SpliTech55088.2022.9854239.

[93] D. Bugden and R. Stedman, "A synthetic view of acceptance and engagement with smart meters in the United States," *Energy Res. Social Sci.*, vol. 47, pp. 137–145, Jan. 2019, doi: 10.1016/j.erss.2018.08.025.

[94] H. Bahsi, H. O. Dola, S. M. Khalil, and T. Korotko, "A cyber attack taxonomy for microgrid systems," in *Proc. 17th Annu. Syst. Syst. Eng. Conf. (SOSE)*, Jun. 2022, pp. 324–331, doi: 10.1109/SOSE55472.2022.9812642.

[95] M. Sanduleac, C. Ionescu, A. Mandis, V. Gropa, C. Efremov, and V. Sanduleac, "Solutions for digital interaction of a resilient energy community in a service-oriented framework," in *Proc. Int. Conf. Expo. Electr. Power Eng. (EPE)*, Oct. 2022, pp. 1–6, doi: 10.1109/EPE56121.2022.9975792.

[96] P. Grünewald and M. Diakonova, "The specific contributions of activities to household electricity demand," *Energy Buildings*, vol. 204, Dec. 2019, Art. no. 109498, doi: 10.1016/j.enbuild.2019.109498.

[97] Y. Losa Rovira, A. Faghih Imani, A. Sivakumar, and J. Pawlak, "Do in-home and virtual activities impact out-of-home activity participation? Investigating end-user activity behaviour and time use for residential energy applications," *Energy Buildings*, vol. 257, Feb. 2022, Art. no. 111764, doi: 10.1016/j.enbuild.2021.111764.

[98] Y. Liu, W. Liu, Y. Shen, X. Zhao, and S. Gao, "Toward smart energy user: Real time non-intrusive load monitoring with simultaneous switching operations," *Appl. Energy*, vol. 287, Apr. 2021, Art. no. 116616, doi: 10.1016/j.apenergy.2021.116616.

[99] M. Ul Hassan, M. H. Rehmani, R. Kotagiri, J. Zhang, and J. Chen, "Differential privacy for renewable energy resources based smart metering," *J. Parallel Distrib. Comput.*, vol. 131, pp. 69–80, Sep. 2019, doi: 10.1016/j.jpdc.2019.04.012.

[100] F. Iqbal et al., "Blockchain-modeled edge-computing-based smart home monitoring system with energy usage prediction," *Sensors*, vol. 23, no. 11, p. 5263, Jun. 2023, doi: 10.3390/s23115263.

[101] S. S. Reka, P. Venugopal, V. Ravi, and T. Dragicevic, "Privacy-based demand response modeling for residential consumers using machine learning with a cloud-fog-based smart grid environment," *Energies*, vol. 16, no. 4, p. 1655, Feb. 2023, doi: 10.3390/en16041655.

[102] C. Inibhunu and C. McGregor, "Edge computing with big data cloud architecture: A case study in smart building," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2020, pp. 3387–3393, doi: 10.1109/BigData50022.2020.9377918.

[103] D. Yang, H. Wei, Y. Zhu, P. Li, and J.-C. Tan, "Virtual private cloud based power-dispatching automation system—Architecture and application," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1756–1766, Mar. 2019, doi: 10.1109/TII.2018.2849005.

[104] G. Alsuhli and A. Khattab, "A fog-based IoT platform for smart buildings," in *Proc. Int. Conf. Innov. Trends Comput. Eng. (ITCE)*, Feb. 2019, pp. 174–179, doi: 10.1109/ITCE.2019.8646480.

[105] R. Jaiswal, R. Davidrajuh, and C. Rong, "Fog computing for realizing smart neighborhoods in smart grids," *Computers*, vol. 9, no. 3, p. 76, Sep. 2020, doi: 10.3390/computers9030076.

[106] M. Zekic-Susac, S. Mitrovic, and A. Has, "Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities," *Int. J. Inf. Manag.*, vol. 58, Jun. 2021, Art. no. 102074, doi: 10.1016/j.ijinfomgt.2020.102074.

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

[107] Y. Himeur et al., "Techno-economic assessment of building energy efficiency systems using behavioral change: A case study of an edge-based micro-moments solution," *J. Cleaner Prod.*, vol. 331, Jan. 2022, Art. no. 129786, doi: 10.1016/j.jclepro.2021.129786.

[108] A. Yassine, S. Singh, M. S. Hossain, and G. Muhammad, "IoT big data analytics for smart homes with fog and cloud computing," *Future Gener. Comput. Syst.*, vol. 91, pp. 563–573, Feb. 2019, doi: 10.1016/j.future.2018.08.040.

[109] N. Hudson, M. J. Hossain, M. Hosseinzadeh, H. Khamfroush, M. Rahnamay-Naeini, and N. Ghani, "A framework for edge intelligent smart distribution grids via federated learning," in *Proc. Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul. 2021, pp. 1–9, doi: 10.1109/ICCCN52240.2021.9522360.

[110] T. Han, K. Muhammad, T. Hussain, J. Lloret, and S. W. Baik, "An efficient deep learning framework for intelligent energy management in IoT networks," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3170–3179, Mar. 2021, doi: 10.1109/JIOT.2020.3013306.

[111] M. Arafeh, A. Hammoud, H. Otrok, A. Mourad, C. Talhi, and Z. Dziong, "Independent and identically distributed (IID) data assessment in federated learning," in *Proc. IEEE Global Commun. Conf.*, Dec. 2022, pp. 293–298, doi: 10.1109/GLOBECOM48099.2022.10001718.

[112] C. Thapa, M. A. P. Chamikara, and S. A. Camtepe, "Advancements of federated learning towards privacy preservation: From federated learning to split learning," in *Studies in Computational Intelligence*, vol. 965. Cham, Switzerland: Springer, 2021, doi: 10.1007/978-3-030-70604-3_4.

[113] H. Luo, H. Cai, H. Yu, Y. Sun, Z. Bi, and L. Jiang, "A short-term energy prediction system based on edge computing for smart city," *Future Gener. Comput. Syst.*, vol. 101, pp. 444–457, Dec. 2019, doi: 10.1016/j.future.2019.06.030.

[114] E. Petrangeli, N. Tonellotto, and C. Vallati, "Performance evaluation of federated learning for residential energy forecasting," *IoT*, vol. 3, no. 3, pp. 381–397, Sep. 2022, doi: 10.3390/iot3030021.

[115] Z. Su et al., "Secure and efficient federated learning for smart grid with edge-cloud collaboration," *IEEE Trans. Ind. Informat.*, vol. 18, no. 2, pp. 1333–1344, Feb. 2022, doi: 10.1109/TII.2021.3095506.

[116] P. Powroznik, P. Szczesniak, and K. Piotrowski, "Elastic energy management algorithm using IoT technology for devices with smart appliance functionality for applications in smart-grid," *Energies*, vol. 15, no. 1, p. 109, Dec. 2021, doi: 10.3390/en15010109.

[117] P. Zhang, S. Ding, and Q. Zhao, "Exploiting blockchain to make AI trustworthy: A software development lifecycle view," *ACM Comput. Surv.*, vol. 56, pp. 1–32, Jan. 2023.

[118] M. Laroui, B. Nour, H. Moungla, M. A. Cherif, H. Afifi, and M. Guizani, "Edge and fog computing for IoT: A survey on current research activities & future directions," *Comput. Commun.*, vol. 180, pp. 210–231, Dec. 2021, doi: 10.1016/j.comcom.2021.09.003.

[119] A. Hasankhani, S. Mehdi Hakimi, M. Bisheh-Niasar, M. Shafie-khah, and H. Asadolahi, "Blockchain technology in the future smart grids: A comprehensive review and frameworks," *Int. J. Electr. Power Energy Syst.*, vol. 129, Jul. 2021, Art. no. 106811, doi: 10.1016/j.ijepes.2021.106811.

[120] A. Pasdar, Y. C. Lee, and Z. Dong, "Connect API with blockchain: A survey on blockchain Oracle implementation," *ACM Comput. Surv.*, vol. 55, no. 10, pp. 1–39, Oct. 2023, doi: 10.1145/3567582.

[121] H. Al-Breiki, M. H. U. Rehman, K. Salah, and D. Svetinovic, "Trustworthy blockchain oracles: Review, comparison, and open research challenges," *IEEE Access*, vol. 8, pp. 85675–85685, 2020, doi: 10.1109/ACCESS.2020.2992698.

[122] S. K. Ezzat, Y. N. M. Saleh, and A. A. Abdel-Hamid, "Blockchain oracles: State-of-the-art and research directions," *IEEE Access*, vol. 10, pp. 67551–67572, 2022, doi: 10.1109/ACCESS.2022.3184726.

[123] H. Al Breiki, L. Al Qassem, K. Salah, M. H. U. Rehman, and D. Sevtinovic, "Decentralized access control for IoT data using blockchain and trusted oracles," in *Proc. IEEE Int. Conf. Ind. Internet (ICII)*, Nov. 2019, pp. 248–257, doi: 10.1109/ICII.2019.00051.

[124] H. Patil, S. Sharma, and L. Raja, "Study of blockchain based smart grid for energy optimization," *Mater. Today, Proc.*, vol. 44, pp. 4666–4670, Jan. 2020, doi: 10.1016/j.matpr.2020.11.013.

[125] S. Abed, R. Jaffal, B. J. Mohd, and M. Al-Shayeji, "An analysis and evaluation of lightweight hash functions for blockchain-based IoT devices," *Cluster Comput.*, vol. 24, no. 4, pp. 3065–3084, Dec. 2021, doi: 10.1007/s10586-021-03324-1.

[126] S. Fu, L. Zhao, X. Ling, and H. Zhang, "Maximizing the system energy efficiency in the blockchain based Internet of Things," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6, doi: 10.1109/ICC.2019.8761539.

[127] O. Debauche, S. Mahmoudi, and Y. Moussaoui, "Internet of Things learning: A practical case for smart building automation," in *Proc. 5th Int. Conf. Cloud Comput. Artif. Intell., Technol. Appl. (CloudTech)*, Nov. 2020, pp. 1–8, doi: 10.1109/CloudTech49835.2020.9365920.

[128] *European Telecommunications Standards Institute*. Accessed: Nov. 14, 2023. [Online]. Available: https://www.etsi.org/standards#Predefined%20Collections

[129] *Institute of Electrical and Electronics Engineers (IEEE)*. Accessed: Nov. 14, 2023. [Online]. Available: https://www.ieee.org/standards/

[130] *Industrial Internet Consortium (IIC)*. Accessed: Nov. 14, 2023. [Online]. Available: https://www.iiconsortium.org/

[131] *OpenFog Consortium*. Accessed: Nov. 14, 2023. [Online]. Available: https://opcfoundation.org/

[132] *Standardization and Commoditization Drive the Evolution of Edge Computing*. Accessed: Nov. 14, 2023. [Online]. Available: https://www.linkedin.com/pulse/standardization-commoditization-drive-evolution-edge-computing/

[133] M. S. Bali, K. Gupta, D. Gupta, G. Srivastava, S. Juneja, and A. Nauman, "An effective technique to schedule priority aware tasks to offload data on edge and cloud servers," *Meas., Sensors*, vol. 26, Apr. 2023, Art. no. 100670, doi: 10.1016/j.measen.2023.100670.

[134] T. Dreibholz and S. Mazumdar, "Towards a lightweight task scheduling framework for cloud and edge platform," *Internet Things*, vol. 21, Apr. 2023, Art. no. 100651, doi: 10.1016/j.iot.2022.100651.

[135] J. Dogani, R. Namvar, and F. Khunjush, "Auto-scaling techniques in container-based cloud and edge/fog computing: Taxonomy and survey," *Comput. Commun.*, vol. 209, pp. 120–150, Sep. 2023, doi: 10.1016/j.comcom.2023.06.010.

[136] X. Li, W. Li, and X. Zhang, "Multi-resource fair allocation with bandwidth requirement compression in the cloud–edge system," *Comput. Electr. Eng.*, vol. 105, Jan. 2023, Art. no. 108510, doi: 10.1016/j.compeleceng.2022.108510.

[137] Y. Jararweh, "Enabling efficient and secure energy cloud using edge computing and 5G," *J. Parallel Distrib. Comput.*, vol. 145, pp. 42–49, Nov. 2020, doi: 10.1016/j.jpdc.2020.06.014.

[138] B. Alotaibi, "Utilizing blockchain to overcome cyber security concerns in the Internet of Things: A review," *IEEE Sensors J.*, vol. 19, no. 23, pp. 10953–10971, Dec. 2019, doi: 10.1109/JSEN.2019.2935035.

[139] A. Gerodimos, L. Maglaras, M. A. Ferrag, N. Ayres, and I. Kantzavelou, "IoT: Communication protocols and security threats," *Internet Things Cyber-Physical Syst.*, vol. 3, pp. 1–13, Jan. 2023, doi: 10.1016/j.iotcps.2022.12.003.

[140] K. Zheng, C. Ding, and J. Wang, "A secure data-sharing scheme for privacy-preserving supporting node–edge–cloud collaborative computation," *Electronics*, vol. 12, no. 12, p. 2737, Jun. 2023, doi: 10.3390/electronics12122737.

[141] L. Ricciardi Celsi and A. Valli, "Applied control and artificial intelligence for energy management: An overview of trends in EV charging, cyber-physical security and predictive maintenance," *Energies*, vol. 16, no. 12, p. 4678, Jun. 2023, doi: 10.3390/en16124678.

[142] *GDPR*. Accessed: Nov. 17, 2023. [Online]. Available: https://gdpr.eu/article-83-conditions-for-imposing-administrative-fines

[143] M. F. Zia, M. Benbouzid, E. Elbouchikhi, S. M. Muyeen, K. Techato, and J. M. Guerrero, "Microgrid transactive energy: Review, architectures, distributed ledger technologies, and market analysis," *IEEE Access*, vol. 8, pp. 19410–19432, 2020, doi: 10.1109/ACCESS.2020.2968402.

[144] J. M. G. Sánchez et al., "Edge computing for cyber-physical systems: A systematic mapping study emphasizing trustworthiness," *ACM Trans. Cyber-Phys. Syst.*, vol. 6, no. 3, pp. 1–28, Jul. 2022, doi: 10.1145/3539662.

[145] N. Schizas, A. Karras, C. Karras, and S. Sioutas, "TinyML for ultra-low power AI and large scale IoT deployments: A systematic review," *Future Internet*, vol. 14, no. 12, p. 363, Dec. 2022, doi: 10.3390/fi14120363.

[146] A. Bahrini et al., "ChatGPT: Applications, opportunities, and threats," in *Proc. Syst. Inf. Eng. Design Symp. (SIEDS)*, Apr. 2023, pp. 274–279, doi: 10.1109/sieds58326.2023.10137850.

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

Olha Boiko received the Ph.D. degree in computer science from Sumy State University, Sumy, Ukraine, in 2017.

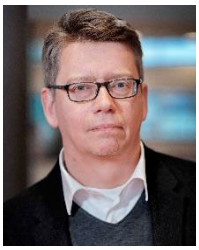
She is a Senior Lecturer with the Department of Information Technologies, Sumy State University, and a Researcher with the Department of Computer Science and Media Technology, Malmö University, Malmö, Sweden. She is with the Internet of Things and People Research Centre, Malmö University. Her research interests include, but are not limited to, artificial intelligence, systems integration, and decision support within the realm of hybrid energy grids.

Anton Komin received the B.S. and M.S. degrees in computer science from Sumy State University, Sumy, Ukraine, in 2010 and 2011, respectively, where he is currently pursuing the Ph.D. degree in computer science.

Since 2010, he is a software engineer of accountant, Customer Relationship Management (CRM), and Enterprise Resource Planning (ERP) systems based on the 1C:Enterprise platform. For the past seven years, his focus has been on the development of iOS mobile applications. His research interests are technical leadership and solution architecture for projects that include mobile applications.

Reza Malekian (Senior Member, IEEE) is a Professor of Computer Science with Malmö University, Malmö, Sweden, and an Extraordinary Professor with the Department of Electrical, Electronic, and Computer Engineering, University of Pretoria, Pretoria, South Africa, where he led the Advanced Sensor Networks Research Group. His research interests include the Internet of Things and cyber–physical systems.

Dr. Malekian received the Vice-Chancellor and Principal's Exceptional Young Researcher's Award. He is also included on the top 100 list of the Royal Swedish Academy of Engineering Sciences (Kungl. Ingenjörsvetenskapsakademien). He is a Chartered Engineer with the Engineering Council of the United Kingdom. He is an Associate Editor of the IEEE SENSORS JOURNAL and an Area Editor of the IEEE INTERNET OF THINGS JOURNAL.



Paul Davidsson received the Ph.D. degree in computer science from Lund University, Lund, Sweden, in 1996.

He is a Professor of Computer Science with Malmö University, Malmö, Sweden, where he is the Director of the Internet of Things and People Research Centre. He is a Co-Founder of the company Noda Intelligent Systems, Karlshamn, Sweden, which focuses on intelligent energy systems, and he is a co-inventor of patents in this area. The results of his work have been reported in more than 250 peer-reviewed scientific articles published in international journals, conference proceedings, and books. His research interests include artificial intelligence, agent-based simulation, system architecture, and data science. His current application areas include smart buildings, transport, and energy systems.

Authorized licensed use limited to: Universidad Nacional de Ingeniería. Downloaded on April 23,2025 at 00:51:18 UTC from IEEE Xplore. Restrictions apply.

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 192 (2021) 3423–3431

Procedia Computer Science

www.elsevier.com/locate/procedia

25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Students' emotion extraction and visualization for engagement detection in online learning

Mohammad Nehal Hasnine[a*], Huyen T. T. Bui[a], Thuy Thi Thu Tran[a], Ho Tran Nguyen[a], Gökhan Akçapınar[b], Hiroshi Ueda[a]

[a]*Reseach Center for Computing and Multimedia Studies, Hosei University, 3-7-2 Kajinocho, Koganei city, Tokyo 184-8584, Japan*
[b]*Department of Computer Education and Instructional Technology, Hacettepe University, 06800, Çankaya Ankara, Turkey*

Abstract

Online learning is growing in various forms, including full-online, hybrid, hy-flex, blended, synchronous, and asynchronous. Assessing students' engagement without having real contact between teachers and students is becoming a challenge for the teachers. Therefore, this paper focuses on analyzing online lecture videos to detect students' engagement without relying on learning management systems produced data. In this regard, an intelligent application for teachers is developed to understand students' emotions and detect students' engagement levels while a lecture is in progress. Real-time and offline lecture videos are analyzed using computer vision-based methods to extract students' emotions, as emotions play essential roles in the learning process. Six types of basic emotions, namely 'angry', 'disgust', 'fear', 'happy', 'sad', 'surprise', 'neutral' are extracted using a pre-trained Convolutional Neural Network (CNN), and those are used to detect 'Highly-engaged', 'Engaged', and 'Disengaged' students in a virtual classroom. This educational application is tested on a 28-second-long lecture video taken from YouTube consisting of 11 students. The results are visualized for engagement detection using visualization methods. Furthermore, this intelligent application, in real-time, is capable of recognizing multiple faces when multiple students share a single camera. Nonetheless, this educational application could be used for supporting collaborative learning, problem-based learning, and emotion-based grouping.

© 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of KES International.

Keywords: emotion-aware learning analytics; engagement; intelligent learning system; lecture video analysis; multimodal learning analytics

* Mohammad Nehal Hasnine (Corresponding author.) Tel.: +81-(0)42-387-6070; fax: +81-(0)42-387-6085.
E-mail address: nehal.hasnine.79@hosei.ac.jp

1877-0509 © 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of KES International.
10.1016/j.procs.2021.09.115

1. Introduction

1.1. Link between student emotion and engagement in higher educational contexts

Student engagement can be defined [1] as: "the interaction between the time, effort and other relevant resources invested by both students and their institutions intended to optimize the student experience and enhance the learning outcomes and development of students and the performance, and reputation of the institution". According to Kuh [2], "engagement premise is straightforward and easily understood: the more students study a subject, the more they know about it, and the more students practice and get feedback from faculty and staff members on their writing and collaborative problem solving, the deeper they come to understand what they are learning".

In higher education, engagement does not represent one variable. Instead, it could come across in many distinct ways. According to Jennifer Fredricks, students' engagement happens in three ways: emotionally, behaviorally, and cognitively. She also stated that emotional engagement is a distinction between boredom and interest. In contrast, behavioral engagement and cognitive engagement are the distinctions between positive behaviors and those we might associate with acting out, and distinction between rote learning and what we might associate with more profound learning, respectively. Besides behavioral, emotional, and cognitive engagements, agentic engagement is vital in learning. Agentic engagement turns students' proactiveness and constructive contributions into learning activities such as making suggestions and offering input. As learning analytics matures, the community continuously monitors students' engagement from their digital learning trails as engagement plays an essential role in the learning process. In the learning process, engagement is known to be one of the qualitative indicators [3]. It is said that students with higher engagement often learn more than those whose engagement level is low.

Emotions are also important factors that influence our learning process. Emotion also has a substantial influence on our memory, reasoning capability, perception, and logical thinking. In education, emotion is essential as it is highly associated with students' attention. Scholars are continuously researching different types of emotions. These include basic emotions such as anger, disgust, fear, happiness, sadness, and surprise. [4]; and academic emotions such as happy, bore, enjoy, lazy, sad, excite, proud, tire, and angry. [5]. Emotions may influence students' engagements. Therefore, emotion-based models aim to establish a link between academic emotions and engagement levels to identify the level of student engagement depending on their emotions [6]. Therefore, it is necessary to understand students' emotions and associate them with various engagement levels for better educational outcomes.

1.2. Research objectives and contributions

This research aims to understand students' emotions and engagement while attending classes online without having real contact with the teachers. An educational application is aimed to develop for the teachers who are conducting lectures using online tools. Therefore, the key contributions of this paper include:

- The use of emotion recognition in the e-learning platform.
- Design the architecture of a model that could understand students' emotions and detect various engagements from those emotions.
- Develop an intelligent educational application based on the model.
- Visualize results to the teachers as the intervention.

2. Literature review

In the multimodal learning analytics literature, there are several applications of automatic emotion extraction and analysis. As education transitions online, emotions are much discussed in the research community for improving students' learning. Many scholars have combined multimodal data, including emotion, pose, and gestures, generated data to predict students' engagements with learning management systems.

To date, several methods have been employed to detect student engagements including engagement tracing, log file analysis, sensor data analysis, and computer vision-based methods. Computer vision-based methods are based on the modalities of three sub-categories, namely facial expression analysis, gesture-and-posture detection, and eye

movement detection [7]. As teaching over web conferencing technologies is becoming popular, computer vision-based methods could be applied to analyze lecture videos.

A recent study analyzed lecture videos to identify 'happy human', 'happy agent', 'bored human', and 'bored agent' [8]. The findings of this study suggest that students are happier and found to be more motivated when they learn from happy instructors than from bored instructors. A lecture video analysis tool [9] is developed that is cable to analyze and digitalize students' facial expressions. This tool can understand disgust, sadness, happiness, fear, contempt, anger, and surprise students from the faces. Also, this tool provides instant feedback to the lecturer by monitoring the students' emotions. Another study [10] used eyes and head movement data and combined it with facial emotions to generating a concentration index. Later on, this concentration index is used to classify three types of engagements, namely 'very engaged', 'nominally engaged', and 'not engaged at all'. Another example is, the Wits Intelligent Teaching System [11], for analyzing online video lectures to understanding students' engagements. In this system, students sitting in large lecture venues are considered where the authors have used behavior and posture detection using AlexNet's convolutional neural network.

3. The model for emotion extraction and engagement detection

The stakeholder of this educational application is the teachers who are teaching using web conferencing tools such as Zoom or WebEx. While teaching over web conferencing tools, it is difficult for the instructors to understand their students' attention and engagement with the content. Therefore, this system is for assisting the instructors in having a better understanding of the students. Figure 1 shows the overall architecture of the model.

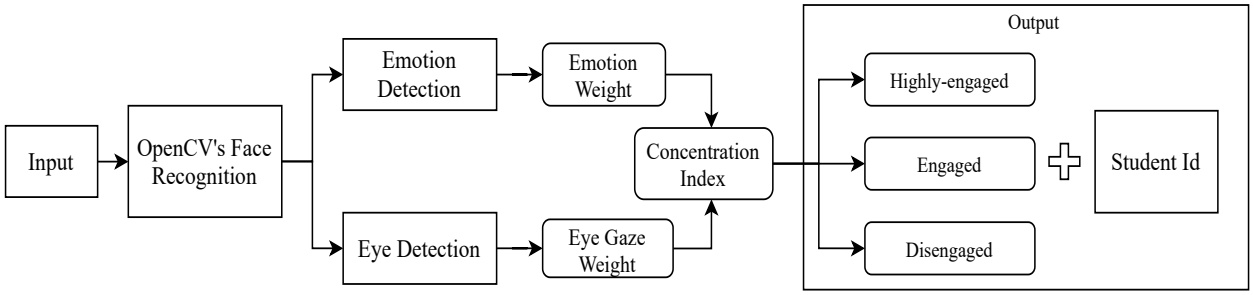


Fig. 1. The overall architecture of the model

3.1. Implementation of the OpenCV Face Recognition

In computer vision and image processing, face recognition is a method that is used to identify and verify people through their facial features. In general, a face recognition algorithm captures, analyzes, and compares patterns depending on the person's facial properties such as color, shape, and size. In general, the face detection process is for detecting and locating human faces in images and videos. Face capture and face match are the processes for transforming analog information (a face) into a set of digital information (data) based on the person's facial features and verifying if two faces belong to the same person [12].

In Artificial Intelligence, several libraries could be used to recognize faces, including Python, YoLo, and OpenCV. For this study, OpenCV is used for face recognition as it is an open-source library and is free of cost. OpenCV also provides ready-to-use methods with advanced capabilities such as facial detection, face tracking, and facial recognition [13]. Figure 2 is the design of the face recognition system.

The face recognition system is comprised of four steps. They are initial dataset construction, facial features extraction, face embedding, and face matching.

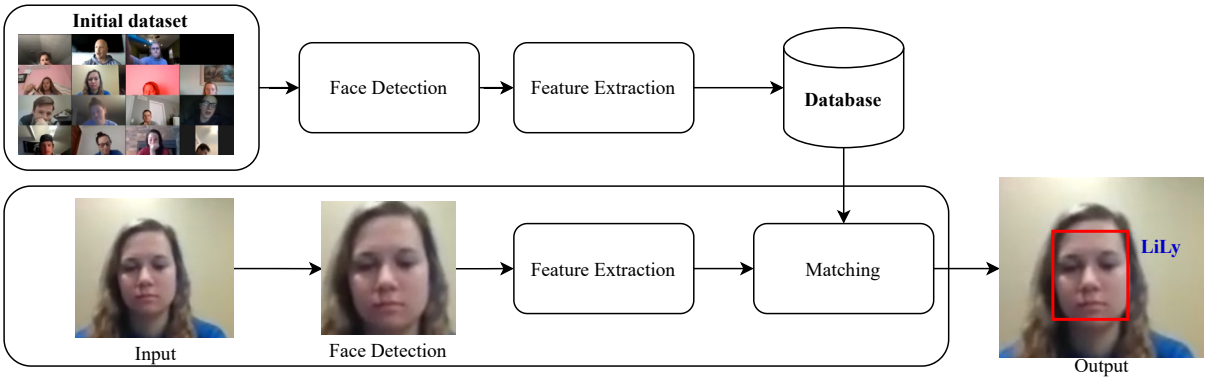


Fig. 2. Face recognition system

- **Initial dataset construction:** The first step is to construct a database with the known students. This can be done by inputting a student name, id, and photo in the database. The initial dataset is necessary to know about the number of students in the class, the facial description (that is, photo) of the students, and identify each student in the database.

```
|--Dataset
    |----- student_name.jpg
    |--------huyen.jpg
    |--------thuy.jpg
    |--------ho.jpg
```

- **Facial feature extraction and face embedding:** The next step is face embedding. This part of the model can detect and locate the face and embed the located face. This task is done with deep metric learning by dlib, a toolkit for face embedding. The dlib libraries provide the pre-trained model used to generate 128 measurements for each input face. The network generates nearly the same numbers when looking at two different pictures of the same person else, it generates different numbers when there are two different pictures of a different person.

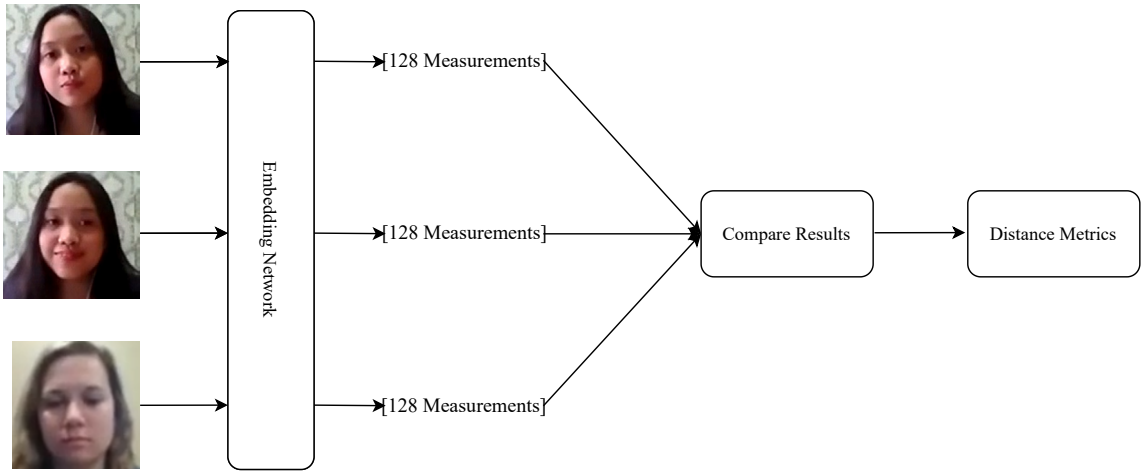


Fig. 3. Face embedding network for face recognition system

- **Face matching:** In testing new student's information, the matching step checks each face in the input images whether the face matches with the face in the database or not. The output of the matching steps is:

if (input image is matched in database) **then**
    return student's name
else if (input image partial match) **then**
    return the student's name whose partial match is found
else (input image does not match in database)
    return unknown student

3.2. Emotion detection

At present, our emotion detection model is based on a research paper by Arriaga O et al. in 2017 [14]. This model can classify the human face within the following classes: angry, disgust, fear, happy, sad, surprise, neutral. For this classification, a 21 layered architecture was trained on the FER 2013 dataset that obtained an accuracy of 68%.

3.3. Eye detection

Detection of eye positions and their movements are analyzed. For this step, we used the guidelines of eye detection provided by Sergio Canu [15]. As guide-lined by Sergio Canu, we detected two eyes separately using the following measurement.

Left eye points: (36, 37, 38, 39, 40, 41)
Right eye points: (42, 43, 44, 45, 46, 47)

After that, using the face landmarks detection approach [16], our model detected 68 specific face landmarks. We assigned a specific index to each of the points. The next step was to create two lines: crossing the eye horizontally and crossing the eye vertically. Finally, for detecting the eyes are open or not, we calculate the left eye ratio, right eye ratio, and blinking ratio based on the horizontal and vertical line. And, to determine the eye gaze, we calculate left_gaze_ratio_lr, left_gaze_ratio_up, right_gaze_ratio_lr, right_gaze_ratio_ud.
Table 1 shows the eye size scores.

Table 1. Eye ratio score

	Open	Semi-open	Close
Eye Size	>0.3	0.2 - 0.3	<0.2

3.4. Engagement detection

In this paper, engagement detection is calculated based on the concentration index calculation method proposed by Sharma P et al. in 2019 [10]. For this study, the concentration index is based on eye gaze and emotion weights. Emotion weight describes a specific emotional state at the time of a person. The value varies from 0 to 1. Table 2 shows the emotion weight to present each of the six emotions.

Table 2. Emotion weight [10]

Emotion	Neutral	Happy	Surprise	Sad	Angry	Fearful
Weight	0.9	0.6	0.6	0.3	0.25	0.3

Eye gaze weight is calculated based on the weights presented in the Table 3.

Table 3. Eye gaze weight [17]

Eye gaze weight	Open	Semi-open	Close
(Look at) Center	5	1.5	0
(Look) Up right	2	1.5	0
(Look) Up left	2	1.5	0
(Look) Right	2	1.5	0
(Look) Left	2	1.5	0
(Look) Down right	2	1.5	0
(Look) Down left	2	1.5	0

Concentration index (CI) is calculated using the following equation [10]

$$CI = (\text{Emotion Weight} \times \text{Gaze Weight}) / 4.5 \quad (1)$$

A student's engagement is calculated based on the CI. The following table is used to calculate engagement from the CI.

Table 4. Engagement detection from CI [10]

Engagement type	CI
Highly-engaged	>65%
Engaged	25-65%
Disengaged	<25%

4. Result visualization

Figure 4 displays the way our application detects students, understands their emotion and detects their engagements.

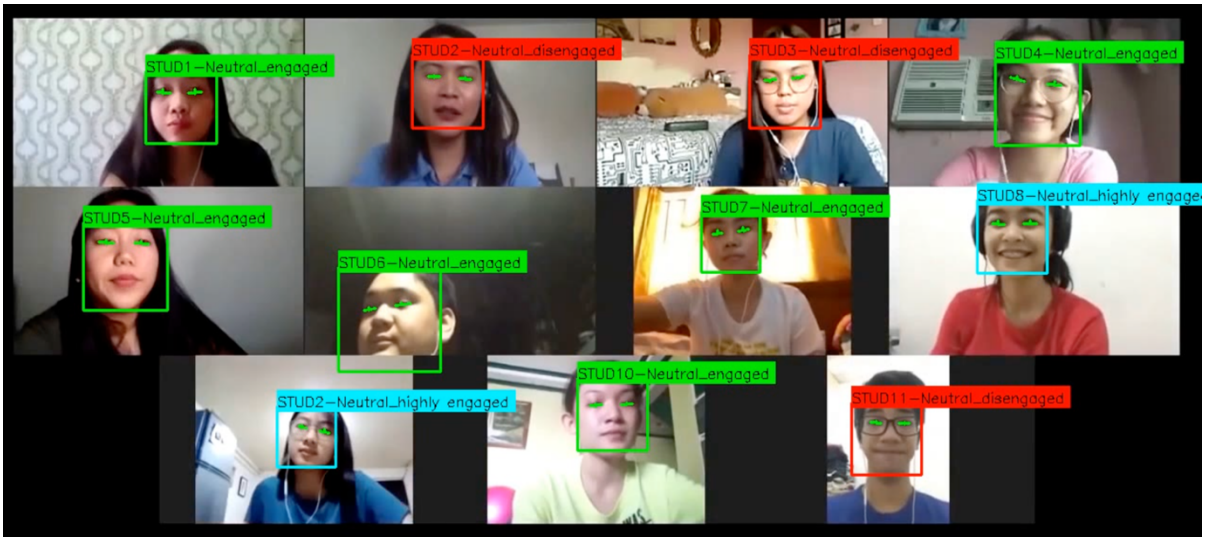
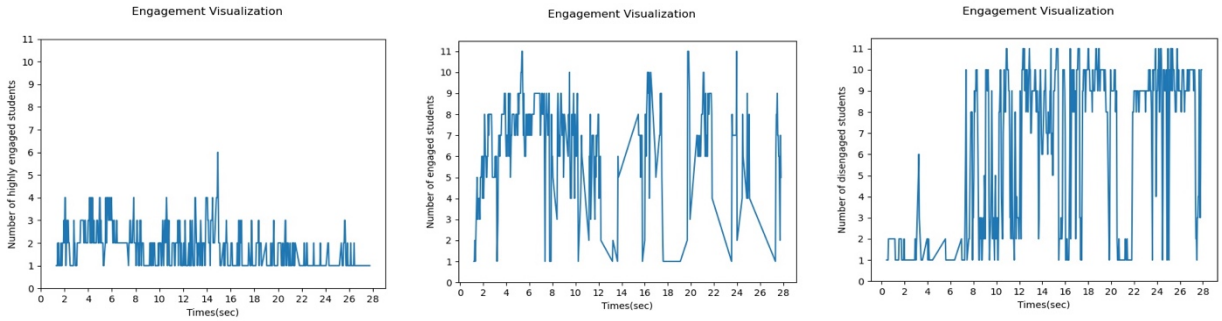


Fig. 4. Lecture video analysis

We used a YouTube video consisting of 11 students engaged in an interactive lecture to visualize the result. The video is 28-second-long. We input the video file to our application to understand those students' behaviors while attending an online lecture. To test the result, our application processed 25 frames per second.

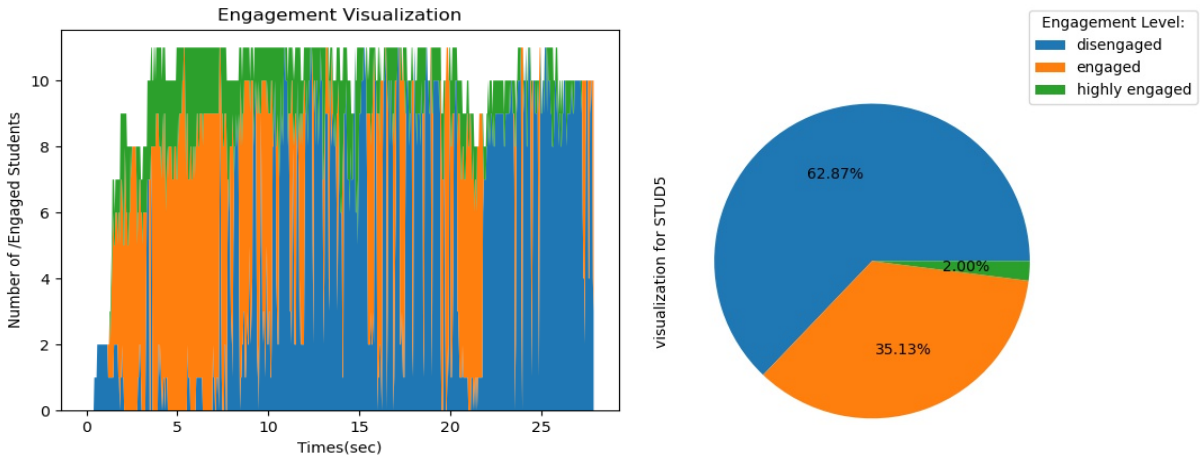
Once the system finishes processing the video, the application generates three types of engagement visualizations to support a teacher.

- First type of visualization is the time-series of engagement detection (refer to Fig. 5).
- Second type of visualization is about the summarization of engagement level (refer to Fig.6(a)).
- The third type of visualization is each student's engagement in the class (refer to Fig.6(b)).



(a) (b) (c)

Fig. 5. Time series of engagement detection for (a) Highly-engaged; (b) Engaged; (c) Disengaged students.



(a) (b)

Fig. 6. (a) Summarization of engagement level; (b) Visualization for each student's engagement in the class.

5. Discussion

Engagement plays an essential role in students' learning. Since it is dynamic and continually changes, detecting multiple students' engagement during the lecture is not an easy task for instructors, especially during the online learning scenarios. Thanks to the advancement in artificial intelligence and related fields (e.g., deep learning), there

are more possibilities in automatically detecting engagement. There are two main approaches in the literature for automatic engagement detection, and they are as follows sensor data analysis-based approach and computer vision-based approach [7]. This study focused especially on emotional engagement and followed a computer vision-based approach. Specifically, students' facial expressions captured from a web camera were used as a proxy to measure their engagement continually across the lecture. As learning analytics matures, it has shown significant advances in several areas, from pedagogy to artificial intelligence in education. Without relying on learning management systems produced data, it also explores sophisticated technologies, including eye-tracking, sentiment analysis, emotion analysis, and video analysis. As virtual classrooms are becoming popular, lecture video analysis is getting attention in the community as lecture videos could reveal learners' emotions, sentiments, and engagement. Visual cues such as eye movement, facial expression, and gestures, and posture can be used to measure one's engagement. However, facial expression is the most potent and effective way to assess emotional engagement [18].

As student engagement and emotions play essential roles in the learning process, lecture videos could be analyzed using computer vision-based methods. This paper proposed an educational application for understanding students' emotions and engagement by analyzing lecture videos. In order to create easy to interpret metric for course instructors, we related students' emotions with their level of engagement. In order to link students' emotions to their engagement, we used the results of the previous study conducted by [10]. Teachers could use this application to understand students' behavior while teaching classes using online tools such as Zoom or WebEx. The contributions of this work are designing a model that understands students' emotions and detects various engagements, develops an educational application based on the model, and visualizes results for better educational insights. The developed model successfully identified multiple students' emotions from the lecture recording. However, to use the developed model in real-life situations, external validation is required by using the sensor data (e.g., EEG, eye, and tracker) or ground truth data labeled by an expert.

This study has some limitations, criticisms, and several challenges that would be addressed in future work. For example, the challenges that we faced include engagement detection from the profile pictures; challenges when objects cover the face; engagement detection under bad light; when speaking or doing off-task behavior, there will be various kinds of emotion showing up on their face; the orientation of the camera may affect the performance of analyzing emotion; and the orientation of the face. The proposed model is not validated with actual data, and therefore, we aim to collect students' data and validate it. Moreover, the result presented in this paper is based on a publicly available lecture video taken from YouTube. Hence, at this point, we could not ensure that the three types of engagement detected by the application match the three types of engagement determined by the person. To solve these limitations, we aim to conduct a user study in the future.

References

[1] V. Trowler, "Student engagement literature review," *The higher education academy*, vol. 11, no. 1, pp. 1–15, 2010.

[2] G. D. Kuh, "The national survey of student engagement: Conceptual and empirical foundations," *New directions for institutional research*, vol. 2009, no. 141, pp. 5–20, 2009.

[3] F. D'Errico, M. Paciello, and L. Cerniglia, "When emotions enhance students' engagement in e-learning processes," *Journal of e-Learning and Knowledge Society*, vol. 12, no. 4, 2016.

[4] P. E. Griffiths, "Basic emotions, complex emotions, Machiavellian emotions," 2002.

[5] A. B. Bernardo, J. A. Ouano, and M. G. C. Salanga, "What is an academic emotion? Insights from Filipino bilingual students' emotion words associated with learning," *Psychological Studies*, vol. 54, no. 1, pp. 28–37, 2009.

[6] K. Altuwairqi, S. K. Jarraya, A. Allinjawi, and M. Hammami, "A new emotion–based affective model to detect student's engagement," *Journal of King Saud University-Computer and Information Sciences*, 2018.

[7] M. A. A. Dewan, M. Murshed, and F. Lin, "Engagement detection in online learning: a review," *Smart Learning Environments*, vol. 6, no. 1, pp. 1–20, 2019.

[8] T. Horovitz and R. E. Mayer, "Learning with human and virtual instructors who display happy or bored emotions in video lectures," *Computers in Human Behavior*, vol. 119, p. 106724, Jun. 2021, doi: 10.1016/j.chb.2021.106724.

[9] G. Tonguç and B. Ozaydın Ozkara, "Automatic recognition of student emotions from facial expressions during a lecture," *Computers & Education*, vol. 148, p. 103797, Apr. 2020, doi: 10.1016/j.compedu.2019.103797.

[10] P. Sharma, S. Joshi, S. Gautam, S. Maharjan, V. Filipe, and M. J. C. S. Reis, "Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning," *arXiv:1909.12913 [cs]*, Dec. 2020, Accessed: Apr. 15, 2021. [Online]. Available: http://arxiv.org/abs/1909.12913

[11] R. Klein and T. Celik, "The Wits Intelligent Teaching System: Detecting student engagement during lectures using convolutional neural

networks," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sep. 2017, pp. 2856–2860. doi: 10.1109/ICIP.2017.8296804.

[12] "Facial recognition: top 7 trends (tech, vendors, markets, use cases & latest news)," *Thales Group*. http://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/facial-recognition (accessed Apr. 22, 2021).

[13] M. Khan, S. Chakraborty, R. Astya, and S. Khepra, "Face Detection and Recognition Using OpenCV," in *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, Oct. 2019, pp. 116–119. doi: 10.1109/ICCCIS48478.2019.8974493.

[14] O. Arriaga, M. Valdenegro-Toro, and P. Plöger, "Real-time Convolutional Neural Networks for Emotion and Gender Classification," *arXiv:1710.07557 [cs]*, Oct. 2017, Accessed: Apr. 22, 2021. [Online]. Available: http://arxiv.org/abs/1710.07557

[15] S. Canu, "Eye detection - Gaze controlled keyboard with Python and Opencv p.1," *Pysource*, Jan. 07, 2019. https://pysource.com/2019/01/07/eye-detection-gaze-controlled-keyboard-with-python-and-opencv-p-1/ (accessed Apr. 22, 2021).

[16] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, doi: 10.1109/CVPR.2014.241.

[17] "GitHub - CaedenZ/distractionModel." https://github.com/CaedenZ/distractionModel (accessed Apr. 30, 2021).

[18] M. Mukhopadhyay, S. Pal, A. Nayyar, P. K. D. Pramanik, N. Dasgupta, and P. Choudhury, "Facial Emotion Detection to Assess Learner's State of Mind in an Online Learning System," in *Proceedings of the 2020 5th International Conference on Intelligent Information Technology*, 2020, pp. 107–115.