# UNIVERSIDAD DE CÓRDOBA

## Programa de doctorado:
## Computación avanzada, energía y plasmas

**TÍTULO:**
**PREDICCIÓN DE RENDIMIENTO ACADÉMICO EN APRENDIZAJE COMBINADO MEDIANTE DATOS MULTIMODALES Y TECNICAS DE FUSIÓN Y MINERÍA DE DATOS**

Tesis presentada por:
**Wilson Gustavo Chango Sailema**

Directores:
**Dr. D. Cristóbal Romero Morales**
**Dra. Dña. Rebeca Cerezo Menéndez**

Córdoba Abril, 2022

TITULO: *PREDICCIÓN DE RENDIMIENTO ACADÉMICO EN APRENDIZAJE COMBINADO MEDIANTE DATOS MULTIMODALES Y TÉCNICAS DE FUSIÓN Y MINERÍA DE DATOS*

AUTOR: *Wilson Gustavo Chango Sailema*

© Edita: UCOPress. 2022
Campus de Rabanales
Ctra. Nacional IV, Km. 396 A
14071 Córdoba

https://www.uco.es/ucopress/index.php/es/
ucopress@uco.es

# UNIVERSITY OF CÓRDOBA

## Doctoral Programme:
## Advanced computing, energy and plasmas

**TITLE:**
**PREDICTING ACADEMIC ACHIEVEMENT IN BLENDED LEARNING USING MULTIMODAL DATA AND FUSION AND DATA MINING TECHNIQUES**

A Thesis presented by:
**Wilson Gustavo Chango Sailema**

Advisors:
**Dr. D. Cristóbal Romero Morales**
**Dra. Dña. Rebeca Cerezo Menéndez**

Córdoba April, 2022

**TÍTULO DE LA TESIS:** PREDICTING ACADEMIC ACHIEVEMENT IN BLENDED LEARNING USING MULTIMODAL DATA AND FUSION AND DATA MINING TECHNIQUES

**DOCTORANDO:** Wilson Gustavo Chango Sailema

## INFORME RAZONADO DEL LOS DIRECTORES DE LA TESIS

El doctorando (Wilson Gustavo Chango Sailema) ha progresado enormemente como investigador desde que comenzara la tesis doctoral en el año 2019 en la Universidad de Córdoba. Durante estos 3 años el doctorando ha realizado todas las actividades obligatorias y opcionales, trabajado duro seguido siempre las pautas de trabajo que le hemos marcado los directores y el plan de investigación que se estableció. Como principales frutos del trabajo realizado se han derivado los tres artículos siguientes publicados en revistas incluidas en los primeros cuartiles de la relación de revistas del ámbito de la especialidad y referenciadas en la última relación publicada por el Journal Citation Reports (SCI y/o SSCI):

1. Chango, W., Cerezo, R., & Romero, C. (2021). Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses. *Computers and Electrical Engineering*, *89,* Article 106908.https://doi.org/10.1016/j.compeleceng.2020.106908. Impact Factor: 3.818 (Q1).
2. Chango, W., Cerezo, R., Sánchez-Santillán, M., Azevedo, R., & Romero, C. (2021). Improving prediction of students' performance in intelligent tutoring systems using attribute selection and ensembles of different multimodal data sources. *Journal of Computing in Higher Education, 33*(3), 614–634. https://doi.org/10.1007/s12528-021-09298-8. Impact Factor: 2.627 (Q1).
3. Chango, W., Lara, J. A., Cerezo, R., & Romero, C. (2022). A Review on Data Fusion in Multimodal Learning Analytics. *WIREs Data Mining and Knowledge Discovery.* Advance online publication. https://doi.org/10.1002/widm.1458. Impact Factor:7.250 (Q1).

Córdoba, abril de 2022

Firma del/de los director/es

ROMERO MORALES CRISTOBAL - 30800430A

Firmado digitalmente por ROMERO MORALES CRISTOBAL - 30800430A
Fecha: 2022.04.08 21:04:13 +02'00'

Fdo.: Dr. Cristóbal Romero Morales

CEREZO MENENDEZ REBECA - 71644426R

Firmado digitalmente por CEREZO MENENDEZ REBECA - 71644426R
Fecha: 2022.04.18 17:14:23 +02'00'

Fdo.: Dra. Rebeca Cerezo Menéndez

La tesis titulada "PREDICCIÓN DE RENDIMIENTO ACADÉMICO EN APRENDIZAJE COMBINADO MEDIANTE DATOS MULTIMODALES Y TECNICAS DE FUSIÓN Y MINERÍA DE DATOS", que presenta Wilson Gustavo Chango Sailema para optar al grado de doctor, ha sido realizada dentro del programa de doctorado Computación Avanzada, Energía y Plasmas, en la línea de investigación Aprendizaje Automático, Modelado de Sistemas y Minería de Datos, del Departamento de Informática y Análisis Numérico de la Universidad de Córdoba, bajo la dirección de los doctores Cristóbal Romero Morales y Rebeca Cerezo cumpliendo, en su opinión, los requisitos exigidos a este tipo de trabajos.

El Doctorando:

Firmado electrónicamente por:
WILSON GUSTAVO
CHANGO SAILEMA

Fdo: Wilson Gustavo Chango Sailema

El Director

ROMERO MORALES
CRISTOBAL -
30800430A

Firmado digitalmente por
ROMERO MORALES
CRISTOBAL - 30800430A
Fecha: 2022.04.08 21:05:33
+02'00'

Fdo: Dr. Cristóbal Romero Morales

El Director

CEREZO
MENENDEZ
REBECA -
71644426R

Firmado digitalmente
por CEREZO MENENDEZ
REBECA - 71644426R
Fecha: 2022.04.18
17:14:44 +02'00'

Fdo: Dra. Rebeca Cerezo Menéndez

**FUENTES DE FINANCIACIÓN**

Esta tesis ha sido parcialmente subvencionada con los proyectos del Ministerio de Ciencia e Innovación: INTENSE: Mejorando la experiencia del usuario de ciencia de datos con inteligencia computacional (PID2020-115832GB-I00) y por Detección Temprana e Intervención en Dificultades del Aprendizaje Específicas (DAE) desde el Modelo RtI (PID2019-107201GB-I00).

# AGRADECIMIENTOS

El aporte investigativo no hubiese sido posible sin la participación de personas e instituciones que han facilitado las cosas para que este trabajo llegue a culminarse de manera satisfactoria, es para mí un verdadero placer utilizar este espacio para ser justo y consecuente con ellas, expresándoles mis agradecimientos, a la Pontificia Universidad Católica Del Ecuador Sede Esmeraldas, por haber financiado gran parte de mi estudio doctoral otorgándome una beca a partir de la convocatoria del año 2019, de la misma manera a la UNIVERSIDAD DE CÓRDOBA (España), UNIVERSIDAD DE OVIEDO (España) , por facilitar toda la información para las publicaciones de los artículos.

Deseo expresar mi agradecimiento a los directores de esta tesis doctoral, Dr. Cristóbal Romero Morales (UCO) y Dr. Rebeca Cereza Menéndez (Universidad de Oviedo), por la dedicación y apoyo que han brindado a este trabajo y por la dirección y el rigor que ha facilitado a las mismas.

A mis padres Luis y María, mis hermanos Freddy, Milton y Edison, a mi Tío Luis que está en el cielo, pero, sobre todo, gracias a mi esposa Cristina y a mis hijos Ariel y Said, por su inmensa paciencia, comprensión y solidaridad con este proyecto, por el tiempo que me han concedido, un tiempo robado a la historia familiar, sin su apoyo este trabajo nunca se habría escrito y, por eso, este trabajo es también de ustedes.

**Wilson Gustavo Chango S.**

*“Sin importar lo mala que parezca ser la vida, siempre hay algo que puedes hacer, y tener éxito. Mientras que haya vida, hay esperanza”*

**Stephen Hawking**

# TABLA DE CONTENIDOS

# RESUMEN

El objetivo general de este trabajo persigue mejorar la predicción del rendimiento académico a través de la minería de datos y técnicas de fusión, provenientes de diferentes modalidades, bien sea por el contexto de aprendizaje o por la metodología de evaluación utilizada. El fin último, pretende posibilitar la prevención e intervención de ese proceso de aprendizaje para paliar las posibles dificultades que puedan acontecer.

El objetivo general se concreta en dos objetivos más concretos. En primer lugar, realizar una revisión teórica de la literatura existente sobre fusión de datos multimodales, y, en segundo lugar, estudiar que aproximación de fusión de datos ofrece mejores resultados para la predicción del rendimiento académico.

Para acometer el primer objetivo específico se realizó una revisión teórica sistemática y para el segundo, se llevaron a cabo dos estudios empíricos. Los estudios empíricos analizaron datos provenientes de dos muestras en las cuales se recogieron datos con diferentes metodologías. Por un lado, un grupo de 57 estudiantes de ingeniería de la Universidad de Córdoba, cuyas sesiones de aprendizaje se desarrollaron en un entorno *blended learning* (clases presenciales y campus virtual), por otro lado, un grupo de 40 estudiantes de diferentes titulaciones de la Universidad de Oviedo, que desarrollaron sesiones de aprendizaje en MetaTutorES, un entorno hipermedia diseñado para evaluar el proceso de aprendizaje a través de metodología multimodal.

Fruto de la revisión, observamos cómo se han utilizado técnicas de fusión de datos en diferentes escenarios de aprendizaje, pero también en base al tipo de datos fusionados y a los enfoques de fusión de datos utilizados. Asimismo, se detectaron los principales retos presentes y futuros en el área de estudio, siendo uno de ellos, la experimentación con diferentes aproximaciones a la fusión de datos, para observar cuál de ellas arroja mejores resultados para predecir el rendimiento académico. En este sentido, los resultados de ambos estudios empíricos coincidieron en que el uso de la técnica *ensemble* junto con la selección de atributos, era la aproximación que mejores índices de precisión y ajuste presentaba.

Estos resultados cobran importancia en el contexto educativo, dado que, cuanto antes y mejor puedan predecirse posibles problemas en el rendimiento, antes se podrá implementar medidas de prevención e intervención, los llamados Sistemas de Detección Temprana.

# ABSTRACT

The main goal of this study is to improve academic performance prediction through the mining of data and data fusion from different modalities, either by the use of learning context or, by the evaluation methodology. The ultimate goal is to enable the prevention and intervention to reduce the possible unexpected difficulties in this learning process.

To accomplish the main objective, two specific objectives have been established. The first is to review the available literature about multimodal data fusion. The second is to study which data fusion approach offers better results in predicting academic performance.

To achieve the first specific objective, a systematic theoretical review was carried out. While for the second, two empirical studies were executed. The empirical studies analyzed data from two samples in which data were collected with different methodologies. On the one hand, a group of 57 students of Engineer career from the University of Córdova, whose learning sessions were carried out through a Blended Learning Environment (face-to-face classes and virtual campus). On the other hand, a group of 40 students of different careers from the University of Oviedo, developed learning sessions in MetaTutorES, a hypermedia environment designed to evaluate the learning process through the multimodal methodology.

As a result, the author observed how data fusion techniques have been used in different learning scenarios also, based on the merged data type and the data fusion approaches used. In addition, present and future challenges in the studied field were detected. One of them is the experimentation with different approaches to data fusion to observe which one obtains better results to predict academic performance. In this sense, the results of both empirical studies agreed that the use of *ensemble* technique with the selection of attributes method was the approach that presented the best accuracy and adjustment rates.

These results become relevant in the educational context since the sooner and better the performance problems are predicted the sooner prevention and intervention measures could be implemented the called Early Detection Systems

# Parte I. Tesis Doctoral

# 1

# INTRODUCCIÓN Y ANTECEDENTES

La presente memoria de tesis se estructura de modo que, a continuación, se ofrece una breve introducción sobre el estado actual y la pertinencia del problema de estudio, para, a continuación, formular los objetivos del trabajo. Objetivos que se acometen a través de tres estudios publicados en sendas revistas de impacto, los cuales se incluyen al final de esta memoria. Se termina con un apartado de conclusión y discusión conjunta de los resultados de sendos trabajos.

Uno de los temas que mayor interés suscitan en Minería de Datos Educativos (en adelante EDM), y que aun implica desafíos presentes y futuros, es la predicción del rendimiento académico. Uno de esos retos pasa por predecir los resultados de aprendizaje a través de técnicas de fusión de datos multimodales, también conocidas como Data Fusión y MMLA (Multimodal Learning Analytics) Análisis del Aprendizaje Multimodal.

Sin embargo, debemos remontarnos varias décadas atrás, cuando, ya con el objetivo de mejorar el proceso de aprendizaje y ayudar a resolver problemas educativos, surgieron diferentes aproximaciones automáticas que utilizan el análisis y la explotación de grandes cantidades de datos generadas durante dicho proceso y que son difíciles de analizar de forma manual. De entre las diferentes aproximaciones, han adquirido especial relevancia, de un lado, el EDM (Educational Data Mining) Minería de Datos en Educación, que consiste en la aplicación de técnicas de minería de datos para analizar los datos generados en el ámbito educativo, con sus particularidades y desafíos. Y de otro lado, el LA (Learning Analytics) Analítica del Aprendizaje, que abarca un espectro más amplio de tareas como la recopilación de los datos educativos, el propio análisis de los mismos y las acciones derivadas de los resultados obtenidos tras el citado análisis (Monés et al., 2020; Romero y Ventura, 2020). En su concepción inicial, los enfoques de análisis de datos educativos se basaban en la explotación de una fuente de datos concreta. Sin embargo, ese enfoque tiene la limitación propia de la fuente de datos empleada, que refleja una porción incompleta de la realidad del proceso educativo.

En este sentido, los modelos de educación a distancia están evolucionando cada vez más y la investigación acerca del aprendizaje en *Computer Based*

*Learning Environments* (en adelante CBLEs) Entornos de Aprendizaje por Computadora, es un tópico que cuenta con un importantísimo corpus teórico. Sin embargo, lejos de agotarse, el campo de estudio se amplía dadas las características propias del objeto de investigación; los CBLEs cambian, avanzan día a día, lo cual supone nuevas implicaciones para el proceso de Enseñanza-Aprendizaje (E-A), y nuevos retos para investigadores, alumnos, profesores e instituciones. El *e-learning (Electronic Learning) Enseñanza y Aprendizaje en Línea*, *b-learning (Blended Learning) Aprendizaje Mixto*, los entornos hipermedia, campus virtuales, *Smart Learning Environments* (SLEs) Entornos de Aprendizaje Inteligentes, etc., ya son agentes determinantes del proceso de E-A de la Educación Superior en todo mundo, sobre manera tras la situación desatada por la pandemia.

La pandemia de COVID-19 ha influido en los sistemas educativos de todo el mundo, provocando el cierre temporal de escuelas y universidades. Hasta agosto de 2020, aproximadamente 1600 millones de estudiantes se vieron afectados por el cierre de escuelas y educación superior en respuesta a la pandemia (Naciones Unidas, 2020). Para superar este problema, la UNESCO recomendó el uso de entornos de aprendizaje a distancia (UNESCO, 2020). Por lo tanto, la pandemia de COVID-19, constituyó un gran desafío para los educadores, pero también para muchas áreas de investigación implicadas en el proceso E-A (Salta et al., 2022).

Como consecuencia de este fenómeno, y del avance intrínseco de la tecnología, cada vez son más, y más complejos, los entornos de aprendizaje, dando lugar a entornos presenciales y virtuales enriquecidos capaces de generar una enorme cantidad de datos de diferentes modalidades que, combinadas, pueden ofrecer un mejor conocimiento del proceso educativo (Chen et al., 2021; Tabuenca et al., 2021).

Esta idea de explotación combinada de fuentes de datos ha dado lugar al *Multimodal Learning Analytics* (MMLA), enfoque que se basa en la captura, integración y análisis de diferentes fuentes de datos educativos que, de forma conjunta, aportan una comprensión holística del proceso de aprendizaje (Sharma y Giannakos, 2020). La combinación de las técnicas de tratamiento de datos multimodales en su intersección con las áreas de EDM y LA ha demostrado ser una línea fructífera en los últimos años (Budaher et al., 2020).

Sin embargo, a pesar de las innumerables ventajas que proporciona, el uso combinado de datos no es un aspecto fácil de abordar ya que se encuentra con importantes desafíos, como la diferente granularidad o la necesidad de alineamiento temporal de los datos recogidos en las diferentes fuentes. En este sentido, el uso de técnicas de Data Fusion resulta necesario y prometedor en el campo de la Educación en general (Sultana et al., 2020) y, particularmente en el campo de EDM/LA (Mu et al., 2020), tal como demuestran trabajos recientes en este sentido (Kaur y Kautish, 2019; Lahat et al., 2015; Poria et al., 2017; Wang et al., 2018).

La fusión de datos es un proceso de múltiples niveles que se ocupa de la asociación, correlación, y combinación de datos de múltiples fuentes para realizar estimaciones y evaluaciones mejoradas respecto a otras técnicas (Castanedo, 2013). En resumen, el proceso de combinar eficientemente datos de diferentes fuentes, de forma que la explotación combinada de esos datos permite obtener un conocimiento de más alto nivel que el proporcionado por cada una de las fuentes por separado. En el ámbito de los CBLEs, esta idea se ha utilizado para intentar explotar datos multimodales de forma conjunta y lograr así un mejor conocimiento del proceso educativo.

Según el ámbito de aplicación, las técnicas de fusión de datos se pueden categorizar de diferentes modos. La clasificación más extendida se basa en considerar el periodo o momento en el que se realiza la fusión, dando lugar a los tres tipos de fusión (Ding et al., 2019). *Feature-level* o *early fusión (*Fusión a nivel de características o fusión temprana*)*: enfoque de fusión consistente en concatenar las diferentes *features* obtenidas de los datos de las diferentes fuentes en un único vector de elementos heterogéneos; *Decision-level* or *later fusión(*Fusión a nivel de decisión o fusión tardía*)*: enfoque de fusión que consiste en crear, en primer lugar, un clasificador con cada una de las fuentes de datos por separado para, posteriormente, fusionar la predicción ofrecida por los diferentes clasificadores; *Hybrid fusión* (Fusión Híbrida): enfoque de fusión que emplea los dos enfoques anteriores en un mismo proceso de fusión.

Pero la fusión de datos comporta sus propios desafíos y es ahí donde se hacen necesario plantearse los objetivos de esta tesis doctoral. Durante el desarrollo de la tesis, hemos llevado a cabo, precisamente, varios experimentos para testar éstas y otras aproximaciones, y concluir cuál de ellas ofrece mejores resultados para conocer el proceso E-A en general, y la predicción del rendimiento académico en particular. En estos experimentos se fusionan datos, a priori, tan alejados, como la asistencia la clase, la toma de apuntes, los *logs* de interacción aprendiz-entorno de aprendizaje, las fijaciones de la mirada, las expresiones del rostro, las emociones, etc. A partir de esos estudios, además, conoceremos de cerca MetaTutorES, una metodología de evaluación multimodal del proceso de aprendizaje que proporciona una gran cantidad de datos con un enorme potencial para su posterior análisis, y con el fin de comprender y optimizar el aprendizaje y los entornos en los que se produce.

## Objetivos

El objetivo principal de esta investigación es predecir el rendimiento académico de estudiantes de educación superior utilizando datos multimodales con técnicas de Minería de Datos y técnicas de fusión de datos. Para cumplir con este objetivo principal, se definen 3 subjetivos:

- **O1:** Realizar una revisión teórica de la literatura existente sobre fusión de datos multimodales para detectar los avances actuales y retos futuros del área de estudio.
- **O2.** Evaluar qué enfoque de fusión de datos y algoritmos de clasificación producen los mejores resultados para predecir el rendimiento en diferentes conjuntos de datos.
- **O3**. Contrastar cómo de útiles son los modelos de predicción que producimos para ayudar a los profesores a detectar a los estudiantes que están en riesgo de fracaso académico.

# 2

# DISCUSIÓN Y CONCLUSIONES

El presente trabajo se planteaba 3 objetivos fundamentales que han guiado la realización de esta tesis.

En el **objetivo 1**, creímos necesario y pertinente realizar una revisión teórica de la literatura existente sobre fusión de datos multimodales para detectar el estado actual y los retos futuros del área de estudio. Esta revisión arrojó luz ante los enfoques más utilizados en fusión de datos educativo, las técnicas, el tipo de datos y el objetivo de la fusión.

En relación al tipo y la fuente de los datos fusionados, se ha apreciado, en primer lugar, que existe un uso bastante equilibrado en los diferentes entornos educativos, ya que la fusión de datos se ha encontrado en 11 artículos centrados en aprendizaje presencial, 8 en aprendizaje online y 7 en entornos híbridos.

También se detectó que la gran mayoría de datos fusionados incluyen algún aspecto concreto relacionado con los aprendices, habiendo una minoría de trabajos centrados en datos del profesor. En este sentido, sería interesante combinar en una misma investigación datos de profesor y de estudiantes para determinar si el comportamiento de los estudiantes puede estar influenciado por las características del profesor, o si, en el otro sentido, el profesor adapta su metodología en función del tipo de estudiante al que enseña, enmarcándolo, por ejemplo, en las clásicas teorías de Biggs (1987).

En cuanto a las fuentes de datos fusionadas presentan gran variedad, habiéndose encontrado principalmente en grabaciones de los estudiantes, mediciones sensoriales de aspectos diversos, y datos numéricos que reflejan alguna magnitud generalmente relacionada con el rendimiento académico. Casi todos los datos encontrados son de naturaleza física o digital, con algunos de tipo fisiológico en menor medida. Cabe reseñar que no se ha encontrado ninguna fusión de datos de tipo psicométrico/ambiental en los procesos de fusión analizados. Sería interesante utilizarlas para, por ejemplo, poder determinar si los procesos psicológicos de los estudiantes, se ven afectados de algún modo por las características ambientales (temperatura, humedad, iluminación, etc.) en las que se desempeña su aprendizaje.

En cuanto a los objetivos de EDM/LA mejorados gracias a la fusión, destacan por número aquellos que persiguen la gestión de las emociones en los estudiantes, los que analizan el comportamiento de los estudiantes y los que predicen, tanto el desempeño académico como el interés o el *engagement*.

Por otra parte, en relación con el enfoque de fusión empleado, una importante mayoría de trabajos realizan fusión de *features* en etapa temprana (*early fusión*), existiendo también un número importante, pero menor, que realizan fusión de las decisiones obtenidas por los diferentes clasificadores en una etapa posterior (*late fusión*). Sin embargo, muy pocos trabajos realizan enfoques híbridos de las dos anteriores y menos aún se salen de este marco de referencia en el área (*early-late-hybrid*). Analizando la técnica de fusión empleada, y en consecuencia con lo anterior, hemos llegado a la conclusión de que la agregación de *features* es el enfoque predominante, seguido de otros basados en el uso de operadores estadísticos y *ensembles*. Y son precisamente estos resultados los que nos conducen a las conclusiones de los objetivos 2 y 3.

En el **objetivo número 2** nos planteamos estudiar qué enfoque de fusión de datos y algoritmos de clasificación producen los mejores resultados para predecir el rendimiento en diferentes conjuntos de datos. Para ello llevamos a cabo los dos estudios empíricos de la tesis, con dos conjuntos de datos diferentes, procedentes de dos experiecnias educativas completamente diferentes, también. En sendos estudios se pusieron a prueba 4 enfoques de fusión diferentes, dos tempranos y dos tardíos (*early fusión vs late fusión*). Para la fusión temprana se realizaron cuatro experimentos en total, dos en cada conjunto de datos, empleando la fusión de todos los atributos y la selección de atributos (Chango et al., 2021a, 2021b). Para la fusión tardía se realizaron tres experimentos, dos en Chango et al., 2021a y uno en Chango et al. 2021b, donde se usó la técnica de *ensembles* aisladamente, pero también combinada con la selección de atributos.

Usando como criterio los índices de precisión y el AUC (*Area under the ROC curve area*) de los algoritmos de clasificación, podemos concluir que los mejores resultados se obtienen usando el enfoque combinado de fusión tardía, que combinaba la técnica de ensembles con la de selección de atributos. Aun siendo unos resultados prometedores, las técnicas de fusión empleadas han sido eminentemente básicas (agregación, *ensembles* y operadores estadísticos). Cabe señalar que la disciplina de *data fusión* trabaja en enfoques mucho más avanzados que permiten mejorar la fusión realizada en diferentes ámbitos, ganando en versatilidad. El uso de técnicas basadas en filtros, enfoques probabilísticos, o el uso de la teoría de la evidencia de Dempster-Shafer se antojan útiles para tal fin aunque no hayan sido empleados para fusionar datos educativos. Del mismo modo, los experimentos han seguido los esquemas de fusión *early-late-hybrid*, pudiendo plantearse en un futuro el uso de otros tipos de esquemas más flexibles que han dado buenos resultados en ciertas investigaciones como Li et al. (2020), Qu et al. (2021) y Worsley (2014).

En otro plano, bien es cierto que no hemos podido concluir que algoritmo de clasificación es el que nos arrojaba mejores resultados, aunque este desenlace podría entrar dentro de la normalidad si tenemos en cuenta el teorema *No-Free-Lunch* (Wolpert, 2002), en el que se asume que ningún algoritmo de aprendizaje supervisado puede superar a otro algoritmo en todos los posibles problemas de aprendizaje o en diferentes conjuntos de datos.

En el último de los objetivos, el **número 3**, aquel que tiene unas mayores implicaciones prácticas, nos propusimos contrastar cómo de útiles son los modelos de predicción que producimos a partir de la fusión de datos para ayudar a los profesores a detectar a los estudiantes que están en riesgo de fracaso académico. En este sentido, los modelos de caja blanca que se obtuvieron aportan a los profesores explicaciones comprensibles (reglas IF-THEN) sobre cómo clasificaron a los alumnos en base a su rendimiento.

Asimismo, en Chango et al 2021a, observamos que los atributos que mejor discriminan en estas reglas fueron los procedentes del comportamiento de los estudiantes en Moodle, y en especial, el nivel de actividad en el foro de Moodle, resultados en consonancia con la literatura previa de Cerezo et al., 2016, y Romero et. al, 2009. Y en Chango et al., 2021b, los atributos que más aparecían en estas reglas eran los registros obtenidos a partir de logs de interacción que denotaban el uso de estrategias de resumen, la coordinación de fuentes de información obtenida a partir de datos de seguimiento ocular, y la sorpresa, de entre todas las 6 emociones básicas medidas en el estudio a través de reconocimiento facial automático. Sobre este particular, resultaría especialmente interesante repetir estos experimentos usando otro tipo de medidas que se encuentran en la literatura y que podrían enriquecer en gran medida estos modelos de predicción destinados a los educadores, como la respuesta psicogalvánica, los auto informes (Azevedo et al., 2010; 2017; Cerezo et al., 2020), y la tasa cardiaca (Huber y Bannert, 2022). En la misma línea, en la mayoría de entornos de aprendizaje existen numerosas y prolijas fuentes de datos textuales acerca del aprendizaje de los estudiantes, tales como informes, anotaciones, transcripciones, etc. Sin embargo, muy pocos trabajos hacen uso de este tipo de datos textuales. Es cierto que el análisis de texto es complejo y requiere de enfoques específicos, pero el uso de técnicas de inteligencia artificial para el procesamiento de dichos textos puede antojarse una línea interesante en el futuro, para medir si esos datos textuales fusionados con otros más comunes (video, audio, calificaciones, etc.), mejora de algún modo el análisis llevado a cabo.

Teniendo en cuenta las conclusiones de este trabajo, la discusión de sus resultados y el actual y más que probable escenario post-pandémico, que urge entornos de aprendizaje híbridos que puedan responder a circunstancias muy cambiantes, las técnicas de *Data Fusion* pueden ser la herramienta adecuada para fusionar datos procedentes de entornos y situaciones de aprendizaje multimodales que nos permitan conocer las peculiaridades del proceso de E-A que tiene lugar en estos entornos.

**Líneas Futuras y Mejoras**

Tras toda la experimentación realizada durante el desarrollo de este trabajo, podemos apuntar las siguientes líneas de trabajo futuras:

- Analizar los vídeos grabados automáticamente en lugar de hacerlo manualmente o de forma semiautomática, el procesamiento automático de las grabaciones de vídeo reuniría la información de forma más eficiente en comparación con la codificación manual. Además, el uso de múltiples cámaras web distribuidas en el aula, en lugar de una sola cámara, nos permitiría utilizar algoritmos más avanzados para detectar la participación de los estudiantes con mayor precisión.
- Utilización de otras técnicas específicas de fusión de datos, ya que hay otras teorías/métodos de fusión de datos como los métodos basados en la probabilidad (PBM) y los métodos de razonamiento de la evidencia (EBM) que podemos utilizar con datos brutos. También podríamos utilizar características de nivel semántico (abstracto) para producir una agregación inteligente de datos.
- Utilizar diferentes variables/atributos adicionales de la interacción multimodal del estudiante con el ITS, como son los datos de pensamiento en voz alta, datos de autoinforme y/o medidas fisiológicas, aspectos como las emociones de logro experimentadas de los estudiantes, sus objetivos y enfoques de aprendizaje, su autoestima y sus creencias epistemológicas pueden ayudar a mejorar los resultados de la predicción.
- También se podría utilizar algoritmos clasificadores adicionales y más avanzados, en particular el aprendizaje profundo, que podrían tener un rendimiento significativamente mejor que los métodos clásicos.

- También debido a la limitada generalidad de los resultados, el siguiente paso sería aplicar la propuesta actual en otros sistemas de aprendizaje, como los sistemas de gestión del aprendizaje (LMS) o los entornos personales de aprendizaje (PLE). Esto permitiría comparar los resultados en diferentes contextos de aprendizaje y con una mayor diversidad de temas.
- Finalmente, la mayoría de las fuentes de datos en esta área proceden de los alumnos, y sólo unos pocos estudios se han centrado en los profesores. Sería interesante combinar los datos de los profesores y los de los de los alumnos en un mismo estudio para determinar si el comportamiento de los alumnos puede estar influenciado por las características del profesor, o si el profesor adapta su metodología en función del tipo de alumno al que enseña.

# REFERENCIAS

Azevedo, R., Johnson, A., Chauncey, A., y Burkett, C. (2010). Self-regulated learning with MetaTutor: Advancing the science of learning with MetaCognitive tools. In M. Khine & I. Saleh (Eds.), *New science of learning* (pp. 225-247). Springer.

Azevedo, R., Taub, M., y Mudrick, N. V. (2017). Understanding and reasoning about real-time cognitive, affective, and metacognitive processes to foster self-regulation with advanced learning technologies. In P. A. Alexander, D. H. Schunk , J. A. Greene (Eds.), *Handbook of Self-Regulation of Learning and Performance.* Routledge.

Castanedo, F. (2013). A review of data fusion techniques. *The scientific world journal, 2013,* Article 704504, 1-19. https://doi.org/10.1155/2013/704504

Chango, W., Cerezo, R., y Romero, C. (2021a). Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses. *Computers and Electrical Engineering*, *89,* Article 106908. https://doi.org/10.1016/j.compeleceng.2020.106908

Chango, W., Cerezo, R., Sanchez-Santillan, M., Azevedo, R., y Romero, C. (2021b). Improving prediction of students' performance in intelligent tutoring systems using attribute selection and ensembles of different multimodal data sources. *Journal of Computing in Higher Education*, *33*(3), 614–634. https://doi.org/10.1007/s12528-021-09298-8

Chango, W., Lara, J. A., Cerezo, R., y Romero, C. (2022). A Review on Data Fusion in Multimodal Learning Analytics. *WIREs Data Mining and Knowledge Discovery.* Advance online publication. https://doi.org/10.1002/widm.1458

Biggs, J. (1987). *Student approaches to learning and studying.* Australian Council for Educational Research

Budaher, J., Brun, A., y Boyer, A. (2020). *Multi-source Data Mining for e-Learning*. arXiv preprint arXiv:2009.08791.

Cerezo, R., Sánchez-Santillán, M., Paule-Ruiz, M. P., y Núñez, J. C. (2016). Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. *Computers & Education*, *96*, 42-54.

Cerezo, R., Fernández, E., Gómez, C., Sánchez-Santillán, M., Taub, M., y Azevedo, R. (2020). Multimodal protocol for assessing metacognition and self-regulation in adults with learning difficulties. *JoVE (Journal of Visualized Experiments*), 163, e60331.

Chen, X., Zou, D., Xie, H., y Wang, F. L. (2021). Past, present, and future of smart learning: a topic-based bibliometric analysis. *International Journal of Educational Technology in Higher Education, 18*(1), 1–29. https://doi.org/10.1186/S41239-020-00239-6

Ding, W., Jing, X., Yan, Z., y Yang, L. T. (2019). A survey on data fusion in internet of things: Towards secure and privacy-preserving fusion. *Information Fusion*, *51*, 129–144. https://doi.org/10.1016/J.INFFUS.2018.12.001

Huber, K., y Bannert, M. (2022). What happens to your body during learning with computer-based environments? Exploring negative academic emotions using psychophysiological measurements. *Journal of Computers in Education*, 1-27.

Kaur, R., y Kautish, S. (2019). Multimodal Sentiment Analysis: A Survey and Comparison. *International Journal of Service Science, Management, Engineering, and Technology (IJSSMET)*, *10*(2), 38–58. https://doi.org/10.4018/IJSSMET.2019040103

Li, J., Zhong, J., Yang, J., y Yang, C. (2020). An Incremental Learning Framework to Enhance Teaching by Demonstration Based on Multimodal Sensor Fusion. *Frontiers in Neurorobotics*, *0*, 55. https://doi.org/10.3389/FNBOT.2020.00055

Martínez Monés, A., Dimitriadis Damoulis, I., Acquila Natale, E., Álvarez, A., Caeiro Rodríguez, M., Cobos Pérez, R., ... y Sancho Vinuesa, T. (2020). Achievements and challenges in learning analytics in Spain: The view of SNOLA. RIED. *Revista Iberoamericana de Educación a Distancia, 23*(2), 187.

Mu, S., Cui, M., y Huang, X. (2020). Multimodal Data Fusion in Learning Analytics: A Systematic Review. *Sensors, 20*(23), 6856. https://doi.org/10.3390/S20236856

Romero, C., y Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), e1355. https://doi.org/10.1002/WIDM.1355

Romero, C., González, P., Ventura, S., del Jesús, M. J., y Herrera, F. (2009). Evolutionary algorithms for subgroup discovery in e-learning: A practical application using Moodle data. *Expert Systems with Applications*, *36*(2), 1632-1644.

Poria, S., Cambria, E., Bajpai, R., y Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, *37*, 98–125. https://doi.org/10.1016/J.INFFUS.2017.02.003

Salta, K., Paschalidou, K., Tsetseri, M., y Koulougliotis, D. (2022). Shift from a traditional to a distance learning environment during the COVID-19 pandemic. *Science & Education, 31*(1), 93-122. https://doi.org/10.1007/s11191-021-00234-x

Sharma, K., y Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology*, *51*(5), 1450–1484. https://doi.org/10.1111/BJET.12993

Sultana, M., Jibon, F. A., y Kashem, A. (2020). An Improved Model of Virtual Classroom using Information Fusion and NS-DBSCAN. *Global Journal of Computer Science and Technology*, *20*(5). https://computerresearch.org/index.php/computer/article/view/1966

Tabuenca, B., Serrano-Iglesias, S., Martin, A. C., Villa-Torrano, C., Dimitriadis, Y., I. Asensio-Perez, J., … Kloos, C. D. (2021). Affordances and Core Functions of Smart Learning Environments: A Systematic Literature Review. *IEEE Transactions on Learning Technologies*, *14*(2), 129–145. https://doi.org/10.1109/TLT.2021.3067946

Qu, J., Liu, A., y Liu, R. (2021). Research on Evaluation and Confirmation of College Students' Learning Behavior Based on Comprehensive Weighted Fusion Algorithm. *2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2021*, 121–125. https://doi.org/10.1109/ICCCBDA51879.2021.9442584

United Nations. (2020). *Policy brief: Education during COVID-19 and beyond.* New York. https://unsdg.un.org/sites/default/files/2020-08/sg_policy_brief_covid-19_and_education_august_2020.pdf. Accessed 24 April 2021

UNESCO. (2020). *Education: From disruption to recovery.* https://en.unesco.org/covid19/educationresponse. Accessed 24 April 2021.

Wang, R., Ji, W., Liu, M., Wang, X., Weng, J., Deng, S., Gao, S. y Yuan, C. (2018). Review on mining data from multiple data sources. *Pattern Recognition Letters*, *109*, 120–128. https://doi.org/10.1016/j.patrec.2018.01.013

Worsley, M. (2014). Multimodal learning analytics as a tool for bridging learning theory and complex learning behaviors. *MLA 2014 - Proceedings of the 2014 ACM Multimodal Learning Analytics Workshop and Grand Challenge, Co-Located with ICMI 2014*, 1–4. https://doi.org/10.1145/2666633.2666634

Wolpert D.H. (2002). The Supervised Learning No-Free-Lunch Theorems. In Roy R., Köppen M., Ovaska S., Furuhashi T., Hoffmann F. (Eds), *Soft Computing and Industry.* Springer. https://doi.org/10.1007/978-1-4471-0123-9_3

# Parte II. Publicaciones

Received: 1 February 2022 | Revised: 12 March 2022 | Accepted: 15 March 2022

DOI: 10.1002/widm.1458

OVERVIEW

WIREs DATA MINING AND KNOWLEDGE DISCOVERY

WILEY

# A review on data fusion in multimodal learning analytics and educational data mining

**Wilson Chango**[1] | **Juan A. Lara**[2] | **Rebeca Cerezo**[3] | **Cristóbal Romero**[4]

[1]Department of Systems and Computing, Pontifical Catholic University of Ecuador, Quito, Ecuador

[2]Department of Computer Science, Madrid Open University, Madrid, Spain

[3]Department of Psychology, University of Ovideo, Oviedo, Spain

[4]Department of Computer Science, University of Cordoba, DaSCI Institute, Córdoba, Spain

**Correspondence**
Cristóbal Romero, Department of Computer Science, University of Cordoba, DaSCI Institute, Córdoba, Spain.
Email: cromero@uco.es

**Funding information**
Universidad de Córdoba/CBUA; Ministry of Sciences and Innovation I + D + I, Grant/Award Numbers: PID2019-107201GB-I00, PID2020-115832GB-I00

**Edited by:** Witold Pedrycz, Editor-in-Chief

**Abstract**

The new educational models such as smart learning environments use of digital and context-aware devices to facilitate the learning process. In this new educational scenario, a huge quantity of multimodal students' data from a variety of different sources can be captured, fused, and analyze. It offers to researchers and educators a unique opportunity of being able to discover new knowledge to better understand the learning process and to intervene if necessary. However, it is necessary to apply correctly data fusion approaches and techniques in order to combine various sources of multimodal learning analytics (MLA). These sources or modalities in MLA include audio, video, electrodermal activity data, eye-tracking, user logs, and click-stream data, but also learning artifacts and more natural human signals such as gestures, gaze, speech, or writing. This survey introduces data fusion in learning analytics (LA) and educational data mining (EDM) and how these data fusion techniques have been applied in smart learning. It shows the current state of the art by reviewing the main publications, the main type of fused educational data, and the data fusion approaches and techniques used in EDM/LA, as well as the main open problems, trends, and challenges in this specific research area.

This article is categorized under:

Application Areas > Education and Learning

KEYWORDS

data fusion, educational data science, multimodal learning, smart learning

## 1 | INTRODUCTION

The current, and more than likely post-pandemic, scenario seems to point toward new hybrid, more flexible and technological learning environments that can respond to changing circumstances. In this regard, blended learning (BL), hybrid learning (HL), and smart learning (SL) are options that comes up repeatedly:

- Hybrid learning (HL) is an educational approach where some individuals participate in person, and some participate online. Instructors and facilitators teach remote and in-person learners at the same time using technology like video conferencing (Raes, 2022).
- Blended learning (BL) is a style of education in which instructors and facilitators combine in-person instruction with online learning activities. Learners complete some components online and do others in person (Sánchez Ruiz et al., 2021).
- Smart learning environments (SLEs) are physical environments enriched with digital, context-aware, adaptive devices which aim to achieve more effective, better-quality learning (X. Chen et al., 2021; Tabuenca et al., 2021). SLEs contain multiple sources of data which, combined together, can offer a better understanding of the educational process.

All these new type of learning environments produce a huge amount of student's data interaction. In the last decade, there were an increasing interest in the analysis and exploitation of large amounts of data produced during the learning process in these new educational environments, which are difficult to analyze manually. In fact, there are two related communities about the same educational data science (EDS) research area (Romero & Ventura, 2020):

- Educational data mining (EDM) can be defined as the application of data mining (DM) techniques to this specific type of dataset that come from educational environments to address important educational questions.
- Learning analytics (LA) can be 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.

Both communities share a common interest in data-intensive approaches to educational research and share the goal of enhancing educational practice. However, LA is more focused on the educational challenge and EDM is more focused on the technological challenge. On the one hand, LA is focused on data-driven decision making and integrating the technical and the social/pedagogical dimensions of learning by applying known predictive models. On the other hand, EDM is generally looking for new patterns in data and developing new algorithms and/or models. Regardless of the differences between the LA and EDM communities, the two areas have significant overlap both in the objectives of investigators as well as in the methods and techniques that are used in the investigation (Romero & Ventura, 2020).

Normally, most of the EDM/LA approaches to analyzing educational data are based on using only one specific data source. However, this means being limited by the data source used, which reflects only part of the reality of the educational process. This is a problem in SLEs which produces a fast quantity of data from different sources that make appropriate the use of data fusion techniques for merging all information to correctly understand the peculiarities of the teaching-learning process occurring in these environments. This idea of combined use of several educational data sources has given rise to multimodal learning analytics (MLA). This approach is based on capturing, integrating, and analyzing different sources of educational data which together provide a holistic understanding of the learning process (Sharma & Giannakos, 2020). During multimodal interaction in education environments, new data collection and sensing technologies are making it possible to capture massive amounts of data about students' activity. These technologies include the logging of computer activities, wearable cameras, wearable sensors, biosensors (e.g., that permit measurements of skin conductivity, heartbeat, and electroencephalography), gesture sensing, infrared imaging, and eye tracking. Such techniques enable researchers to have unprecedented insight into the minute-by-minute students' activities, especially those involving multiple dimensions of activity and social interaction (Blikstein & Worsley, 2016). The combination of multimodal data treatment techniques and the intersection with EDM and LA has been shown to be a productive line of study in recent years (Budaher et al., 2020; Kaur & Kautish, 2019; Lahat et al., 2015; Poria et al., 2017; Wang et al., 2018). For example, Di Mitri et al. (2019) proposed a mechanism allowing annotation of multimodal data for subsequent analysis, and Järvelä et al. (2021) gave many examples of the advantages offered by multimodal data with regard to self-regulated learning. Despite numerous advantages, combined use of educational data is not easy, and there are notable challenges, such as differing granularity or the need to align the different timescales for the data collected from different sources.

Data fusion can be defined as the process of effectively combining data from different sources so that using that data in combination produces more information than each of the sources would separately. In SLEs, this idea has been used to try and exploit multimodal data and better understand the educational process. The general approach of multimodal learning data fusion and mining in smart classroom is shown in Figure 1. Multimodal data come from different educational environment such as traditional classroom, e-learning or blended and hybrid learning, and different sources or data type. The fusion point and the used fusion technique depend on the educational problem to solve and the DM/LA objective. Finally, new knowledge can be discovered after applying this process for improving our smart classroom.

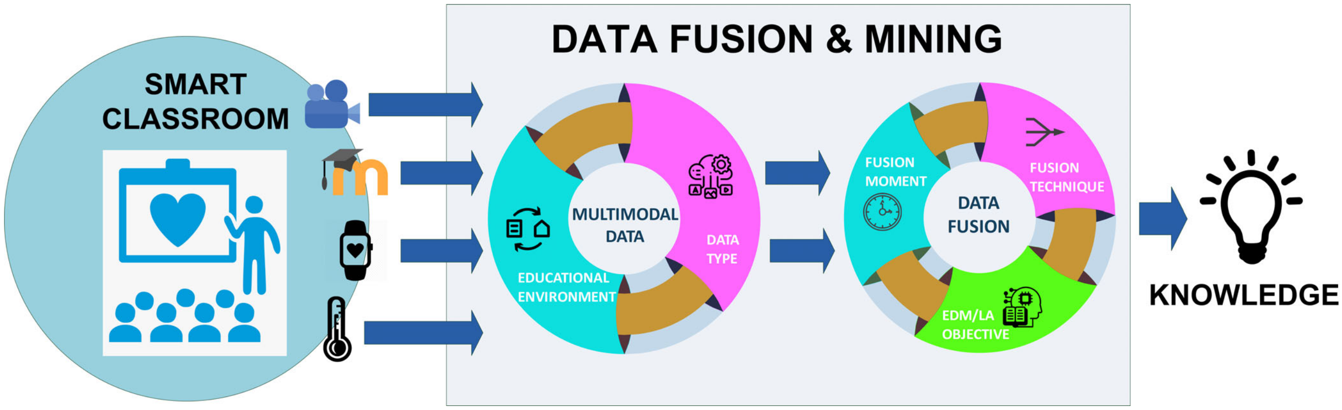


FIGURE 1 General multimodal data fusion approach for EMD/LA

In recent years, there have been an increasing number of survey papers about multimodal educational data (Blikstein & Worsley, 2016; Ochoa, 2017; Shankar et al., 2018). These works examined the application of EDM/LA in multimodal educational data, but which barely touched on data fusion, focusing instead on complex learning tasks (Blikstein & Worsley, 2016), the study of LA architectures (Shankar et al., 2018), and the study of learning environments where multimodal LA is usually applied (Ochoa, 2017). There are also a few review papers more focused in the specific application of data fusion in EDM/LA (Dewan et al., 2019; Han et al., 2020; Nandi et al., 2020). However, they only focused on some specific aspects, including emotion recognition (Nandi et al., 2020), engagement detection (Dewan et al., 2019), or sentiment analysis (Han et al., 2020). Finally, the survey that is most closely related with our current review is from Mu et al. (2020), which focused only on LA, without examining EDM bibliography. That survey only presented the papers analyzed quantitatively, without deeper analysis of the different studies and without establishing the challenges and lines of future study for researchers in this area. So, it is clear that these previous existed and related reviews give an incomplete picture, meaning that there is a need for an up-to-date, comprehensive review of the literature on studies about the specific use of data fusion techniques in SLEs for the application both in LA and EDM. Our objective in this review is to provide a in depth analyses of all the multimodal data used (types, capture methods, etc.), a description of all the data fusion methods and techniques used, the LA and EDM objectives and successful applications, and to identify a set of open challenges and problems. In this way, we will provide the scientific community with a thorough, up-to-date understanding of the current state of this discipline.

We followed the systematic literature review procedure proposed by Tranfield et al. (2003). We used Google Scholar, Web of Science, and Scopus search engines to search for academic papers up to December 2021. In our search we used the following search terms: "Data Fusion" AND ("Multimodal Learning" OR "E-learning" OR "Online learning" OR "Web-based learning" OR "Blended Learning" OR "Hybrid learning" OR "Smart Learning" OR "Education"). This preliminary search identified 56 papers whose titles or abstracts included the defined keywords. Then, the papers were selected by reading both the full content of the papers initially downloaded from the search and applying the following inclusion and exclusion rule. We only considered studies in which there was a real educational data fusion process with the aim of applying LA or EDM techniques. It did not include studies which merely used multimodal data from different sources separately, such as Järvelä et al. (2021), nor studies that did use fusion of educational data but without the aim of applying LA/EDM techniques. In this way, we finally selected only 31 papers (20 journal papers and 11 conference papers) published between 2015 and 2021, which confirms the relatively novel nature of this topic.

The remainder of this article is organized as follows: Section 2 provides an analysis of the selected studies according to the type of fused multimodal data; Section 3 analyzes those studies from the perspective of the fusion approach or technique used; and finally, Section 4 presents our conclusions and outlines the identified challenges and open problems in this area.

## 2 | MULTIMODAL DATA

In this section, we analyze what are the most common data used when fusing MLA data. We have differentiated two different viewpoints or fundamental aspects: the educational environment used (in-person, online, and hybrid/blended) and the type of data fused.

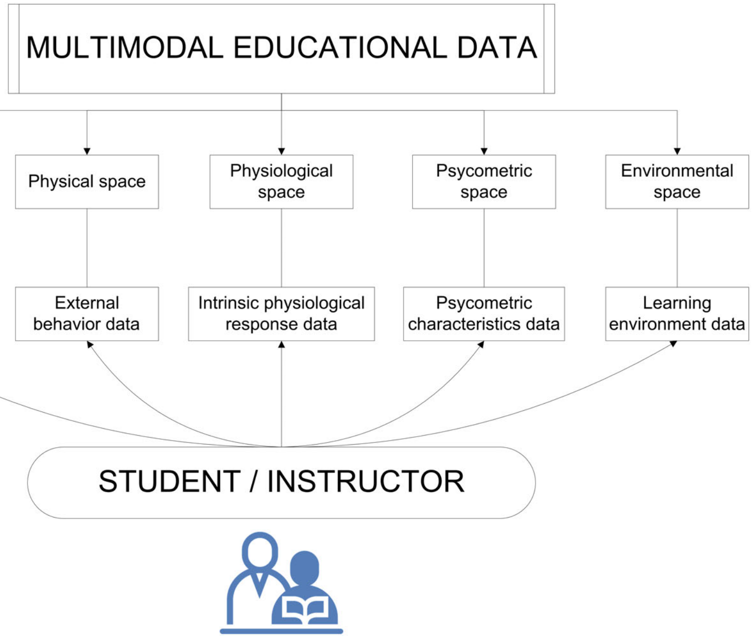


**FIGURE 2** Categories of multimodal educational data, based on Mu et al. (2020)

In the next two subsections, we have analyzed the previously selected papers. For each source of fused data identified, we show its name, type (audio, video, numerical, etc.), method of capture (camera, microphone, log, etc.) and category. The category will be analyzed using the classification from Mu et al. (2020), summarized in Figure 2, which establishes five different categories of data: digital, physical, physiological, psychometric, and environmental. Digital space referred to various digital traces generated on the system platform during the learning process, such as an online learning platform, virtual experiment platform, or STEAM educational software. Physical space was about the data obtained by various sensors, such as gesture, posture, and body movement. Physiological space referred to the data related to human internal physiological reflection, including EEG and ECG, which objectively reflected students' learning status. In contrast, psychometric space, a relatively common source of learning data, referred to various self-report questionnaires that subjectively reflected the learner's mental state. Environmental space referred to the data about a learning environment where a learner was physically located, such as temperature and weather.

## 2.1 | Traditional in-person classroom data

This section presents the different data source from face-to-face traditional teaching. With classroom-based learning, students go to a physical classroom where the teaching and much of the learning takes place. Table 1 shows the papers that used these types of data for fusing them. It presents the reference in the first column, the different sources of fused data in the second, and for each source, the type and category (according to the taxonomy in Figure 2), and finally the capture method (Webcam, SMI Eye Tracking Glasses, Electrode, Different Sensors, CSV files, and Platform).

As the table shows, there is a wide variety of data sources to fuse in face-to-face teaching. In Giannakos et al., 2019, different physical and physiological student data were fused from multiple sensors, including heart rate, body temperature, and blood volume. Some studies used fusion of video and text data (Daoudi et al., 2021), while others fused data from recordings made using 180-degree video cameras (Gadaley et al., 2020). Some of the studies used many different types of sources (Olsen et al., 2020), whereas others focused on specific types of data, such as images in Mao et al. (2019). Most of the studies focused on data collected from students, but some, such as Prieto et al. (2018), used data gathered from wearables worn by the teacher. It is also interesting to note the fusion of multimedia data (audio and video) together with data from students' hands while they performed certain tasks (Worsley, 2014). The only researchers who contributed with several papers to this survey were Henderson et al., whose studies focused on student posture and movement along with other data that varied from one study to the next (N. Henderson et al., 2020; N. L. Henderson, Rowe, Mott, & Lester, 2019). It was also interesting to see the fusion between physical and digital data gathered via

**TABLE 1** Sources of multimodal data from in-person classroom

| Paper | Source | Type | Category | Capture method |
|---|---|---|---|---|
| Giannakos et al., 2019 | (Student) heart rate | Time series | Physical | Sensor |
| | Electrodermal activity | Time series | Physiological | Sensor |
| | Body temperature | Time series | Physiological | Sensor |
| | Blood volume | Time series | Physical | Sensor |
| | Electroencephalogram | Numerical | Physiological | Electrode |
| | Eye tracking | Video | Physiological | Webcam |
| Daoudi et al., 2021 | Real-time video recordings | Video | Physical | Webcam |
| | Exchanged messages during playing | Text | Digital | CSV |
| Gadaley et al., 2020 | Student attention to a video source | Video | Physical | Webcam |
| | Student head position to determine attention | Time series | Physical | Webcam |
| Olsen et al., 2020 | Student audio | Audio | Physical | Webcam |
| | Eye tracking | Video | Physical | Webcam |
| | Questionnaire type test | Numerical | Digital | Platform |
| | Cognitive load by gaze analysis | Time series | Physical | Sensor |
| | Dialogue between students | Text | Digital | Platform Log |
| Mao et al., 2019 | Student images | Photographs | Physical | Webcam |
| Prieto et al., 2018 | Teacher eye-tracking | Video | Physical | SMI glasses |
| | Teacher movement in the classroom | Time series | Physical | Sensor |
| | Teacher's presentation | Audio | Physical | SMI glasses |
| | Teacher's video lessons | Video | Physical | SMI glasses |
| Worsley, 2014 | Oral test for student | Audio | Physical | Webcam |
| | Student behavior | Video | Physical | Webcam |
| | Hand movements | Time series | Physical | Sensor |
| N. Henderson et al., 2020 | Student posture (Kinect) | Time series | Physical | Sensor |
| | Following movement | Time series | Physical | Sensor |
| | Interaction between students | Video | Physical | Webcam |
| | Student actions | Text | Digital | Platform |
| N. L. Henderson, Rowe, Mott, & Lester, 2019 | Student posture (Kinect) | Time series | Physical | Sensor |
| | Following movement | Time series | Physical | Sensor |
| | Interaction between students | Video | Physical | Webcam |
| | Recordings to identify student frustration | Video | Physical | Webcam |
| | Students' hands (temperature, electrodermal activity and 3D coordinates) | Time series | Physical/ physiological | Sensor |
| Ma et al., 2015 | Recording of the class | Video | Physical | Webcam |
| | Interaction between students and teacher | Video | Physical | Webcam |
| | User and course information | Text | Digital | Platform |
| | Session start time | Numerical | Digital | Platform |
| | Browser history | Text | Digital | Platform |
| Andrade et al., 2016 | Interviews with students | Text | Digital | Logs |
| | Gaze tracking | Time Sees | Physical | Sensor |
| | Student gestures | Video | Physical | Webcam |

(Continues)

TABLE 1 (Continued)

| Paper | Source | Type | Category | Capture method |
|---|---|---|---|---|
| Monkaresi et al., 2017 | Writing activity | Text | Digital | Platform |
| | Facial expressions | Video | Physical | Webcam |
| | Hear rate | Time series | Physiological | Sensor |

webcam and the learning platform in Ma et al. (2015). The study by Andrade et al. (2016) stood out as it included the analysis of children's gazes and gestures during certain learning activities. Finally, automated detection of student's engagement has done by fusing information from writing activity, videos of their faces after the activity and heart rate (Monkaresi et al., 2017).

## 2.2 | Online classroom data

This section presents the different data sources from online learning. Online education uses the Internet and information and communications technology (ITC) to provide students with tools like chats, blogs, video conferences and shared documents. Table 2 shows the papers that used these types of data for fusing them, using the same column structure as in the previous section.

In online education, there was also variation between the different studies. The work presented in Wu et al. (2020) stood out for focusing on the teacher and for analyzing teacher gestures, behavior, and especially body position/pose, as well as proposing a general model of human joint positions which they used to model teacher movement in an open online classroom. Some studies, such as Brodny (2017), fused various video sources and student data from the platform. The study by Peng and Nagao (2021) stood out for the wide range of different sources, including heart rate and mental states via video and text; Luo et al. (2020) also fused video and text data, including data about posture, facial expressions, and thoughts. In N. L. Henderson, Rowe, Mott, Brawner, et al. (2019), the researchers fused data about posture and electrodermal activity, this time in a fully online environment (the same authors have also looked at in-person settings). The study by Liu et al. (2019) had a wide range of different types of data with numerous means for capturing it, whereas Nam Liao et al. (2019) went in the opposite direction, as it only included numerical digital data. Yue et al. (2019) used a wide range of fused data (facial expressions, eye tracking, grades, etc.) and was the only study to include an open data source in the fusion. Di Mitri et al., 2017 is the only work that used learning environment data such as temperature, pressure, precipitation, and weather type together with heart rate and step count in self-regulated learning. Finally, two works used psychometrics data. Sharma et al. (2019) proposed stimuli-based gaze analytics to enhance motivation and learning in MOOCs while the student's eye-movements were recorded. They also used a motivation scale from a 5-point Likert questionnaire. Hussain et al. (2011) detect learners' affective states from multichannel physiological signals (heart activity, respiration, facial muscle activity, and skin conductivity) during tutorial interactions with AutoTutor, an intelligent tutoring systems (ITS) with conversational dialogues. They also asked learners to provide self-reports of affect based on both categorical and dimensional (valence/arousal) models.

## 2.3 | Hybrid and blended classroom

This section presents the different data fused from hybrid and blended learning environments. Both types of learning involve a mix of in-person and online learning, but the who differs in the two scenarios. With hybrid learning, the in-person learners and the online learners are different individuals. With blended learning, the same individuals learn both in person and online. Table 3 show the papers that used these types of data for fusing them, using the same structure as previous sections.

In hybrid or semi-in-person education, the work by Chango, Cerezo, and Romero (2021); Chango, Cerezo, Sanchez-Santillan, et al. (2021) stands out for the fusion of different types of class recordings with data obtained through Moodle, while Xu et al. (2019) fused video and text of the teacher both explaining various ideas in class and answering students' questions. The study by J. Chen et al. (2014) stood out by including probably the greatest number and widest variety of

**TABLE 2** Sources of multimodal data in online settings

| Study | Source | Type | Category | Capture method |
|---|---|---|---|---|
| Wu et al., 2020 | Teacher gestures (indicative, descriptive, or rhythmical) | Video | Physical | Webcam |
| | Teacher behavior (writing on the board, asking questions, demonstrating, instructing, describing, and non-gesture behavior) | Video | Physical | Webcam |
| | Teacher body movement | Time series | Physical | Sensor |
| Brodny, 2017 | Facial expressions | Video | Physical | Webcam |
| | Self-report (key-presses and mouse movement patterns) | Video | Physical | Webcam |
| | Physiological signals | Video | Physiological | Webcam |
| | Moodle course (activities, questions, and forum) | Numerical | Digital | Platform |
| Peng & Nagao, 2021 | Student heart rate | Time series | Physical | Sensor |
| | Conversations with the teacher | Text | Digital | Microphone |
| | Students' mental states | Video | Physical | Webcam |
| Luo et al., 2020 | Head position to measure cognitive attention | Video | Physical | Webcam |
| | Facial expressions (smiles) | Video | Physical | Webcam |
| | Student thoughts | Text | Digital | Platform |
| N. L. Henderson, Rowe, Mott, Brawner, et al., 2019 | Student posture (Microsoft Kinect) | Time series | Physical | Sensor |
| | Student skin temperature and electrodermal activity | Time series | Physiological | Sensor |
| Liu et al., 2019 | Speech between students | Audio | Physical | Microphone |
| | Student interaction with the system interface | Video | Physical | Webcam |
| | Student interactions with the teacher | Video | Physical | Webcam |
| | Student activities | Text | Digital | Platform |
| | Evaluation records | Numerical | Digital | Record in CSV |
| Nam Liao et al., 2019 | Student pre-requisites | Numerical | Digital | Platform |
| | Multiple-choice questionnaires | Numerical | Digital | Platform |
| | Individual student tasks | Numerical | Digital | Platform |
| Yue et al., 2019 | Facial expressions | Video | Physical | Webcam |
| | Eye movement | Time series | Physical | Sensor |
| | Open-source image dataset for learning Asian faces | Image | Digital | Open source |
| | Dynamic mouse records for performance analysis | Time series | Digital | Log |
| | Student scores | Numerical | Digital | Log |
| Di Mitri et al., 2017 | Leg step count and Heart rate | Time series | Physical | Sensor |
| | Weather condition | Time series | Environmental | Platform |
| Sharma et al., 2019 | Eye movement | Time series | Physical | Sensor |
| | Motivation from questionnaire | Numerical | Psychometric | Platform |
| | Student scores | Numerical | Digital | Log |

(Continues)

**TABLE 2** (Continued)

| Study | Source | Type | Category | Capture method |
|---|---|---|---|---|
| Hussain et al., 2011 | Electrocardiogram<br>Facial electromyogram<br>Respiration<br>Galvanic<br>Skin response<br>Facial expressions<br>Self-report affective state | Time series<br>Time series<br>Video<br>Numerical | Physiological<br>Physiological<br>Physiological<br>Physiological<br>Physical<br>Psychometric | Sensor<br>Sensor<br>Sensor<br>Sensor<br>Webcam<br>Platform |

**TABLE 3** Sources of multimodal data in hybrid and blended settings

| Study | Source | Type | Category | Capture method |
|---|---|---|---|---|
| Chango, Cerezo, & Romero, 2021; Chango, Cerezo, Sanchez-Santillan, et al., 2021 | Theory classes (attendance, attention, and notetaking) | Video | Physical | Webcam |
| | In-person practical classes (attendance and scores) | Video | Physical | Webcam |
| | Online student interactions with the platform | Numerical | Digital | Platform |
| | Final exam score | Numerical | Digital | Platform |
| Xu et al., 2019 | Classes given by the teacher | Video | Physical | Webcam |
| | Teacher speech | Text | Digital | Log |
| | Questions asked in class | Text | Digital | Log |
| J. Chen et al., 2014 | Head position | Video | Physical | Webcam |
| | Gaze tracking | Video | Physical | Webcam |
| | Facial expression | Video | Physical | Webcam |
| | Student electrodermal activity | Time series | Physiological | Sensor |
| | Student evaluation (attempts to answer questions, correct/incorrect responses, and final score) | Text | Digital | Log |
| Bahreini et al., 2016 | Facial features to detect emotions | Video | Physical | Webcam |
| | Vocal features to detect emotions | Audio | Physical | Microphone |
| Li et al., 2020 | Teacher/demonstrator body movement (Kinect) | Time series | Physical | Sensor |
| | Teacher/demonstrator joint positions (Myo armbands) | Time series | Physical | Sensor |
| Qu et al., 2021 | Classroom teaching data (performance, exam results) | Numerical | Digital | Log |
| | Online teaching data (performance, exam results) | Numerical | Digital | Log |
| | Offline teaching data (performance, exam results) | Numerical | Digital | Log |
| Shankar et al., 2019 | Digital tool adaptors | Numerical | Digital | CSV |
| | IoT adaptors | Numerical | Digital | CSV |

data sources to fuse, including posture, gaze, electrodermal activity, and student evaluation data. In contrast, Bahreini et al. (2016) performed emotion detection from the fusion of video of student faces and recordings of their voices. The study by Li et al. (2020) was innovative in the EDM/LA field, recording the body movement of the teacher/

demonstrator using sensors and armbands, starting with a model of a human being emulated by the movement of a robot. Finally, Qu et al. (2021) fused different numerical student performance data, presenting little variety of data types in the study. Numerical data were also fused in Shankar et al. (2019), although in this case from what the authors called adaptors, one of which gathered data from the students' digital environments, while the other—the IoT adaptor—collected data from sensors physically located in the learning environment. That makes this study a good example of a hybrid using a physical-digital fusion.

## 3 | DATA FUSION TECHNIQUES IN MULTIMODAL LA/EDM

This section aims to analyze the fusion process of multimodal educational data. In the next subsection we describe three fundamental aspects of this process: when the fusion is done or fusion point, what are the most used data fusion techniques, and in which EDM/LA applications/objectives data fusion has been used more.

### 3.1 | When fusion is done

Data fusion techniques can be characterized in different ways depending on the area of application. Figure 3 shows the most widespread—practically standardized—categorization of MLA data fusion.

As we can see in Figure 3, data fusion techniques can be classified based on when the fusion is done, giving rise to the three following three main types (Ding et al., 2019):

- Feature-level or early fusion: a fusion approach consisting of concatenating the various features of the data from the different sources in a single vector of heterogeneous elements.
- Decision-level or later fusion: a fusion approach which consists of first creating a classifier with each of the data sources separately in order to subsequently fuse the prediction offered by the different classifiers.
- Hybrid fusion: a fusion approach which uses the two approaches above in a single fusion process.

Table 4 categorize the selected papers by the fusion point or the moment in which the fusion is done (early, later and hybrid fusion). It is important to note that some papers may appear in multiple categories due to they have used different time points where fusion was done. We also found some studies that do not fit into none of those three groups or in which the fusion point was not specified (Others category).

Five of the studies which appeared in the early fusion category stand out for going beyond simple concatenation of features with rather more detailed procedures. Four of those studies were configured to select the best features of each

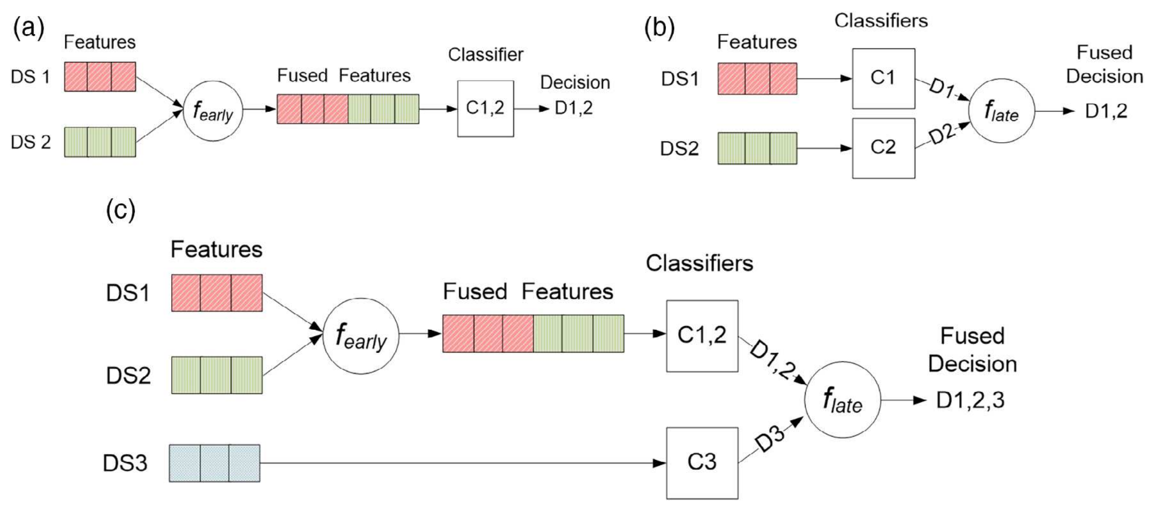


FIGURE 3 Multimodal data fusion schema according to when fusion is done. (a) Feature or early fusion. (b) Decision or late fusion. (c) Hybrid fusion

**TABLE 4** Categorization of papers by fusion point

| Fusion point | Explanation | Studies |
|---|---|---|
| Early | Concatenation of the features of the different data sources | Andrade et al., 2016; Bahreini et al., 2016; Chango, Cerezo, & Romero, 2021; Chango, Cerezo, Sanchez-Santillan, et al., 2021; Gadaley et al., 2020; Giannakos et al., 2019; N. L. Henderson, Rowe, Mott, & Lester, 2019; N. L. Henderson, Rowe, Mott, Brawner, et al., 2019; N. Henderson et al., 2020; Liu et al., 2019; Mao et al., 2019; Nam Liao et al., 2019; Olsen et al., 2020; Peng & Nagao, 2021; Prieto et al., 2018; Shankar et al., 2019; Wu et al., 2020; Xu et al., 2019; Yue et al., 2019; Di Mitri et al., 2017; Sharma et al., 2019; Hussain et al., 2011 |
| Later | Fusion of the predictions of each classifier (each created from a data source) | Chango, Cerezo, & Romero, 2021; Chango, Cerezo, Sanchez-Santillan, et al., 2021; J. Chen et al., 2014; Daoudi et al., 2021; Peng & Nagao, 2021; Wu et al., 2020; N. L. Henderson, Rowe, Mott, & Lester, 2019; N. L. Henderson, Rowe, Mott, Brawner, et al., 2019; Monkaresi et al., 2017 |
| Hybrid | A mix of the two approaches above | Brodny, 2017; Chango, Cerezo, & Romero, 2021; Chango, Cerezo, Sanchez-Santillan, et al., 2021; Luo et al., 2020 |
| Others | Approaches that do not fit within the three described above | Li et al., 2020; Qu et al., 2021; Worsley, 2014; Ma et al., 2015 |

data source (Chango, Cerezo, & Romero, 2021; Chango, Cerezo, Sanchez-Santillan, et al., 2021; N. L. Henderson, Rowe, Mott, Brawner, et al., 2019; N. Henderson et al., 2020). In contrast, N. L. Henderson, Rowe, Mott, and Lester (2019), reduced the dimensionality of the features using principal component analysis (PCA) in two different configurations: (a) they concatenated all of the features of the sources and applied PCA to the resulting vector; (b) they applied PCA to the features of each source first and concatenated the results following the reduction of dimensionality. Yue et al. (2019) selected the best features first and then reduced dimensionality using two approaches, PCA and a Kolmogorov–Smirnov test. The other studies in the early category based fusion on mere concatenation of the features extracted from each source into a single vector of features which fed into the subsequent analysis.

There were also a number of studies in the later or decision fusion category, based on fusing the predictions made by the different classifiers constructed from the different data sources. Three studies used the "ensembles" approach to fuse the decisions (Chango, Cerezo, & Romero, 2021; Chango, Cerezo, Sanchez-Santillan, et al., 2021; Wu et al., 2020). There were also four studies which based decision fusion on the decision made by the classifier with the best predictive ability (N. L. Henderson, Rowe, Mott, & Lester, 2019; N. Henderson et al., 2020; Peng & Nagao, 2021; Monkaresi et al., 2017). The fusion in N. L. Henderson, Rowe, Mott, Brawner, et al. (2019) was done by consolidating partial decisions from each classifier into a single value, whereas the result of the fusion in Daoudi et al. (2021) was produced by weighting each classifier's decision. J. Chen et al. (2014) did not specify the details of their decision fusion, merely indicating that it was done. Finally, Monkaresi et al. (2017) used individual channel base classifier to make a classification by using the decision of whichever base classifier had the highest decision probability.

Various studies appeared in both the early and decision categories because the researchers made comparisons between the two approaches to determine which gave the best results. In contrast, only three studies appeared in the hybrid fusion category, using early and decision fusion in combination. Brodny (2017) proposed a conceptual model with feature fusion at the beginning followed by decision fusion, although it was laid out very broadly. In addition to the basic schemes of early and decision fusion described above, Chango, Cerezo, and Romero (2021); Chango, Cerezo, Sanchez-Santillan, et al. (2021) also used hybrid configurations at some point in which the features of some sources were fused at the beginning to produce classifiers which were subsequently fused by means of ensembles. This study also stood out for the richness of the experiments, as it compared the hybrid approach with purely early and late fusion approaches.

Finally, four studies did not fit in any of the previous categories. Li et al. (2020) used a fusion model of sensors from various data sources in order to produce more accurate data during a demonstration by an instructor for a robot to learn to do certain tasks. Qu et al. (2021) proposed a five-step fusion process which affected the features produced by the data sources via a weighting technique but which did not fit within the early fusion category because there was no concatenation of features as such. The work presented in Worsley (2014) was a general article proposing three generic fusion approaches which differed from the traditional early-decision-hybrid categorization. In contrast, the author talked about naïve fusion (equivalent to feature fusion), low-level fusion (where the researcher is intentional about enacting multimodal data fusion on very small-time scales because they may have prior knowledge that the various modalities have time-specific relevance to one another), and high-level fusion (which takes the data to a higher level of meaning, equating features to states). Ma et al. (2015) produced a process model of multisource data fusion analysis and, according to the authors, put it into practice from the dimension of data fusion. There is only one paragraph in the article explaining the fusion, but it is so abstract that it is impossible to determine when the fusion was done or what data were affected.

## 3.2 | Fusion technique

We have classified the selected papers by the used fusion technique in the main types or categories (see Table 5). The criteria for classification were the fundamental data fusion techniques used, from the purist perspective of data fusion: aggregation, ensembles, statistical operators, mathematical operators, similarity-based, probability, and filters. Again, the categories were not exclusive as some studies used more than one technique. There were also studies which did not fit into the standard data fusion categories and studies which did not specifically state the fusion technique used (Others in the table).

Perhaps the most elementary data fusion technique consists of combining data in the most basic sense of aggregating or concatenating. A large number of studies fell within this category because it is specifically the idea of concatenation of data which underlies early fusion studies in which features are aggregated or concatenated. Work which stood

TABLE 5 Categorization of papers based on fusion technique

| Technique | Explanation | Studies |
|---|---|---|
| Aggregation | Fusion consists of aggregating (in the sense of concatenating) data from the different sources | Andrade et al., 2016; Bahreini et al., 2016; Chango, Cerezo, & Romero, 2021; Chango, Cerezo, Sanchez-Santillan, et al., 2021; Gadaley et al., 2020; Giannakos et al., 2019; N. L. Henderson, Rowe, Mott, & Lester, 2019; N. L. Henderson, Rowe, Mott, Brawner, et al., 2019; N. Henderson et al., 2020; Liu et al., 2019; Nam Liao et al., 2019; Olsen et al., 2020; Peng & Nagao, 2021; Prieto et al., 2018; Shankar et al., 2019; Worsley, 2014; Wu et al., 2020; Xu et al., 2019; Yue et al., 2019; Di Mitri et al., 2017; Sharma et al., 2019; Hussain et al., 2011 |
| Ensembles | Applies the idea of ensembles (machine learning) to combine the data from the various classifiers' decisions | Chango, Cerezo, & Romero, 2021; Chango, Cerezo, Sanchez-Santillan, et al., 2021; Wu et al., 2020 |
| Statistical operators | Uses statistical operators to combine the data from the different sources | Daoudi et al., 2021; N. L. Henderson, Rowe, Mott, & Lester, 2019; N. Henderson et al., 2020; Qu et al., 2021 |
| Mathematical operators | Uses mathematical operators to combine the data from the different sources | Qu et al., 2021 |
| Similarity-based | Fusion is based on the calculation of similarities | Qu et al., 2021 |
| Probability | The fusion uses the concept of probability, normally linked to the concept of "certainty" provided by each data source | Peng & Nagao, 2021; Monkaresi et al., 2017 |
| Filters | Data are fused via the use of filters, generally used for estimating the hidden state of a dynamic system | Li et al., 2020 |
| Others | Non-standard fusion techniques which do not fit in any of the other categories | Luo et al., 2020; N. L. Henderson, Rowe, Mott, Brawner, et al., 2019; Brodny, 2017; J. Chen et al., 2014; Mao et al., 2019; Ma et al., 2015 |

out in this category included Worsley (2014) for the terminology used, which associated the term "naïve" fusion with aggregation (of features in this case); Prieto et al. (2018), for the aggregation of data from wearables, with the peculiarities that went with it, and the various studies by N. L. Henderson, Rowe, Mott, and Lester (2019); N. Henderson et al. (2020) who demonstrated a consistent line of research in this type of educational data fusion.

Ensembles were also an interesting fusion approach, combining the results offered by various classifiers. This category included studies such as Wu et al. (2020), who used the "stacking" method, and Chango, Cerezo, and Romero (2021); Chango, Cerezo, Sanchez-Santillan, et al. (2021) who used the "vote" approach provided by the Weka tool. Details of both approaches may be found in each of the articles cited.

Another common way of combining data was to calculate some statistic from the data which summarized it. Daoudi et al. (2021) calculated the weighted means of the different classifiers constructed in the decision fusion process. N. L. Henderson, Rowe, Mott, and Lester (2019); N. Henderson et al. (2020) calculated the maximum value from the values for predictive reliability from the various classifiers. Qu et al. (2021) used the Spearman coefficient as a measure of correlation between the data sources during the fusion process.

That same study by Qu et al. (2021) appears in the category of fusion methods based on mathematical operators and the category of those based on similarity because, after calculating the Spearman coefficient, the authors used various mathematical formulas to weight the data sources, including some which used the concept of similarity. That means that this study has an advanced, rather than elementary, fusion approach which might open the door for other researchers to not restrict themselves to the classical aggregation of features.

Probabilistic theory is also applicable in data fusion. We found only one of the studies in this category, Peng and Nagao (2021), who chose the classifier in a late fusion process according to the probability distributions of each of the classifiers being correct. It is interesting to note that in this study the authors also spoke of a type of fusion called "single-channel level" fusion, which in our opinion does not actually refer to any kind of fusion as it is based on just the construction of classifiers from each of the data sources. Monkaresi et al. (2017) used individual channel base classifier to make a classification by using the decision of whichever base classifier had the highest decision probability.

In data fusion, filters are one of the most well-known and widely applied approaches as they allow environmental data to be fused in order to predict future states in dynamic systems. Despite that, we only found one study falling within this category (Li et al., 2020), which used the Kalman filter to fuse data of different modalities and the dynamic time warping algorithm to align the data in the same timeline. The idea was that the fused data, coming from sensors placed on an instructor's body during a demonstration of a task, would allow a robot to predict the actions it needed to do to imitate that as faithfully as possible.

Two of the studies used other non-standard fusion techniques. There was a very elaborate fusion strategy in Luo et al. (2020). On the one hand, the features obtained from student interaction data on the platform were weighted according to their entropy and then fused by means of aggregation using that weighting, constructing a first classifier from those features which reflected the students "thinking" aspect. Subsequently, using video recordings, two classifiers were created reflecting "attention" and "emotion" aspects, with those two classifiers being fused at the decision feature level using the analytic hierarchy process (AHP) technique. That fused decision of attention and emotion was then fused with the thinking classifier at the decision level using AHP. This is summarized in Figure 4. N. L. Henderson, Rowe, Mott, Brawner, et al. (2019) used a feature fusion configuration via aggregation of features (both in the most basic version and in the more advanced version which selected the best features) and, the most interesting aspect of the article, the decision fusion configuration used the Match-score fusion technique (Rahman & Gavrilova, 2018).

It was not possible to determine the fusion techniques used in some studies. Despite Brodny (2017) proposing a hybrid fusion technique, it was in broad conceptual terms without any specific fusion techniques being mentioned, it lacked specific test results or tangible results. J. Chen et al. (2014) indicated that their study dealt with decision fusion but did not offer specific details about the fusion approach used. The title of Mao et al. (2019) indicated that there was multi-feature fusion but the article did not give details of the features or the fusion method used. Finally, the study by Ma et al. (2015), from which, as already noted, it was not possible to determine the exact point of fusion, also failed to provide information to identify the data fusion technique used.

## 3.3 | EDM/LA objective

We have classified the selected papers based on the EDM/LA objective or education application that want to resolve the paper that use multimodal data fusion. There are a wide range of popular application or objectives in LA/EDM

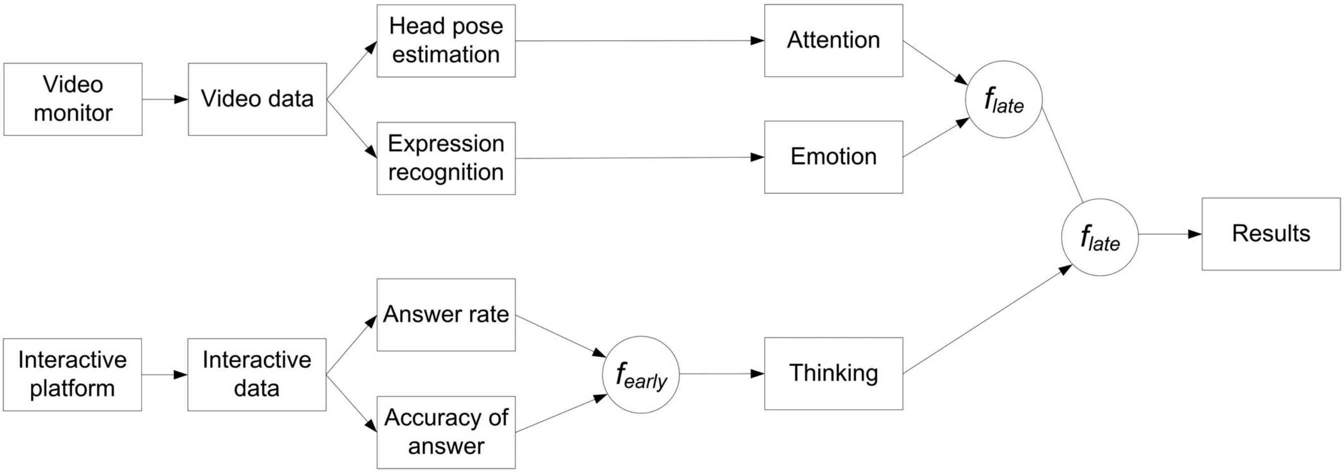


FIGURE 4 Example of the advanced fusion approach used in Luo et al. (2020)

TABLE 6 Categorization of papers according to EDM/LA objective

| EDM/LA objective | Explanation | Studies |
|---|---|---|
| Analysis of students' learning processes | To analyze the student's behavior and style during learning and discovering patterns | Andrade et al., 2016; Liu et al., 2019; Ma et al., 2015; Qu et al., 2021; Shankar et al., 2019 |
| Prediction of students performance | To infer the students final performance/mark/grade variable from some combination of other variables | Chango, Cerezo, & Romero, 2021; Chango, Cerezo, Sanchez-Santillan, et al., 2021; Giannakos et al., 2019; Nam Liao et al., 2019; Olsen et al., 2020; Di Mitri et al., 2017 |
| Students' emotional state evaluation/recognition | To study affect during learning and the importance of students' emotions to learning | Bahreini et al., 2016; Brodny, 2017; J. Chen et al., 2014; Daoudi et al., 2021; N. L. Henderson, Rowe, Mott, & Lester, 2019; N. L. Henderson, Rowe, Mott, Brawner, et al., 2019; N. Henderson et al., 2020; Mao et al., 2019; Peng & Nagao, 2021; Hussain et al., 2011 |
| Prediction of students' engagement | To predict students' engagement, motivation, interest, and so forth | Gadaley et al., 2020; Luo et al., 2020; Yue et al., 2019; Monkaresi et al., 2017; Sharma et al., 2019 |
| Modeling teacher behavior | To analyze teacher behavior during instruction and interaction with students | Prieto et al., 2018; Wu et al., 2020 |
| Teacher discourse classification | To analyze instructors' text data from forums, chats, social networks, and so forth | Xu et al., 2019 |
| Others | Other objectives or applications | Li et al., 2020; Worsley, 2014 |

(Romero & Ventura, 2013) (Romero & Ventura, 2020) for solving educational problems or goals. Table 6 shows the categorization of the papers based on commonly sought objectives in the areas of EDM and LA. There were also other different objectives and some papers which did not specifically indicate the EDM/LA objective (Others in the table).

There was a group of studies in which the objective was to conduct analyses of students' learning processes. Liu et al. (2019) sought to understand student learning processes, incorporating insights from data collected in multiple modalities and contexts. Ma et al. (2015) aimed to analyze the learning process in a smart classroom. Qu et al. (2021) evaluated college students' learning behavior, providing a basis for adaptive learning environments. The objective for Shankar et al. (2019) was to better understand the learning process considering the contextual information of the situation. Andrade et al. (2016) modeled student behavior in order to identify whether clusters of observable behaviors could be used to identify and characterize behavioral frames in rich video data of student interviews.

Other studies sought to predict students' final performance, such as López Zambrano et al. (2021). Chango, Cerezo, and Romero (2021); Chango, Cerezo, Sanchez-Santillan, et al. (2021) predicted the final academic performance of university students in a blended learning environment. In contrast, Giannakos et al. (2019) sought to accurately predict

TABLE 7 Relationship between fusion technique and EDM/LA objective

| Technique objective | Aggregation | Ensembles | Statistical operators | Math operators | Similarity-based | Probability | Filters | Others |
|---|---|---|---|---|---|---|---|---|
| Analysis of students' learning processes | Andrade et al., 2016; Liu et al., 2019; Shankar et al., 2019 | | Qu et al., 2021 | Qu et al., 2021 | Qu et al., 2021 | | | Ma et al., 2015 |
| Prediction of students performance | Chango, Cerezo, & Romero, 2021; Chango, Cerezo, Sanchez-Santillan, et al., 2021; Giannakos et al., 2019; Nam Liao et al., 2019; Olsen et al., 2020; Di Mitri et al., 2017 | Chango, Cerezo, & Romero, 2021; Chango, Cerezo, Sanchez-Santillan, et al., 2021 | | | | | | |
| Students' emotional state evaluation/ recognition | Bahreini et al., 2016; N. L. Henderson, Rowe, Mott, & Lester, 2019; N. L. Henderson, Rowe, Mott, Brawner, et al., 2019; N. Henderson et al., 2020; Peng & Nagao, 2021; Hussain et al., 2011 | | Daoudi et al., 2021; N. L. Henderson, Rowe, Mott, & Lester, 2019; N. Henderson et al., 2020 | | | Peng & Nagao, 2021 | | N. L. Henderson, Rowe, Mott, Brawner, et al., 2019; Brodny, 2017; J. J. Chen et al., 2014; Mao et al., 2019 |
| Prediction of students' engagement | Gadaley et al., 2020; Yue et al., 2019; Sharma et al., 2019 | | | | | Monkaresi et al., 2017 | | Luo et al., 2020 |
| Modeling teacher behavior | Prieto et al., 2018; Wu et al., 2020 | Wu et al., 2020 | | | | | | |
| Teacher discourse classification | Xu et al., 2019 | | | | | | | |
| Others | Worsley, 2014 | | | | | | Li et al., 2020 | |

users' acquisition of skills, commonly called movement-motor learning. The aim in Olsen et al. (2020) was to predict students' collaborative learning gains, while in Nam Liao et al. (2019), it was to arrive at early predictions of students' overall performance on a course. Di Mitri et al. (2017) uses a machine learning approach for predicting performance in self-regulated learning. Sharma et al. (2019).

There was a group of studies in which fusion was used to try and improve the evaluation of emotions during the learning process. The main objectives sought in this regard were: the evaluation of learners' affective states in collaborative serious games (Daoudi et al., 2021); emotion recognition and integration of emotional states in educational applications with consideration of uncertainty (Brodny, 2017); recognizing student mental states in conversations (Peng & Nagao, 2021); emotion recognition (frustration) in game-based learning (N. L. Henderson, Rowe, Mott, Brawner, et al., 2019); recognition of students' affective states (J. Chen et al., 2014); recognition of students' facial micro expressions (Mao et al., 2019); real-time, continuous, unobtrusive emotion recognition (Bahreini et al., 2016); improved detection of affect (N. Henderson et al., 2020); detection of learner affect in game-based learning (N. L. Henderson, Rowe, Mott, & Lester, 2019); and (Hussain et al., 2011) detect learners' affective states in ITS.

In other studies, fusion was used to improve predictions of student engagement or interest. Gadaley et al. (2020) predicted engagement in classes where students were allowed to have digital devices during lectures. Luo et al. (2020) modeled student interest using multimodal natural sensing technology in order to provide an effective basis for improving teaching in real time. The aim in Yue et al. (2019) was to detect learners' emotional and eye-based behavioral engagement in real-time as well as to predict learners' performance after completing a short video course. Monkaresi et al. (2017) explored how computer vision techniques can be used to detect engagement while students completed a structured writing activity (draft-feedback-review). Sharma et al. (2019) proposed stimuli-based gaze variables as student's attention indicators (i.e., with-me-ness) in order to enhance motivation and learning in MOOCs.

One group of studies was notable because they modeled teacher behavior, with fusion being used to assist that modeling. The aim in Wu et al. (2020) was to recognize teacher behavior in order to solve problems of time-consumption and information overload in teaching and then help teachers optimize teaching strategies and improve teaching efficiency. The objective in Prieto et al. (2018) was pedagogical modeling of a teacher in class in order to provide automated tagging of classroom practice that could be used in everyday practice with multiple teachers. Finally, there was one study in which fusion helped to classify teacher speech (Xu et al., 2019). The aim was to automatically classify teacher discourse in a Chinese classroom. One intriguing objective (in the "others" category) was the automatic reproduction of learned tasks by a robot following a teachers' demonstrations of certain physical tasks (Li et al., 2020). Lastly, there was one study in which the author indicated that they did LA, but did not specify any LA objectives, simply discussed how decisions about data fusion have a significant impact on how the research relates to learning theories (Worsley, 2014).

We have also conducted an analysis to reveal the relationship between the data fusion technique used in the different works and the EDM/LA objectives achieved in each of those work (see Table 7).

If we analyze the above table from the perspective of data fusion techniques, we can see that aggregation techniques cover all the range of EDM/LA objectives, having most works located on the categories of analysis of students' learning process, prediction of performance, emotional recognition, and engagement prediction, which shows the wide applicability of aggregation-based fusion approaches. From the perspective of EDM/LA objective, we have noticed that analysis

**TABLE 8** Publicly available EDM/LA multimodal datasets

| Name | URL |
|---|---|
| Dataset of Multimodal Interface for Solving Equations | https://pslcdatashop.web.cmu.edu/Project?id=33 |
| MUTLA: A Large-Scale Dataset for Multimodal Teaching and Learning Analytics | https://paperswithcode.com/dataset/mutla |
| NUS Multi-Sensor Presentation (NUSMSP) Dataset | https://scholarbank.nus.edu.sg/handle/10635/137261 |
| PE-HRI: A Multimodal Dataset for the study of Productive Engagement in a robot mediated Collaborative Educational Setting | https://zenodo.org/record/4288833#.Yd4OO_7MKUk |
| Student Life Dataset | https://studentlife.cs.dartmouth.edu/ |
| VLEngagement: A Dataset of Scientific Video Lectures for Evaluating Population-based Engagement | https://github.com/sahanbull/context-agnostic-engagement |

of students' performance and emotional state recognition are two objectives covered by a large number of different fusion approaches (five and four respectively), having found that aggregation approach is the most used when trying to achieve those two objectives. When it comes to prediction of students' performance, aggregation and ensembles are the two only approaches employed, which indicates that both seems to be the reference standard approaches for achieving those two objectives. Focusing on engagement prediction and students' behavior modeling, aggregation-based fusion also is the most employed approach. Finally, emotional state recognition is an objective for which the use of aggregation, statistical operators and other non-standard approaches have demonstrated to obtain positive results.

## 4 | CONCLUSIONS AND FUTURE TRENDS

Data fusion of multimodal data seems promising in the field of Education in general (Sultana et al., 2020), and particularly in the field of EDM/LA (Mu et al., 2020), as we have shown in this review. We have analyzed all the related papers in the bibliography from the perspective of the data being fused, the fusion approaches used, and the EDM/LA educational objective or application and we have obtained the next conclusions:

- In terms of the data fused, there was a relatively balanced use in the different educational environments, with data fusion being found 12 papers focused on in-person learning, 11 on online learning, and 8 on hybrid and blended environments. Most of the data being fused are focused specifically on learners, while only a minority focused on teacher data. The data came from a wide range of sources, mainly recordings of students, sensor readings of various aspects, and numerical data indicating some magnitude generally related to academic performance. Almost all of the data were physical or digital, a minority were physiological.
- In terms of the fusion approach, the majority of the papers used early fusion of features, while a large number used late fusion or decisions produced by different classifiers in previous stages. Very few studies used hybrids of those two approaches and even fewer went outside this framework (early-late-hybrid) summarized in Figure 2. Looking at the fusion techniques used, aggregation of features is the predominant method, followed by others based on the use of statistical operators and ensembles.
- In terms of EDM/LA objectives or educational application/problem in papers that used data fusion, the most notable were those seeking to manage student emotions, analyze student behavior, and those that aimed to predict academic performance, interest, or engagement.

It is important to notice that only one paper (Olsen et al., 2020) used a free available public multimodal dataset from Pittsburgh Science of Learning Center (PSLC) DataShop. All the other selected papers used their own private datasets. However, there are an increasing number of Publicly Available EDM/LA Datasets (Mihaescu & Popescu, 2021) and Table 8 shows a list of specific multimodal datasets/repositories that could be used for researching on EDM/LA data fusion.

Finally, after doing this review of the literature about data fusion of multimodal data, we identify the next opportunities or challenges for future research in this area:

- Most of the data sources examined were from students, with only a few studies focusing solely on teachers. It would be interesting to combine both teacher and student data in the same study in order to determine whether student behavior could be influenced by teacher characteristics, or whether the teacher adapts their methodology based on the type of student they are teaching, such as in the framework of the classical theories from Biggs (1987) about student and teacher approaches to learning.
- We found only some works about fusion of psychometric/environmental data in the processes we examined. It would be interesting to see more works using this type of data to, for example, be able to determine whether student psychological processes are affected in any way by the nature of the environment in which they are learning (temperature, humidity, lighting, etc.).
- Most fusion techniques used in EDM/LA are basic and fundamental, and the most widely used are simple aggregation, ensembles, and statistical operators. It is clear that is restricted to early-late-hybrid fusion schema. However, data fusion is such a rich, versatile field that much potential is lost by that restriction. Some studies proposed using other, more flexible processes, which produced good results in some studies and should be explored (Li et al., 2020; Qu et al., 2021; Worsley, 2014). Other advanced approaches that allow improved fusion in different fields are: the use

of techniques based on filters, probabilistic approaches, possibilistic approaches, and the use of the Demptster–Shafer theory of evidence seem useful for that and have been little used for educational data fusion.

- Data fusion of multimodal data has been used mainly in some EDM/LA applications such as prediction of performance and engagement, analysis of student behavior, and the management of student emotions. However, there are much other EDM/LA objectives, applications or educational problems that have not been addressed by using data fusion such as classroom planning, learning strategy recommendations, and course construction and organization, and so forth.
- It is also important to mention the connection between multimodal data fusion to the knowledge management (KM) area and incorporate its large experience and vast literature data management process. In this line, it would be interesting to see works showing for example: how data fusion would be benefitted in cloud-based knowledge management frameworks in higher education institutions (Noor et al., 2019) or in visualizing educational information in the context of modern education megatrend (Izotova et al., 2021).

## ACKNOWLEDGMENTS

This work would not have been possible without the funding from the Ministry of Sciences and Innovation I + D + I PID2020-115832GB-I00 and PID2019-107201GB-I00 and funding for open access charge: Universidad de Córdoba/CBUA.

## CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

## AUTHOR CONTRIBUTIONS

**Wilson Chango:** Visualization (lead); writing – original draft (lead); writing – review and editing (lead). **Juan A. Lara:** Resources (lead); software (lead); supervision (lead); validation (lead). **Rebeca Cerezo:** Data curation (lead); formal analysis (lead); funding acquisition (lead). **Cristóbal Romero Morales:** Conceptualization (lead); methodology (lead); project administration (lead).

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## ORCID

*Cristóbal Romero* https://orcid.org/0000-0003-4180-4948

## REFERENCES

Andrade, A., Delandshere, G., & Danish, J. A. (2016). Using multimodal learning analytics to model student behavior: A systematic analysis of epistemological framing. *Journal of Learning Analytics*, *3*(2), 282–306. https://doi.org/10.18608/jla.2016.32.14

Bahreini, K., Nadolski, R., & Westera, W. (2016). Data fusion for real-time multimodal emotion recognition through webcams and microphones in e-learning. *International Journal of Human–Computer Interaction*, *32*(5), 415–430. https://doi.org/10.1080/10447318.2016.1159799

Biggs, J. (1987). *Student approaches to learning and studying*. Australian Council for Educational Research.

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, *3*(2), 220–238. https://doi.org/10.18608/jla.2016.32.11

Brodny, G. (2017). *Uncertainty and integration of emotional states in e-learning*. https://www.scitepress.org/Papers/2017/64022/64022.pdf

Budaher, J., Brun, A., & Boyer, A. (2020). *Multi-source data mining for e-learning*. arXiv preprint arXiv:2009.08791.

Chango, W., Cerezo, R., & Romero, C. (2021). Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses. *Computers & Electrical Engineering*, *89*, 106908. https://doi.org/10.1016/J.COMPELECENG.2020.106908

Chango, W., Cerezo, R., Sanchez-Santillan, M., Azevedo, R., & Romero, C. (2021). Improving prediction of students' performance in intelligent tutoring systems using attribute selection and ensembles of different multimodal data sources. *Journal of Computing in Higher Education*, *33*, 614–634. https://doi.org/10.1007/s12528-021-09298-8

Chen, J., Luo, N., Liu, Y., Liu, L., Zhang, K., & Kolodziej, J. (2014). A hybrid intelligence-aided approach to affect-sensitive e-learning. *Computing*, *98*(1), 215–233. https://doi.org/10.1007/S00607-014-0430-9

Chen, X., Zou, D., Xie, H., & Wang, F. L. (2021). Past, present, and future of smart learning: A topic-based bibliometric analysis. *International Journal of Educational Technology in Higher Education*, *18*(1), 1–29. https://doi.org/10.1186/S41239-020-00239-6

Daoudi, I., Chebil, R., Tranvouez, E., Lejouad Chaari, W., & Espinasse, B. (2021). Improving learners' assessment and evaluation in crisis management serious games: An emotion-based educational data mining approach. *Entertainment Computing*, *38*, 100428. https://doi.org/10.1016/J.ENTCOM.2021.100428

Dewan, M. A. A., Murshed, M., & Lin, F. (2019). Engagement detection in online learning: A review. *Smart Learning Environments*, *6*(1), 1–20. https://doi.org/10.1186/S40561-018-0080-Z

Di Mitri, D., Scheffel, M., Drachsler, H., Börner, D., Ternier, S., & Specht, M. (2017). Learning pulse: A machine learning approach for predicting performance in self-regulated learning using multimodal data. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 188–197). ACM Digital Library.

Di Mitri, D., Schneider, J., Klemke, R., Specht, M., & Drachsler, H. (2019). Read between the lines: An annotation tool for multimodal data for learning. *ACM International Conference Proceeding Series*, *10*(LAK19), 51–60. https://doi.org/10.1145/3303772.3303776

Ding, W., Jing, X., Yan, Z., & Yang, L. T. (2019). A survey on data fusion in internet of things: Towards secure and privacy-preserving fusion. *Information Fusion*, *51*, 129–144. https://doi.org/10.1016/J.INFFUS.2018.12.001

Gadaley, R. S., Duraisamy, P., Van Haneghan, J., & Jackson, S. (2020). Classroom engagement evaluation using multi-sensor fusion with a 180° camera view. *Communications in Computer and Information Science*, *1319*, 123–131. https://doi.org/10.1007/978-981-33-4601-7_13

Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, *48*, 108–119. https://doi.org/10.1016/J.IJINFOMGT.2019.02.003

Han, Z., Wu, J., Huang, C., Huang, Q., & Zhao, M. (2020). A review on sentiment discovery and analysis of educational big-data. *WIREs Data Mining and Knowledge Discovery*, *10*(1), e1328. https://doi.org/10.1002/WIDM.1328

Henderson, N., Rowe, J., Paquette, L., Baker, R. S., & Lester, J. (2020). Improving affect detection in game-based learning with multimodal data fusion. *Lecture Notes in Computer Science*, *12163*, 228–239. https://doi.org/10.1007/978-3-030-52237-7_19

Henderson, N. L., Rowe, J. P., Mott, B. W., Brawner, K., Baker, R., & Lester, J. C. (2019). 4D affect detection: Improving frustration detection in game-based learning with posture-based temporal data fusion. *Lecture Notes in Computer Science*, *11625*, 144–156. https://doi.org/10.1007/978-3-030-23204-7_13

Henderson, N. L., Rowe, J. P., Mott, B. W., & Lester, J. C. (2019). Sensor-based data fusion for multimodal affect detection in game-based learning environments. In *Paper presented at 12th international conference on educational data mining 2019—Workshop on EDM & games: Leveling up engaged learning with data-rich analytics* (pp. 44–50). CEUR Workshop Proceedings.

Hussain, M. S., AlZoubi, O., Calvo, R. A., & D'Mello, S. K. (2011). Affect detection from multichannel physiology during learning sessions with AutoTutor. In *International conference on artificial intelligence in education* (pp. 131–138). Springer. https://doi.org/10.1007/978-3-642-21869-9_19

Izotova, N., Klimenko, M., & Nikolaenko, E. (2021). Information visualization in context of modern education megatrends. In *E3S web of conferences* (Vol. 284, p. 09011). EDP Sciences. https://doi.org/10.1051/e3sconf/202128409011

Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2021). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction*, *72*, 101203. https://doi.org/10.1016/J.LEARNINSTRUC.2019.04.004

Kaur, R., & Kautish, S. (2019). Multimodal sentiment analysis: A survey and comparison. *International Journal of Service Science, Management, Engineering, and Technology (IJSSMET)*, *10*(2), 38–58. https://doi.org/10.4018/IJSSMET.2019040103

Lahat, D., Adali, T., & Jutten, C. (2015). Multimodal data fusion: An overview of methods, challenges, and prospects. *Proceedings of the IEEE*, *103*(9), 1449–1477. https://doi.org/10.1109/JPROC.2015.2460697

Li, J., Zhong, J., Yang, J., & Yang, C. (2020). An incremental learning framework to enhance teaching by demonstration based on multimodal sensor fusion. *Frontiers in Neurorobotics*, *14*, 1–16. https://doi.org/10.3389/FNBOT.2020.00055

Liu, R., Stamper, J., Davenport, J., Crossley, S., McNamara, D., Nzinga, K., & Sherin, B. (2019). Learning linkages: Integrating data streams of multiple modalities and timescales. *Journal of Computer Assisted Learning*, *35*(1), 99–109. https://doi.org/10.1111/JCAL.12315

López Zambrano, J., Lara Torralbo, J. A., & Romero Morales, C. (2021). Early prediction of student learning performance through data mining: A systematic review. *Psicothema*, *33*(3), 456–465. https://doi.org/10.7334/psicothema2021.62

Luo, Z., Jingying, C., Guangshuai, W., & Mengyi, L. (2020). A three-dimensional model of student interest during learning using multimodal fusion with natural sensing technology. *Interactive Learning Environments*, *1*, 1–14. https://doi.org/10.1080/10494820.2019.1710852

Ma, H., Wang, X., & Liang, D. (2015). Research on analysis of multi-source data fusion model in Smart classroom. *Journal of Information Technology and Application in Education*, *4*, 82. https://doi.org/10.14355/jitae.2015.04.013

Mao, L., Wang, N., Wang, L., & Chen, Y. (2019). Classroom micro-expression recognition algorithms based on multi-feature fusion. *IEEE Access*, *7*, 64978–64983. https://doi.org/10.1109/ACCESS.2019.2917230

Mihaescu, M. C., & Popescu, P. S. (2021). Review on publicly available datasets for educational data mining. *WIREs Data Mining and Knowledge Discovery*, *11*(3), e1403. https://doi.org/10.1002/widm.1403

Monkaresi, H., Bosch, N., Calvo, R. A., & D'Mello, S. K. (2017). Automated detection of engagement using video-based estimation of facial expressions and heart rate. *IEEE Transactions on Affective Computing*, *8*(1), 15–28. https://doi.org/10.1109/TAFFC.2016.2515084

Mu, S., Cui, M., & Huang, X. (2020). Multimodal data fusion in learning analytics: A systematic review. *Sensors*, *20*(23), 6856. https://doi.org/10.3390/S20236856

Nam Liao, S., Zingaro, D., Alvarado, C., Griswold, W. G., Porter, L., & Gris-wold, W. G. (2019). Exploring the value of different data sources for predicting student performance in multiple CS courses. In *Proceedings of the 50th ACM technical symposium on computer science education*. ACM Digital Library. https://doi.org/10.1145/3287324

Nandi, A., Xhafa, F., Subirats, L., & Fort, S. (2020). A survey on multimodal data stream mining for e-Learner's emotion recognition. *2020 international conference on omni-layer intelligent systems, COINS 2020* (pp. 1–6). IEEE Xplore. https://doi.org/10.1109/COINS49042.2020.9191370

Noor, A. S. M., Younas, M., & Arshad, M. (2019). A review on cloud based knowledge management in higher education institutions. *International Journal of Electrical & Computer Engineering*, *9*(6), 5420–5427. http://doi.org/10.11591/ijece.v9i6.pp5420-5427

Ochoa, X. (2017). Multimodal learning analytics. In C. Lang, G. Siemens, A. F. Wise, & D. Gašević (Eds.), *The handbook of learning analytics* (pp. 129–141). Society for Learning Analytics Research (SoLAR). https://doi.org/10.18608/HLA17.011

Olsen, J. K., Sharma, K., Rummel, N., & Aleven, V. (2020). Temporal analysis of multimodal data to predict collaborative learning outcomes. *British Journal of Educational Technology*, *51*(5), 1527–1547. https://doi.org/10.1111/BJET.12982

Peng, S., & Nagao, K. (2021). Recognition of students' mental states in discussion based on multimodal data and its application to educational support. *IEEE Access*, *9*, 18235–18250. https://doi.org/10.1109/ACCESS.2021.3054176

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, *37*, 98–125. https://doi.org/10.1016/J.INFFUS.2017.02.003

Prieto, L. P., Sharma, K., Kidzinski, Ł., Rodríguez-Triana, M. J., & Dillenbourg, P. (2018). Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of Computer Assisted Learning*, *34*(2), 193–203. https://doi.org/10.1111/JCAL.12232

Qu, J., Liu, A., & Liu, R. (2021, 2021). Research on evaluation and confirmation of college students' learning behavior based on comprehensive weighted fusion algorithm. In *2021 IEEE 6th international conference on cloud computing and big data analytics, ICCCBDA* (pp. 121–125). IEEE Xplore. https://doi.org/10.1109/ICCCBDA51879.2021.9442584

Raes, A. (2022). Exploring student and teacher experiences in hybrid learning environments: Does presence matter? *Postdigital Science and Education*, *4*, 138–159. https://doi.org/10.1007/s42438-021-00274-0

Rahman, M. W., & Gavrilova, M. L. (2018). Emerging EEG and kinect face fusion for biometric identification. In *2017 IEEE symposium series on computational intelligence, SSCI 2017—Proceedings, 2018-January, 1–8*. IEEE Xplore. https://doi.org/10.1109/SSCI.2017.8285231

Romero, C., & Ventura, S. (2013). Data mining in education. *WIREs Data Mining and Knowledge Discovery*, *3*(1), 12–27. https://doi.org/10.1002/widm.1075

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, *10*(3), e1355. https://doi.org/10.1002/WIDM.1355

Sánchez Ruiz, L. M., Moll-López, S., Moraño-Fernández, J. A., & Llobregat-Gómez, N. (2021). B-learning and technology: Enablers for university education resilience. An experience case under COVID-19 in Spain. *Sustainability*, *13*(6), 3532.

Shankar, S. K., Prieto, L. P., Rodriguez-Triana, M. J., & Ruiz-Calleja, A. (2018). A review of multimodal learning analytics architectures. In *Proceedings—IEEE 18th international conference on advanced learning technologies, ICALT 2018* (pp. 212–214). IEEE Xplore. https://doi.org/10.1109/ICALT.2018.00057

Shankar, S. K., Ruiz-Calleja, A., Prieto, L. P., Rodríguez-Triana, M. J., & Chejara, P. (2019). An architecture and data model to process multimodal evidence of learning. *Lecture Notes in Computer Science*, *11841*, 72–83. https://doi.org/10.1007/978-3-030-35758-0_7

Sharma, K., Dillenbourg, P., & Giannakos, M. (2019). Stimuli-based gaze analytics to enhance motivation and learning in MOOCs. In *Proceedings of the 2019 IEEE 19th international conference on advanced learning technologies (ICALT), Macei, Brazil, 15–18 July 2019* (pp. 199–203). IEEE Xplore. https://doi.org/10.1109/ICALT.2019.00052

Sharma, K., & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology*, *51*(5), 1450–1484. https://doi.org/10.1111/BJET.12993

Sultana, M., Jibon, F. A., & Kashem, A. (2020). An improved model of virtual classroom using information fusion and NS-DBSCAN. *Global Journal of Computer Science and Technology*, *20*(5), 26–33. https://computerresearch.org/index.php/computer/article/view/1966

Tabuenca, B., Serrano-Iglesias, S., Martin, A. C., Villa-Torrano, C., Dimitriadis, Y., Asensio-Perez, I. J., & Kloos, C. D. (2021). Affordances and core functions of smart learning environments: A systematic literature review. *IEEE Transactions on Learning Technologies*, *14*(2), 129–145. https://doi.org/10.1109/TLT.2021.3067946

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, *14*, 207–222. https://doi.org/10.1111/1467-8551.00375

Wang, R., Ji, W., Liu, M., Wang, X., Weng, J., Deng, S., Gao, S., & Yuan, C. (2018). Review on mining data from multiple data sources. *Pattern Recognition Letters*, *109*, 120–128. https://doi.org/10.1016/j.patrec.2018.01.013

Worsley, M. (2014). Multimodal learning analytics as a tool for bridging learning theory and complex learning behaviors. In *MLA 2014—Proceedings of the 2014 ACM multimodal learning analytics workshop and grand challenge, co-located with ICMI* (Vol. 2014, pp. 1–4). ACM Digital Library. https://doi.org/10.1145/2666633.2666634

Wu, D., Chen, J., Deng, W., Wei, Y., Luo, H., & Wei, Y. (2020). The recognition of teacher behavior based on multimodal information fusion. *Mathematical Problems in Engineering*, *10*(13), 1–8. https://doi.org/10.1155/2020/8269683

Xu, L., He, X., Zhang, J., & Li, Y. (2019). Automatic classification of discourse in Chinese classroom based on multi-feature fusion. In *CITS 2019—Proceeding of the 2019 international conference on computer, information and telecommunication systems* (pp. 1–5). IEEE Xplore. https://doi.org/10.1109/CITS.2019.8862101

Yue, J., Tian, F., Chao, K. M., Shah, N., Li, L., Chen, Y., & Zheng, Q. (2019). Recognizing multidimensional engagement of E-learners based on multi-channel data in E-learning environment. *IEEE Access*, *7*, 149554–149567. https://doi.org/10.1109/ACCESS.2019.2947091

**How to cite this article:** Chango, W., Lara, J. A., Cerezo, R., & Romero, C. (2022). A review on data fusion in multimodal learning analytics and educational data mining. *WIREs Data Mining and Knowledge Discovery*, e1458. https://doi.org/10.1002/widm.1458

Computers and Electrical Engineering 89 (2021) 106908

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

# Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses

Wilson Chango [a], Rebeca Cerezo [b], Cristóbal Romero [c,*]

[a] Pontifical Catholic University of Ecuador, Department of Systems and Computing, Ecuador
[b] University of Oviedo - Spain, Department of Psychology, Spain
[c] University of Córdoba – Spain, Department of Computer Science, Spain

ARTICLE INFO

*Keywords:*
Blended learning
Predicting academic performance
Multisource data
Multimodal learning
Data fusion

ABSTRACT

In this paper we apply data fusion approaches for predicting the final academic performance of university students using multiple-source, multimodal data from blended learning environments. We collect and preprocess data about first-year university students from different sources: theory classes, practical sessions, on-line Moodle sessions, and a final exam. Our objective is to discover which data fusion approach produces the best results using our data. We carry out experiments by applying four different data fusion approaches and six classification algorithms. The results show that the best predictions are produced using ensembles and selecting the best attributes approach with discretized data. The best prediction models show us that the level of attention in theory classes, scores in Moodle quizzes, and the level of activity in Moodle forums are the best set of attributes for predicting students' final performance in our courses.

## 1. Introduction

Blended learning (b-learning) is a method of teaching approach that combines online learning with traditional in face-to-face classroom methods. In research literature [1]. the terms blended learning, hybrid learning and mixed-mode instruction are often used to refer b-learning. Its main goal is to overcome the drawbacks of pure online learning and it remains a priority issue in technology enhaced education despite having been put into practice on all over the word for 20 years ago. Nowadays, in the current pandemic scenario, blended instruction has become more important since it is going to be the new normal in terms of teaching-learning in higher education. COVID-19 has led to the sudden suspension of teaching activities in many countries and the scramble to find new ways to resume classes with restrictive space and hygiene requirements. Many institutions do not have enough space to implement the necessary measures and are forced to rethink their face-to-face learning plans as blended learning plans.

Rapid advances in technology have let us capture all student actions in their interactions with virtual and traditional learning environments. Blended learning environments gather a huge amount of data about students' multimodal interactions in traditional classrooms and on-line environments from a wide range of data sources [2]. So, these data sources need to be fused and mined to shed light on educational issues such as prediction of student performance. In this line, Educational data mining (EDM) [3] can be applied to discover and improve educational processes from information extracted from educational data, which is then used to understand the educational process [4]. EDM has been widely used to improve and enhance learning quality, as well as in the pursuit of research objective to understand the teaching-learning process [5]. In this line, one of the most frequent and the oldest studied tasks/problems

* Corresponding author.
*E-mail address:* cromero@uco.es (C. Romero).

https://doi.org/10.1016/j.compeleceng.2020.106908
Received 11 February 2020; Received in revised form 5 October 2020; Accepted 6 November 2020
Available online 20 November 2020
0045-7906/

in EDM is the prediction of learners' performance. It is still a challenge to predict student learning achievement in blended learning environments combining online and offline learning [6] making data fusion techniques necessary.

In this study we do a classification task for predicting the value of a categorical/nominal attribute (the class or final academic status of the student: Pass, Fail or Drop out) based on other attributes (the predictive attributes from various available data sources). We propose applying different data fusion approaches and classification algorithms to data gathered from several sources (theory classes, practical sessions, online sessions and final exams) in a blended university course in order to predict the students' final academic performance. The research questions posed by this study are:

Question 1.- Which data fusion approach and classification algorithms produce the best results from our data?
Question 2.- How useful are the prediction models we produce to help teachers detect students who are at risk of drop out or fail the course?

This paper is organized as follows. The first section covers the background of the related research areas. Following that, we describe the proposed methodology, and describe the data used and how it was preprocessed. Next, we describe the experiments we performed and the results they produced. Finally, we discuss the implications, conclusions and future research.

## 2. Background

Multimodal Learning Analytics (MLA) is a subfield of Learning Analytics (LA) that uses data from different sources about learning traces for doing a single analysis. MLA is much related with multi-view, multi-relational data, and data fusion. It is used for understanding and optimizing teaching-learning process in which the use of videos has now been consolidated, from traditional courses to mixed and online courses [7]. It has become increasingly broadly applied in both online and in face-to-face learning environments where interactions are not solely mediated by digital devices [8]. MLA uses log-files and gaze data, biosensors, interactions with videos, audio and digital documents, and any other data source for understanding or measuring the learning process. So, one important issue to resolve is how to combine, or fuse, the data extracted from several sources/modalities in order to provide a better and more comprehensive view of teaching-learning processes [9].

Data/Information fusion is the process of efficiently transforming and integrating information gathered from various sources at various times, either automatically or semi-automatically, into a form that can provide practical support to a decision-making process, be that human or automatic. Data fusion is used for reducing the dimension of size of the data, optimizing how much data/info there is, and extracting information that is useful [10]. Multimodal data fusion is the combination/integration of data from different/several sources/modalities/contexts in order to obtaining a better understanding of the teaching-learning process [11]. There are three main types of multimodal fusion approaches [12]:

- **Naïve fusion** is the simplest approach. It builds several classifiers using features summary statistics obtained from each of the different data sources/streams.
- **Low-level fusion frame (or feature fusion)** merges raw data. It synchronizes the data sources/streams at each time stage and it analyses the features after their integration together.
- **High-level fusion frame (or quasi feature-level)** uses semantic analysis first to attempt to make sense of the raw data. It extracts one or several abstract or high level features starting from one or more data sources/streams before integrating them.

A different way to group/classify fusion methods is by considering the fusion periods/steps. In terms of period-level fusions, to date there are the next subtypes of MLA fusion [13]:

- **Feature-level or early fusion.** This happens at the first steps of multimodal data fusion in which there is concatenation and no overlapping. So, the obtained feature/attribute vectors are heterogeneous due to concatenate different data sources/streams.
- **Decision-level or later fusion.** In contrast to feature-level fusion, decision-level fusion is conducted at a later step. It allows each data source/stream to use the most appropriate classifier for its features. The drawback is that it can be hard and time consuming to fuse different classifiers at this step.
- **Hybrid fusion**. This type of fusion propose to use in a hybrid way both feature-level and decision-level fusions.

Finally, most data fusion schemes have four stages; preprocessing the data set, shrinking the dimensions of the data and using data correlation to identify the most fruitful feature sets, training classification algorithms, and finally forecasting new data based on classification algorithms. Feature selection algorithms are normally used in data fusion for classification problems in order to reduce dimensions of data and to produce the best results [14].

In this study we apply several multimodal fusion approaches based on Naïve and decision-level fusion. To our knowledge, there are no studies examining how data fusion approaches can help for predicting students' performance in blended learning.

## 3. Proposal

This paper proposes to use a data fusion and mining methodology for predicting students' final performance starting from multi-source and multimodal data (see Fig. 1).

There are two main stages in our methodology as we can see Fig. 1.

- First stage: It gathers data from several sources: theory classes, practical sessions, online sessions with Moodle, and the course final exam. It also applies some pre-process tasks (anonymization; attribute normalization and discretization; and format transformation) for generating datasets in two formats: numerical and categorical.
- Second stage: It uses different data fusion approaches (merging all attributes; selecting the best attributes; using ensembles; and using ensembles and selecting the best attributes) and several white-box classification algorithms with the datasets. Then, we compare the predictions produced by the models in order to discover the best approach and classification model so that it is used for predicting students' final performance.

## 4. Data

We used information from 57 electrical engineering first-year students from University of Cordoba (UCO-Spain) in the Introduction to Computer Science course during the first semester of academic year 2017–2018. The main contents of the course were: History of Computer Science; Introduction to Operating Systems, Databases, Internet and Office Applications; and Introduction to Programming.

### *4.1. Gathering data*

We have gathered all the information from four data sources of the same course: theory classes, practical classes, on-line sessions and final exam. The first three data sources gave us the input attributes and the final exam, the output attribute or class to predict. In this course there was only one group for theory classes and two groups for practical classes. A single teacher collected all the data and video recorded the theory classes because the same teacher was assigned to all the groups in this course. The students all gave their written consent to being recorded, after being informed about the study, and to have their data from practical and online sessions in Moodle collected for the study.

#### *4.1.1. Theory classes*

Theory classes are the traditional face-to-face sessions in which the teacher teaches the theoretical content of the course using blackboards/whiteboards/projectors.

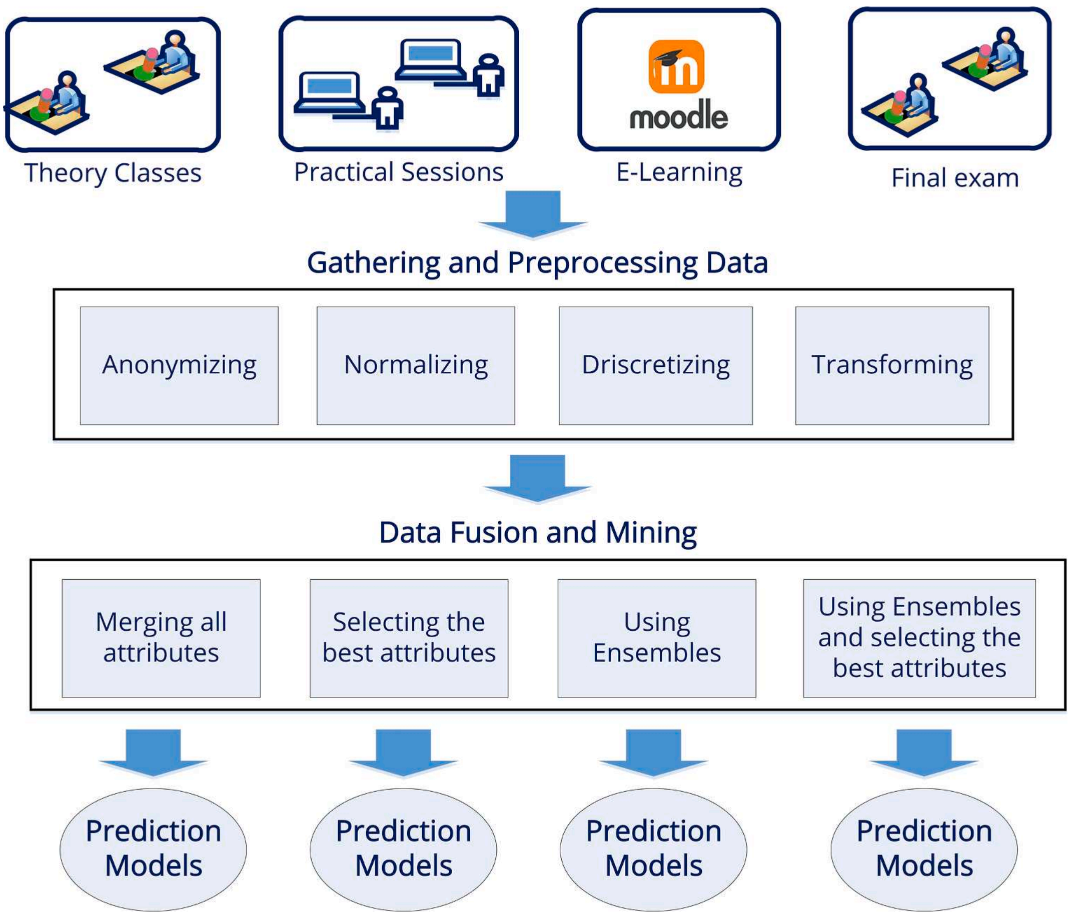


**Fig. 1.** Proposed data fusion and mining methodology for predicting students' performance from multiple data sources.

We collected the following information by extracting it from videos (see Fig. 2) recorded during the 15 in person theory classes given by the course teacher:

- Theory.Attendance: This was gathered manually from the videos. The value was 1 for a student attending a session, and 0 for a student not attending a session.
- Theory.Location: This was gathered manually from the videos. This value was which row the student sat in (from 1 to 12 rows of chairs or 0 if the student did not attend) in the classroom.
- Theory.Attention: This was gathered semi-automatically from the videos. This value measured how much time the student spent looking at the instructor on each theory class (out of 110 minutes of each lesson).
- Theory.TakeNotes: This was gathered semi-automatically from the videos. This value measured how much time the student spent typing notes or writing during each theory class (out of 110 minutes of each lesson).

The teacher recorded all of the theory classes using a camera on the lecturer's desk (see Fig. 2). We also used the individual photos of each student provided in Moodle in order to recognize them. Two researchers involved in the study analyzed the 1650 minutes of recorded video to ensure reliability.

We created two specific programs in Python to semi-automatically produce the attention and note-taking variables. The first program detected the proportion of time a student's face was facing forwards. The second program detected the proportion of time a student's pen was vertical. It was not possible to detect these for all of the students simultaneously, so these two programs were executed for specific coordinates for each student's head and hands (57 executions per program). Although the time values produced were not one hundred percent accurate, they were very close to what we observed. When looking at the videos we noted that there were times when students had their eyes closed, or were looking forwards but not at the teacher or the blackboard or slides, and occasionally students had their pens in their hands without writing.

An Excel file was produced with the values of each attribute for each student in each theory class session.

### 4.1.2. Practical sessions

Practical sessions were those in which the students applied their theoretical learning, such as using two operating systems (Windows and Linux), two office applications (Excel and Access) and a visual programming interface (IDE in Python). We selected Python due to it is a high-level and general-purpose programming language for Engineers. In fact, it is one of the current most popular programming languages thanks to the language versatility, scripting & automation, and it is simple and easy to learn

The own professor of practices collected the following information about the 10 face-to-face practical sessions:

- Practical.Attendance: This was obtained starting from the signing sheet used to monitor each day's attendance. In this course, there were 5 practical subjects spread over 10 h of lessons on 10 face-to-face practical sessions. The value was 1 for a session the student attended, and 0 for a session the student did not attend.
- Practical.Score: This was each students' score from each practical subject, graded by the teacher for each of the five practical subjects. The values were between 1 and 10.

**Fig. 2.** A snapshot of a theory classroom.

The teacher provided us an Excel file with student attendance and scores.

#### 4.1.3. Online sessions

Students also interacted with Moodle for accessing all of the complementary online resources provided by the teacher, including slide-show files for each theory class, a description of each practical, forum discussions, online activities, and quizzes.

The following information was obtained from Moodle logs [15] about student interaction with the online course:

- Moodle.Quiz: This was the students' scores obtained in a Moodle multiple-choice test set by the teacher to test each students' performance in the middle of the course. This was a value between 0 and 10.
- Moodle.Forum: This was the number of contributions/actions each student made to the Moodle discussion forum for the course, either consulting their peers, asking, or answering questions. This ranged from 0 to a maximum value provided by the most active student in the forum.
- Moodle.Task: This was the number of activities that each student uploaded into the Moodle. The instructor requested the students to complete 5 compulsory and 3 optional activities. This variable ranged from 0 to 8.
- Moodle.Time: This was the total time that each student spent logged/connected in Moodle. Each time that a student login to Moodle began a new work session, and the connection time was recorded. This value ranged from 0 to the time spent by the student who spent most time connected to the platform. In some cases user do not explicitly close the session but instead directly close the browser window which produces false values. We solved this kind of problem with outliers in our data files by using specific pre-processing algorithms [16].

The teacher downloaded the log file of the course from the Moodle interface and we automatically gathered the values for each student by using a specific tool for preprocessing Moodle logs that we had developed for a previous study [17]. This tool generated an Excel file with these four attributes for each student who accessed Moodle.

#### 4.1.4. Final exam

The final exam is the in situ final examination that the students had to do at the end of the course. The exam had two parts: a theory part, on paper, with 6 questions (3 multiple choice, 3 open answer) in one hour; and a computer-based practical part, requiring the students to solve 4 problems in 1 hour. The final score from the exam was the sum of the scores in each part, which was given as a score out of 10.

The teacher provided an Excel file with the students' marks in the final exam.

### 4.2. Preprocessing data

We preprocessed [18] all of the data in the aforementioned Excel files. Firstly, the data were anonymized, We implemented a basic solution, using a randomly generated number as a user Identification (ID) rather than the users' names, and replaced the students' names with the random ID in the four Excel files.

Then the input attributes were normalized/rescaled. In this case we rescaled/normalized all of the input attribute values to the same range [0-1] by using the well-known Min-Max method, which is a linear conversion of the data using the formula: $Zi = Xi - \frac{min(X)}{\max(X)} - min(X)$, where $\mathrm{X} = (\mathrm{x}1, \ldots, \mathrm{xn})$.

Next, the output attributes and input attributes were discretized. We stored the 10 input attributes both in numerical and categorical formats. In order to do discretization we used the well-known Equal-Width binning with the following 3 bins/labels: *High, Medium* and *Low*. This method divides all the possible values into only *N* subranges of the same size using the equation: $bin_{width} = \frac{\max Value - - \mathrm{Min} Value}{N}$.

We also discretized the output attribute or class to predict (the students' final academic performance or status). We used a manual method in which the own user/instructor had to specify the cut-off points. In our case, the class had the following 3 values and cut-off points:

- *PASS*: Students scoring 5 or more out of 10 in the final test. In our case, this was 19 out of 57 students (33.33%).
- *FAIL*: Students scoring less than 5 out of 10 in the final test. In our case, this was 17 out of 57 students (29.82%).
- *DROPOUT*: Non-completing students who chose not to do the compulsory final test, and thus did not successfully complete the course [19]. In our case, this was 21 out of 57 students (36.84%).

Finally, we converted the files from Excel to CSV (Comma-separated values) files. It is a delimited text file in which each line of the file is a full data record and it uses a comma character/symbol in order to separate values. We transformed each of the two versions of the four Excel files (numerical and categorical values) into CSV files because they can be directly opened and used by the Weka data mining framework that we used in the experiments [20].

## 5. Experiments and discussion

We carried out four different experiments using four data fusion approaches and several classification algorithms with the preprocessed numerical and discretized data to predict academic performance in a university course (see Fig. 3).

We used two types of white box classification models: rule induction algorithms and decision trees. The obtained models by these algorithms (IF-THEN rules de Decision Trees) are simple and clear and so, they are easy to understand by humans. On the one hand, IF-THEN classification rules provide a high-level knowledge representation that is used for decision making. On the other hand, decision trees can also be converted into a set of IF-THEN classification rules. In our experiments, we selected six well-known classification algorithms integrated in Weka data mining tool [20]: three decision trees algorithms (J48, REPTree and RandomTree) and three rule induction algorithms (JRip, Nnge and PART).

We executed each algorithm using a k-fold cross-validation ($k = 10$). In this way, the dataset is randomly divided into 10 disjointed subsets of equal size in a stratified manner. Of the k (10) subsamples, a single subsample is used as the validation data for testing the model and the remaining other k-1 (9) subsets are used/combined to form the training data. This process is repeated k (10) times and the result is averaged in one single estimation. In order to compare the prediction performance of the algorithms, we needed to select some specific classification metrics from all those previously used in the literature [21]. Some of the most popular evaluation metrics for classification are: Accuracy(Acc), Precision, Recall, Specificity, F-measure, F1-score, Log Loss, Geometry mean, Area Under the Curve(AUC), etc. We selected the following:

- **Accuracy(ACC)** is the most used traditional method to evaluate classification algorithms. It provides a single-number summary of performance. In our case, it is obtained by the next equation: $Acc = \frac{\text{Number of students correctly classified}}{\text{Total number of students}}$. This metric show the percentage of correctly classified students.
- **Area under the ROC curve(AUC)** measures the two dimensional area underneath the entire Relative Operating Characteristic (ROC) curve. ROC curve lets to find possibly optimal models and to discard the suboptimal ones. AUC is often used when the goal of classification is to obtain a ranking because ROC curve construction requires to produce a ranking.

### 5.1. Experiment 1: merging all attributes

In experiment 1 we applied the classification algorithms to a single file with all the attributes merged.

Firstly, we fused the different values (for each session) of the 6 attributes collected in the theory and practical sessions in order to

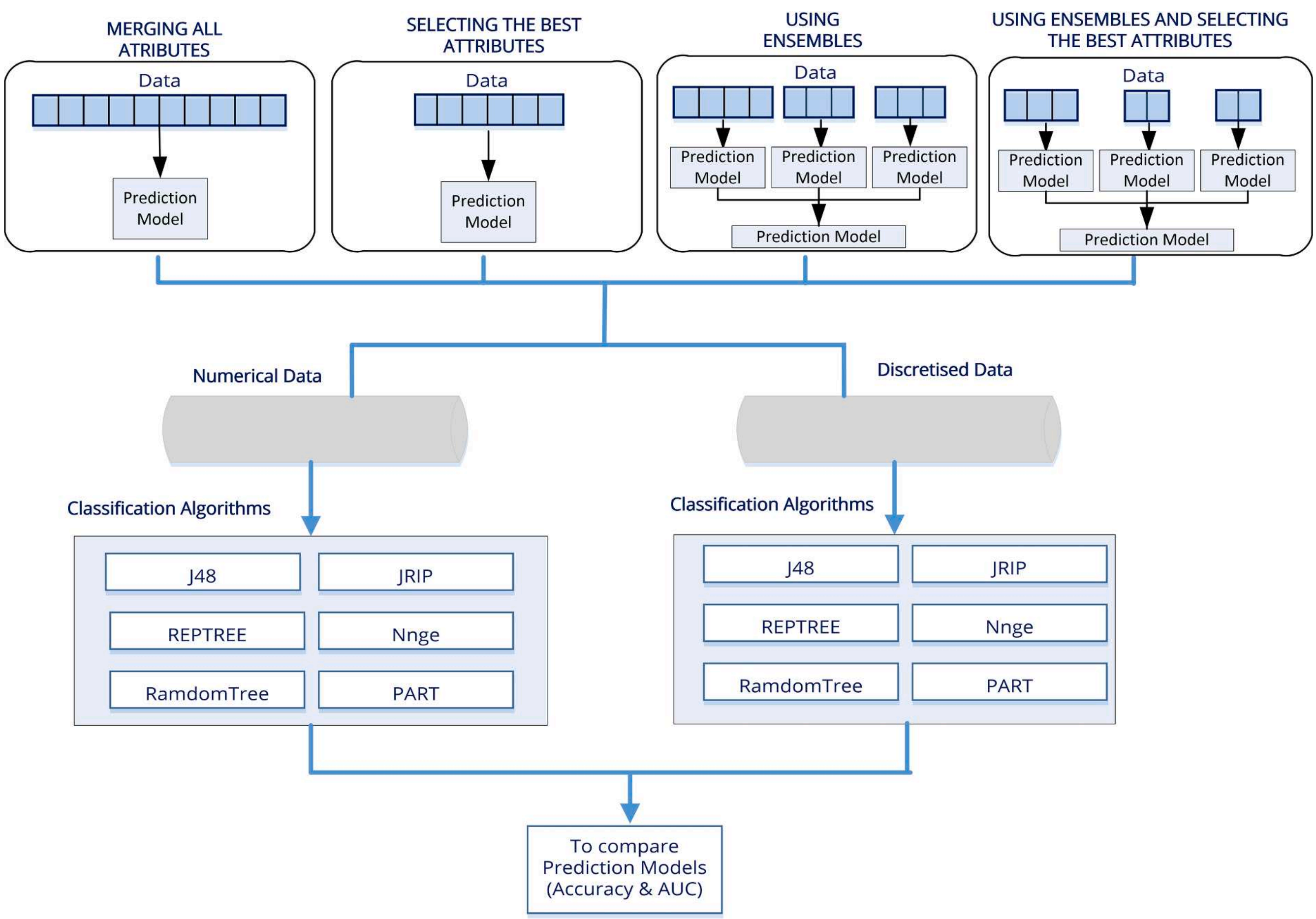


**Fig. 3.** Visual description of the experiments.

have just one single value for each attribute. In our case, we had 15 values (15 lectures) for each one of the 4 attributes collected in the face-to-face theory classes and 10 (10 sessions) and 5 (5 practicals) values respectively for each of the 2 attributes for face-to-face practice sessions. Fusing the 4 4 values about the on-line sessions was not necessary because the specific tool that we used for pre-processing the Moodle logs [17] gave a single value for each attribute directly. To fuse the numerical values, we averaged, that is, we calculated the arithmetic mean by summing of all of the values and dividing by the number of values. In order to merge the discrete or categorical values, we used the mode; the value that appeared most often. It was not necessary to do anything to the files containing the students' academic performance or course status. Following that, we merged the four CSV files into a single CSV file by combining the fused values from each row with the same ID number (without adding the ID number itself) for each file. The same procedure was used for the numerical and discrete/categorical CSV files in order to generate two summary datasets. Each dataset has ten input attributes (in numerical or discrete format) and only one output attribute or class. Finally, we executed six classification algorithms on the two summary datasets and we produced the results (%Accuracy and ROC Area) shown in Table 1.

Table 1 shows that the best results (highest values) were produced by Nnge (80.45 %Acc) and Part (80.45%Acc and 0.91 AUC) algorithms. On average, most of the algorithms exhibited slightly improved performance in both measures when using discretized data.

### 5.2. Experiment 2: selecting the best attributes

In experiment 2 we applied the classification algorithms to a single file with only the best attributes.

Firstly, we applied attribute selection algorithms to the summary files from experiment 1 in order to eliminate redundant or irrelevant attributes. That helps to find the optimal feature set most strongly correlated with the class to predict. The selection of characteristics is important in the classification process by reducing not only the dimensions of the characteristic set but also the additional calculation time required by the classification algorithms. We used the well-known *CfsSubsetEval* (Correlation-based Featured Selection) method provided by the Weka tool [20]. This assesses the merit of attribute subsets by looking at the predictive capacity of each feature in the subset and how redundant they are. In this way, it selects the features that are more correlated with the class. Starting from our initial 10 input attributes, we produced two sets of 3 different optimal attributes (see Table 2) for the numerical and discretized datasets.

Following that, we executed the six classification algorithms with the two new summary datasets producing the results (%Accuracy and ROC Area) shown in Table 3.

Table 3 shows that the best results (highest values) were produced by Jrip (82.45%Acc), Nnge (80.45 %Acc), and J48 (82.45 %Acc and 0.92 AUC) algorithms. Again, on average most of the algorithms exhibited slightly improved performance in both measures when using discretized data.

### 5.3. Experiment 3: using ensembles

In experiment 3 we applied an ensemble of classification algorithms to each different source of data.

First, we created three different sets of datasets starting from the fused attribute values in experiment 1. However, instead of merging all of the attributes from the 4 data sources into a single file, we added the students' final academic status to each dataset. This produced three sets of datasets (6 files in total): two files (numerical and discrete version) for the theory classes with 4 input attributes and 1 output attribute or class; two files (numerical and discrete version) for the practical session with 2 input attributes and 1 output attribute or class; and two files (numerical and discrete version) for the online Moodle sessions with 4 input attributes and only one output attribute or class.

Following that, we applied an ensemble or combination of multiple classification base models generated for each of our different sources of data [22]. We used the well-known Vote approach provided by WEKA for automatic combination of machine learning algorithms. This approach tries to combine the probability distributions of these base classifiers. It produces better results than individual classification models if the classifiers of the sets are accurate and diverse. It has demonstrated better results than homogeneous models for standard datasets. Vote adaptively resamples and combines so that resampling weights are increased for those cases more often misclassified and the combination is done by weighted vote. In order to select the best weighting (for each individual classification model) we tested it by giving the same weight (1) or double that (2) to each individual model. The best result with our data was obtained when combining a weight of 1 for Theory and Practical with a weight of 2 for Moodle by using the average as the

**Table 1**
Results produced by merging all attributes.

| | Numerical data % Accuracy | AUC | Discretized data %Accuracy | AUC |
|---|---|---|---|---|
| Jrip | 77.1930 | 0.8440 | 78.9474 | 0.8880 |
| Nnge | **80.4561** | 0.8760 | 75.4386 | 0.8630 |
| PART | 78.9474 | 0.8640 | **80.4561** | **0.9170** |
| J48 | 75.4386 | 0.8640 | 78.9474 | 0.8780 |
| REPTree | 75.4386 | 0.8630 | 76.6667 | 0.8480 |
| Randomtree | 70.1754 | 0.7820 | 73.6842 | 0.8180 |
| Avg. | 76.2748 | 0.8488 | 77.3567 | 0.8686 |

**Table 2**
Results of the attribute selection with CFSSubsetEval.

| Dataset | # selected features | Name of Selected features |
|---|---|---|
| Normalized | 3 | Theory.Location<br>Moodle.Quiz<br>Theory.Notes |
| Discretized | 3 | Theory.Attention<br>Moodle.Quiz<br>Moodle.Forum |

**Table 3**
Results obtained when selecting the best attributes.

| | Numerical data<br>% Accuracy | AUC | Discretized data<br>%Accuracy | AUC |
|---|---|---|---|---|
| Jrip | 80.7018 | 0.8490 | **82.4561** | 0.9140 |
| Nnge | **82.4561** | 0.9140 | 78.9474 | 0.8430 |
| PART | 77.1930 | 0.8750 | 80.7018 | 0.9140 |
| J48 | 80.7018 | 0.8680 | **82.4561** | **0.9230** |
| REPTree | 77.1930 | 0.8940 | 78.9474 | 0.8880 |
| Randomtree | 75.4386 | 0.8320 | 82.4561 | 0.9170 |
| Avg. | 78.9473 | 0.8720 | 80.9941 | 0.8998 |

combination rule for weights.

We executed the six classification algorithms as base or individual classification models of our Voting method for the 6 previously generated summary datasets. Table 4 shows the results (%Accuracy and ROC Area).

Table 4 shows that the best results (highest values) were produced by Jrip (85.96%Acc and 0.93 AUC). Once again, on average most of the algorithms exhibited slightly improved performance in both measures when using discretized data.

### 5.4. Experiment 4: using ensembles and selecting the best attributes

In experiment 4 we applied an ensemble of classification algorithms to the best attributes from each different source of data.

Firstly, we selected the best attributes for each of the three different sets of datasets (6 files in total) generated in experiment 3. For that, we again used the well-known *CfsSubsetEval* attribute selection algorithm, producing the list of attributes shown in Table 5.

Following that, we applied an ensemble or combination of multiple classification base models by again using the *Vote* automatic combining machine learning algorithm. To find the best weights (for each individual classification model) we tested it by giving the same weight (1) or double that (2) to each individual model. The best result with our data was obtained when combining a weight of 1 for Theory and Practical with a weight of 2 for Moodle by using the average as combination rule for weights.

We executed the six classification algorithms as base or individual classification models of our Voting method for the 6 previously generated summary datasets. Table 6 shows the obtained results (%Accuracy and ROC Area).

Table 6 shows that the best results (highest values) were produced by REPTree (87.47 %Acc and 0.94 AUC). Again, on average, most of the algorithms exhibited slightly improved performance in both measures when using discretized data.

### 5.5. Discussion

Following, we address the two initial research questions by discussing the results produced by our four experiments.

**Table 4**
Results obtained when using ensembles.

| | Numerical data<br>% Accuracy | AUC | Discretized data<br>%Accuracy | AUC |
|---|---|---|---|---|
| Jrip | 82.4561 | 0.9230 | **85.9649** | **0.9380** |
| Nnge | 77.1930 | 0.8770 | 77.1930 | 0. 8770 |
| PART | 80.7018 | 0.9040 | 82.4561 | 0.9130 |
| J48 | 82.4561 | 0.9110 | 82.4561 | 0.9220 |
| REPTree | 82.4561 | 0.9230 | 82.4561 | 0.9220 |
| Randomtree | 77.1930 | 0.8360 | 79.9474 | 0.9170 |
| Avg. | 80.4093 | 0.8956 | 81.7456 | 0.9185 |

**Table 5**
Results of attribute selection with CFSSubsetEval.

| Dataset | Type | # selected features | Name of Selected features |
|---|---|---|---|
| Theory | Numerical | 2 | Theory.Attendance<br>Theory.Attention |
| | Discretized | 1 | Theory.Attention |
| Practice | Numerical | 2 | Practice.Attendance<br>Practice.Score |
| | Discretized | 2 | Practice.Attendance<br>Practice.Score |
| Moodle | Numerical | 2 | Moodle.Quiz<br>Moodle.Forum |
| | Discretized | 2 | Moodle.Quiz<br>Moodle.Forum |

**Table 6**
Results obtained when using ensembles and selection of the best attributes.

| | Numerical data | | Discretized data | |
|---|---|---|---|---|
| | % Accuracy | AUC | %Accuracy | AUC |
| Jrip | 82.4561 | 0.9170 | 84.2105 | 0.9310 |
| Nnge | 80.7018 | 0.9020 | 78.9474 | 0.8900 |
| PART | 80.7018 | 0.9010 | 82.4561 | 0.9350 |
| J48 | 82.4561 | 0.8990 | 84.2105 | 0.9350 |
| REPTree | 84.2105 | 0.9130 | **87.4737** | **0.9420** |
| Randomtree | 77.1930 | 0.9160 | 82.4561 | 0.9330 |
| Avg. | 81.2865 | 0.9080 | 83.2923 | 0.9276 |

#### 5.5.1. *Answering question 1*

Our first research question was: Which data fusion approach and classification algorithms produce the best results from our data? We used four different data fusion approaches and six white-box classification algorithms to answer this question. The four proposed data fusion approaches were not completely different. They were consecutive, or incremental approaches, each one was a modified or extended version of one or more of the previous approaches:

1 Merging all attributes. Our first data fusion approach which uses a simple approach to Naïve fusion in which general summary statistics are generated by combining the different data sources.
2 Selecting the best attributes. Our second approach (modifying the first approach) in which we applied a reduction of features by selecting the best attributes starting from the previous general summary statistics.
3 Using ensembles. Our third approach (which modified the first approach) applied decision-level fusion to combine the results of 3 classifiers, one for each individual statistical summary of our 3 data sources.
4 Using ensembles and selection of the best attributes. Our fourth approach (Implementation of a hybrid between attribute selection algorithms, classification algorithms and automatic learning algorithms) applied decision-level fusion again combining the results of 3 classifiers but this time having previously selected the best attributes in each of the 3 individual statistical summaries of each data source.

Table 7 shows that the average prediction performance (Average of % Accuracy and AUC) of the classification algorithms increased in each new approach. The second approach improved on the first approach, the third approach improved on the second approach and the best result was produced using the fourth approach of using ensembles and selection of the best attributes. In all the approaches the

**Table 7**
Average results obtained in the four data fusion approaches.

| Average | Numerical data | | Discretized data | |
|---|---|---|---|---|
| | % Accuracy | AUC | %Accuracy | AUC |
| Merging all attributes | 76.2749 | 0.8488 | 77.3567 | 0.8687 |
| Selecting the best attributes | 78.9474 | 0.8720 | 80.9942 | 0.8998 |
| Using ensembles | 80.4094 | 0.8957 | 81.7456 | 0.9185 |
| Using ensembles and selection of the best attributes | 81.2866 | 0.9080 | **83.2924** | **0.9277** |

average values were higher when using discretized data than numerical data.

We were unable to find a single best algorithm that would win in all cases in our experiments (8 cases = 4 experiments * 2 different datasets, numerical and discretized). This is logical and is in line with the No-Free-Lunch theorem [23], in which it is generally accepted that no single supervised learning algorithm can beat another algorithm over all possible learning problems or different datasets. In the first experiment, the algorithm that produced the highest prediction values was PART (80.4561% Accuracy and 0.9170 AUC), in the second experiment it was J48 (82.4561 Accuracy and 0.9230 AUC), in the third it was Jrip (85.9649 Accuracy and 0.9380 AUC), and finally the algorithm that produced the highest prediction values of Accuracy (87.4737%) and AUC (0.9420) was REPTree when using an ensemble and selection of the best attributes from the discretized data in the fourth experiment.

#### 5.5.2. Answering question 2

Our second research question was: How useful are the prediction models we produce to help teachers detect students at risk of drop out or fail the course? To answer that, we will demonstrate and describe the prediction model that produced the highest values of Accuracy and AUC in each of our 4 experiments.

In experiment 1, the prediction model that produced the best prediction was generated by the PART algorithm using discretized data (see Table 8).

This prediction model (see Table 8) consists of 5 rules that show that the students who had high scores in Moodle quizzes or who had medium scores in Moodle quizzes and also paid attention in theory classes, were the students who passed the course. The students who failed the course were those who got low scores in the Moodle quizzes. The students who dropped out from the course were those who pied little attention in theory classes and also showed low activity in the Moodle forum. The remaining students were classified as passing.

In experiment 2, the prediction model that produced the highest prediction values used the J48 algorithm with the discretized data (see Table 9).

This prediction model (see Table 9) is a decision tree with 9 leaves that can be transformed into 9 prediction rules. These rules show that the students who passed the course are those who had medium scores in Moodle quizzes and also paid medium to high attention in theory classes, or those who simply had high scores in Moodle quizzes. The students who dropped out from the course are those who had low scores in Moodle quizzes, showed low activity in the Moodle forum, and also paid little attention in theory classes. In addition, students who failed are those who had low scores in Moodle quizzes, showed low activity in the Moodle forum and paid medium to high attention in theory classes. There are also other failing student profiles: students who had medium scores in Moodle quizzes and also paid little attention in theory classes; students who had low scores in Moodle quizzes, showed low activity in the Moodle forum, and paid medium to high attention in theory classes.

In experiment 3, the prediction model that produced the highest prediction values used the JRIP algorithm with discretized data (see Table 10).

This prediction model (see Table 10) is a combination of three models that show differential student behavior related to theory, practice and Moodle. The students who regularly attended theory classes passed the course; the students who exhibited low attendance finally dropped out. The students who regularly attended practical classes and exhibited high performance in those practical classes then passed the entire course. In contrast, the students who rarely attended practical classes and had low performance in practicals then failed the entire course. The students who uploaded a moderate number of activities to the Moodle platform or got high scores in Moodle quizzes are students who passed the course; and logically, the students who uploaded a low number of activities to the Moodle platform and got low scores in Moodle quizzes are students who failed the course, but the students with medium performance in quizzes and low contributions to the forum also failed.

In experiment 4, the prediction model that produced the highest prediction values used the RepTree algorithm with discretized data (see Table 11).

This prediction model (see Table 11) is also a combination of three models that show differential student behavior related to theory, practicals and Moodle. Exhibiting low attention in theory classes, low practical attendance, or low scores in Moodle quizzes plus little forum participation seems to lead to students dropping out. At the same time, students exhibiting medium or high attention, or medium to high Moodle forum participation, fail; those demonstrating medium practical attendance or high practical attendance plus low or medium practice score also fail. The students that demonstrated high practical attendance and performance passed, as did the students with medium to high scores in Moodle quizzes.

In general, we can see that these white-box models are very useful for explaining to the teacher how the predictions of pass, fail or dropout are arrived at. The teacher can discover what the main predictive attributes and values are directly from the background of the

**Table 8**
PART decision list when merging all attributes.

| |
|---|
| IF Moodle.Quiz = High THEN Pass |
| IF Moodle.Quiz = Medium AND Theory.Attention = Medium THEN Pass |
| IF Moodle.Quiz = Low THEN Fail |
| IF Theory.Attention = Low AND Moodle.Forum = Low THEN Dropout |
| ELSE Pass |
| Number of Rules: 5 |

**Table 9**
J48 pruned tree when selecting the best attributes.

```
IF Moodle.Quiz = Low
    | Moodle.Forum = Low
    | | Theory.Attention = Low THEN Dropout
    | | Theory.Attention = Medium THEN Fail
    | | Theory.Attention = High THEN Fail
    | Moodle.Forum = Medium THEN Fail
    | Moodle.Forum = High THEN Fail
    ELSE IF Moodle.Quiz = Medium
    | Theory.Attention = Low THEN Fail
    | Theory.Attention = Medium THEN Pass
    | Theory.Attention = High THEN Pass
    ELSE IF Moodle.Quiz = High THEN Pass
    Number of Leaves: 9
    Size of the tree: 13
```

IF-THEN rules. In this sense, the presence of the attributes attention in the classroom, forum participation and score in Moodle quizzes is notable. It is important to notice that all the models produced only had attributes from the theory and online sessions, not from the practical sessions. This may be due to the variables provided/obtained from the practicals were not discriminating in predicting the students' final performance. Our results also revealed that the most discriminant information source was student behavior in Moodle. In this regard, we saw that the two ensemble approaches had optimal weighting when giving greater weight to the on-line data source. Although it is not clear as to how much online learning is inherent in blended learning [24], these results seem to point to the conclusion that the use of distance learning platforms in the b-learning educational experience is positive [25].

## 6. Conclusions

This paper uses different data fusion approaches in blended learning for answering two research questions:

- Which data fusion approach and classification algorithms produce the best results from our data? The use of ensembles and selecting the best attributes approach from discretized summary data produced our highest/best results in Accuracy and AUC values. The REPTree classification algorithm obtained the highest/best results in this approach from discretized summary data.
- How useful are the prediction models we produce to help teachers detect students who are at risk of drop out or fail the course? The white-box models we produced give teachers very understandable explanations (IF-THEN rules) of how they classified the students' final performance or classification. They showed that the attributes that appear most in these rules were attention in theory classes, scores in Moodle quizzes, and the level of activity in the Moodle forum.

As next step, we intend to investigate and do new experiments for trying to improve our process and to overcome some limitations:

- Analyzing the video automatically rather than manually or semi-automatically. Processing the video recordings automatically would gather information more efficiently compared to manual coding. The use of multiple web-cams distributed around the classroom, rather than a single camera, will let us use more advanced algorithms for detecting student engagement more accurately.
- Using raw data and other specific data fusion techniques. We used a basic Naïve and knowledge-based fusion method that uses summary data. However, there are other fusion theories/methods data such as Probability-based methods (PBM) and Evidence reasoning methods (EBM) that we can use with raw data. We could also use semantic (abstract) level features in order to produce intelligent data aggregation.

## Author statement

Wilson Chango, Rebeca Cerezo and Cristóbal Romero certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript.

## Declaration of Competing Interest

The Authors of this paper (Wilson Chango, Rebeca Cerezo and Cristóbal Romero) do not have any conflicts of interest to declare about this paper title: "Multi-source and Multimodal data fusion for predicting academic performance in blended learning university courses" submitted to the special issue "Special Section on Recent Advancements in Big Data Fusion" of the journal Computers & Electrical Engineering.

**Table 10**
JRIP when using ensembles.

| |
|---|
| JRIP rules (Theory): |
| =========== |
| IF (Theory.Attendance = High) THEN Pass |
| IF (Theory.Attention = Low) THEN Dropout |
| ELSE Dropout |
| Number of Rules: 3 |
| JRIP rules (Practice): |
| =========== |
| IF (Practice.Attendance = High) and (Practice.Score = High) THEN Pass |
| IF (Practice.Attendance = Low) and (Practice.Score = Low) THEN Fail |
| ELSE Dropout |
| Number of Rules: 3 |
| JRIP rules (Moodle): |
| =========== |
| IF (Moodle.Task = Low) and (Moodle.Quiz = Low) THEN Fail |
| IF (Moodle.Quiz = Medium) and (Moodle.Forum = Low) THEN Fail |
| IF (Moodle.Task = Medium) THEN Pass |
| IF (Moodle.Quiz = High) THEN Pass |
| ELSE Dropout |
| Number of Rules: 5 |

**Table 11**
RepTree when using ensembles with selecting the best attributes.

| |
|---|
| REPTree (Theory) |
| ============ |
| IF Theory.Attention = Low THEN Dropout |
| IF Theory.Attention = Medium THEN Fail |
| IF Theory.Attention = High THEN Pass |
| Size of the tree: 4 |
| REPTree (Practice) |
| ============ |
| IF Practice.Attendance = Low THEN Dropout |
| IF Practice.Attendance = Medium THEN Fail |
| IF Practice.Attendance = High |
| \| AND Practice.Score = Low THEN Fail |
| \| OR Practice.Score = Medium THEN Fail |
| \| OR Practice.Score = High THEN Pass |
| Size of the tree: 7 |
| REPTree (Moodle) |
| ============ |
| IF Moodle.Quiz = Low |
| \| AND Moodle.Forum = Low THEN Dropout |
| \| OR Moodle.Forum = Medium THEN Fail |
| \| OR Moodle.Forum = High THEN Fail |
| ELSE IF Moodle.Quiz = Medium THEN Pass |
| ELSE IF Moodle.Quiz = High THEN Pass |
| Size of the tree: 7 |

## Acknowledgments

This work would not be possible without the funding from the Ministry of Sciences and Innovation I+D+I (TIN2017-83445-P and PID2019-107201GB-100) and the European Regional Development Fund and the Principality of Asturias (FC-GRUPIN-IDI/2018/000199).

## References

[1] Thompson K, Jowallah R, Cavanagh TB. Solve the big problems: leading through strategic innovation in blended teaching and learning. Technology leadership for innovation in higher education. IGI Global; 2019. p. 26–48. https://10.4018/978-1-5225-7769-0.ch002.
[2] Maryam T, Marlia P. Blended learning or E-learning? Imacst 2012;3(1):103–10. https://doi.org/10.1016/j.iheduc.2012.12.001.
[3] Romero C, Ventura S. Educational data science in massive open online courses. Wiley Interdiscip Rev 2017;7(1). https://doi.org/10.1002/widm.1187.
[4] Bogarín A, Cerezo R, Romero C. A survey on educational process mining. Wiley Interdiscip Rev 2018;8(1). https://doi.org/10.1002/widm.1230.
[5] Romero C, Ventura S. Educational data mining and learning analytics: an updated survey. WIREs Data Min Knowl Discov 2020:e1355. https://doi.org/10.1002/widm.1355.
[6] Zacharis NZ. Predicting student academic performance in blended learning using Artificial Neural Networks. Int J Artif Intell Appl 2016;7(5):17–29. https://doi.org/10.5121/IJAIA.2016.7502.

[7] Greene J, Azevedo R. The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. Educ Psychol 2010;45(4):203–9. https://doi.org/10.1080/00461520.2010.515935.
[8] Blikstein P, Worsley M. Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. J Learn Anal 2016;3(2):220–38. https://doi.org/10.18608/jla.2016.32.11.
[9] Ochoa X. Multimodal learning analytics. The handbook of learning analytics. Handbook of Learning Analytics; 2017. p. 129–41. https://doi.org/10.18608/hla17.
[10] Worsley M. Multimodal learning analytics as a tool for bridging learning theory and complex learning behaviors. In: ACM workshop on multimodal learning analytics workshop and grand challenge; 2014. p. 1–4. https://doi.org/10.1145/2666633.2666634.
[11] Lahat D, Adali T, Jutten C. Multimodal data fusion: an overview of methods, challenges, and prospects. Proc IEEE 2015;103(9):1449–77. https://doi.org/10.1109/JPROC.2015.2460697.
[12] Han Z, Wu J, Huang C, Huang Q, Zhao M. A review on sentiment discovery and analysis of educational big-data. Wiley Interdiscip Rev 2020;10(1):e1328. https://doi.org/10.1002/widm.1328.
[13] Ding W, Jing X, Yan Z, Yang LT. A survey on data fusion in internet of things: towards secure and privacy-preserving fusion. Inf Fusion 2019;51:129–44. https://doi.org/10.1016/j.inffus.2018.12.001.
[14] Jesus J, Araújo D, Canuto A. Fusion approaches of feature selection algorithms for classification problems. In: IEEE 5th Brazilian conference on intelligent systems; 2016. p. 379–84. https://doi.org/10.1109/BRACIS.2016.075.
[15] Romero C, Ventura S, García E. Data mining in course management systems: moodle case study and tutorial. Comput Educ 2008;51(1):368–84. https://doi.org/10.1016/j.compedu.2007.05.016.
[16] Martínez JA, Campuzano J, Sancho-Vinuesa T, Valderrama E. Predicting student performance over time. A case study for a blended-learning engineering course. LASI 2019;13.
[17] López-Zambrano J, Martinez JA, Rojas J, Romero C. A tool for preprocessing moodle data sets. In: Proceedings of the 11th international conference on educational data mining; 2018. p. 15–8.
[18] Romero C, Romero JR, Ventura S. A survey on pre-processing educational data. Educational data mining. Cham: Springer; 2014. p. 29–64. https://doi.org/10.1007/978-3-319-02738-8_2.
[19] Levy Y. Comparing dropouts and persistence in e-learning courses. Comput Educ 2007;48(2):185–204. https://doi.org/10.1016/j.compedu.2004.12.004.
[20] Witten I, Frank E, Hall M. Data mining: practice machine learning tools and techniques. second ed. 2011. https://doi.org/0120884070, 9780120884070.
[21] Espejo, P. G., Gibaja, E., Menéndez, V. H., Zapata, A., Romero, C. Improving multi-label classification for learning objects categorization by taking into consideration usage information. 25 (13): 1687-1716. 2019. https://doi.org/10.3217/jucs-025-13-1687.
[22] Ludmila I, Kuncheva. Combining pattern classifiers: methods and algorithms. John Wiley and Sons, Inc.; 2004. https://doi.org/10.1002/9781118914564.
[23] Wolpert DH. The supervised learning no-free-lunch theorems. Soft computing and industry. London: Springer; 2002. p. 25–42.
[24] Garrison DR, Kanuka H. Blended learning: Uncovering its transformative potential in higher education. Internet Higher Educ 2004;7(2):95–105. https://doi.org/10.1016/j.iheduc.2004.02.001.
[25] Lara JA, De Sojo AA, Aljawarneh S, Schumaker RP, Al-Shargabi B. Developing big data projects in open university engineering courses: lessons learned. IEEE Access 2020;8:22988–3001. https://doi.org/10.1109/ACCESS.2020.2968969.

**Wilson Chango** is a Associate Professor at the Pontifical Catholic University of Ecuador – Esmeraldas. He is currently a PhD student in advanced computing at the University of Córdoba in Spain.

**Rebeca Cerezo** started to work in 2007 as research fellow in The ADIR Research Group from the University of Oviedo and is Associate Professor at the Department of Psychology since 2010.

**Cristóbal Romero** is a Full Professor at the University of Córdoba in Spain and member of the KDIS (Knowledge Discovery and Intelligent Systems) research group.

Journal of Computing in Higher Education (2021) 33:614–634
https://doi.org/10.1007/s12528-021-09298-8

# Improving prediction of students' performance in intelligent tutoring systems using attribute selection and ensembles of different multimodal data sources

**Wilson Chango[1] · Rebeca Cerezo[2] · Miguel Sanchez-Santillan[2] · Roger Azevedo[3] · Cristóbal Romero[4]**

Accepted: 11 August 2021 / Published online: 28 October 2021


## Abstract

The aim of this study was to predict university students' learning performance using different sources of performance and multimodal data from an Intelligent Tutoring System. We collected and preprocessed data from 40 students from different multimodal sources: learning strategies from system logs, emotions from videos of facial expressions, allocation and fixations of attention from eye tracking, and performance on posttests of domain knowledge. Our objective was to test whether the prediction could be improved by using attribute selection and classification ensembles. We carried out three experiments by applying six classification algorithms to numerical and discretized preprocessed multimodal data. The results show that the best predictions were produced using ensembles and selecting the best attributes approach with numerical data.

**Keywords** Predicting academic performance · Intelligent tutoring systems · Multisource data · Multimodal learning · Data fusion

## Introduction

The rapid growth of technology has meant that computer learning has increasingly integrated artificial intelligence techniques in order to develop more personalized educational systems. These systems are known as Intelligent Tutoring Systems (ITS).

MetaTutorES (Cerezo et al., 2020a,b), a Spanish adaptation of MetaTutor (Azevedo et al., 2011) is an ITS designed to detect, model, trace, and foster students' self-regulated learning while learning various science topics (e.g., by modeling and scaffolding metacognitive monitoring, facilitating the use of effective

✉ Rebeca Cerezo
cerezorebeca@uniovi.es

Extended author information available on the last page of the article

learning strategies, and setting and coordinating relevant learning goals). The system uses human-like avatar technology that allows pedagogical agents to track student behavior and provide interaction on this basis. Tracking students' behavior is also a powerful research tool used to collect data on students' cognitive, metacognitive, affective, and motivational processes deployed during learning (Azevedo et al., 2018; Greene & Azevedo, 2010; Taub et al., 2021). These different data sources can be fused and mined to reveal learning-related information such as student performance. In this regard, Educational data mining (EDM) and Learning Analytics (LA) can be applied to understand educational processes using information extracted from educational data, which is then used to improve the educational process and the quality of learning (Romero & Ventura, 2020).

One of the oldest and most commonly studied issues in EDM/LA is the prediction of learners' performance. It is still a challenge to predict student learning achievement in ITSs using Multimodal Learning Analytics (MLA) with learning data from different sources and doing a single analysis (Blikstein & Worsley, 2016). MLA uses log-files and gaze data, biosensors, interactions with videos, audio and digital documents, and any other relevant data source to measure and understand the learning process.

One important issue in MLA is how to combine, or fuse, the data extracted from various sources/modalities in order to provide a better, more comprehensive view of teaching–learning processes (Bogarín et al., 2018; Chango et al., 2021). The most common and simplest data fusion approach for combining all the data sources is to build a machine-learning classifier from the summary statistics produced from each of the data sources. An important task when fusing data is to reduce the dimensions of the variables/attributes and to identify the most fruitful feature sets. Feature selection algorithms are normally used in data fusion for classification problems in order to reduce the data dimensions and produce the best results (Jesus et al., 2016). Finally, classification ensembles have demonstrated very good results in predicting student academic performance from multimodal data sources (Adejo & Connolly, 2018).

In this paper we perform a classification task, predicting the value of a categorical/nominal attribute (the class or final knowledge status of the student (Pass, Fail) based on other attributes (the predictive attributes from various available data sources). We propose applying classification algorithms, feature selection algorithms, and ensembles to data gathered from a variety of sources (learning strategies from ITS logs, emotions from face recording videos, and interaction zones from eye tracking) in order to predict the students' final performance in the ITS. In this sense, the ultimate contribution of this study is to analyze the learning process through resources, allowing a more personalized response to each learner.

The research questions posed by this study are:

Question 1.- Can attribute selection and classification ensemble algorithms improve the prediction of students' final performance from our ITS data?

Question 2.- How useful are the models produced and what are the best variables to help teachers understand how to predict students' final performance in the ITS?

This paper is organized as follows. The first section covers the background of the related research area of MLA. Subsequently, we describe the proposed methodology, the data used, and how it was preprocessed. Then, we describe the experiments we performed and the results they produced. Finally, we discuss the implications, conclusions, and lines for future research.

## Background

MLA aims to combine different sources of learning traces into a single analysis, it is a subfield of EDM related to multi-view and multi-relational data and data fusion. It aims to understand and optimize learning in digital where the use of videos is currently consolidated, from traditional courses to mixed and online courses (Chan et al., 2020). MLA can generate distinctive insights into what happens when students create unique solution paths to problems, interact with peers, and act in both physical and digital environments. It has become increasingly broadly applied in both digital and in real-world scenarios where interactions are not solely mediated through computers or digital devices (Blikstein & Worsley, 2016). In MLA, learning traces are extracted not only from log-files but also from digital documents, recorded video and audio, pen strokes, position tracking devices, biosensors, and any other data sources that could be useful for understanding or measuring the learning process. Below, we describe the data sources used in the present study.

### Learning strategies from ITS logs

There is empirical evidence about performance prediction through computer learning environment log data (Cerezo et al., 2016; Lerche & Kiel, 2018; Li & Tsai, 2017), including predicting performance in offline courses from logs of online behavior (Zhou et al., 2015). As computer-based learning environments, ITSs allow us to see what learning strategies users deploy while they are studying, and are part of a new trend in the measurement of learning in general, and self-regulated learning in particular—the so called third wave—, characterized by combined use of measurement and Advanced Learning Technologies (Panadero et al., 2016; Winne & Azevedo, 2021). These performance analytics include data on the student's performance and different learning metrics. Example include completion time, successful or unsuccessful completion of assignments, speed of task resolution, the number of attempts or failures, and the complexity of the problem-solving process (Crescenzi-Lanna, 2020). All of these data are normally produced by the computer during the student's interaction with the learning environment and are stored in database or log files (Cristóbal Romero et al., 2008). This technology overcomes the limitations of self-report methodology, making it possible to detect, model, and trace students' learning, with the added benefit of not interfering with student activity, because even though a huge amount of data is generated, it is processed automatically by the computer.

### Interaction zones from eye tracking

Eye-tracking devices provide information that can be used to infer the student's attention level, engagement, preference, or understanding. It provides an understanding of what attracts immediate attention, which target elements are ignored, what order elements are noticed in, and how elements compare to others (Cerezo et al., 2020a,b; Taub & Azevedo, 2019). In this sense, gaze data can provide very useful, accurate information for predicting student learning during interaction with ITSs (Bondareva et al., 2013), and multiple researchers have suggested that the duration of fixations are indicators of cognitive processing during learning (Antonietti et al., 2015).

There are different options for collecting eye-tracking data such as saccade amplitude, direction change, fixations, etc. (Crescenzi-Lanna, 2020). In the current study, we are interested in analyzing fixations, particularly the number of fixations in areas that could be related to the learner's final performance. For that purpose, we defined three Areas of interest (AOIs) in our ITS interface: AOI1 Learning session timer, AOI2 ITS agent/avatar, and AOI3 Supporting image/graphics content. These are areas of interest because, in terms of the interface configurations, fixations on AOI1 may denote time management or resource management strategies, while reduced or excessive fixations on AOI1 might indicate poor time management skills. Fixation on AOI2, the agent, would show that the participant is making use of the prompts and feedback provided by the agents during the learning session and has established an interaction with the agent. Fixations on AOI3 may point to participants using a strategy of coordinating information sources (text-images), associated with learning gains (Azevedo, 2009; Cerezo et al., 2020a,b).

### Emotions from face recording videos

Emotions are a critical component of learning and problem solving, especially when it comes to interacting with computer-based learning environments (Harley et al., 2015), and there is a relationship between negative learning emotion and learning performance (Chen & Wang, 2011). In this context, studies from affective computing literature suggest that facial expressions may be the best single method for accurately identifying emotional states (D'Mello & Kory, 2012). Techniques for automatic detection of emotions (Blanchard et al., 2009) are capable of isolating a learner's mood via artificial intelligence facial recognition systems, and there are tools available that can process video data, such as the Microsoft Emotion API (2019), Face API (2019), and Affectiva (2019). In this line, including the learner's emotional states may help enhance ITS quality and efficacy. Previous research has indicated that academic emotions are significantly related to students' motivation, learning strategies, cognitive resources, self-regulation, and academic achievement (Pekrun et al., 2011).

In previous studies (Chango et al., 2020), student emotions as recognized by an API during a learning session with an ITS have been used as the sole data source for predicting the student's final performance. The best models demonstrated a prediction accuracy of 63.82% and 0.67 AUC, figures that we aim to improve on by using more student features and variables from various multimodal data sources, together with ensembles and selection of the best attributes.

## Proposal

The current study proposes a two-stage methodology for predicting students' final performance from multimodal data (see Fig. 1).

As Fig. 1 shows, the two main stages in our methodology are:

- First stage. Collecting data from various sources: learning strategies from MetaTutorES logs, number of fixations from gaze data, and emotions from face recording videos. It also includes some pre-processing tasks (anonymization, attribute normalization and discretization, and format transformation) to generate numerical and categorical datasets.

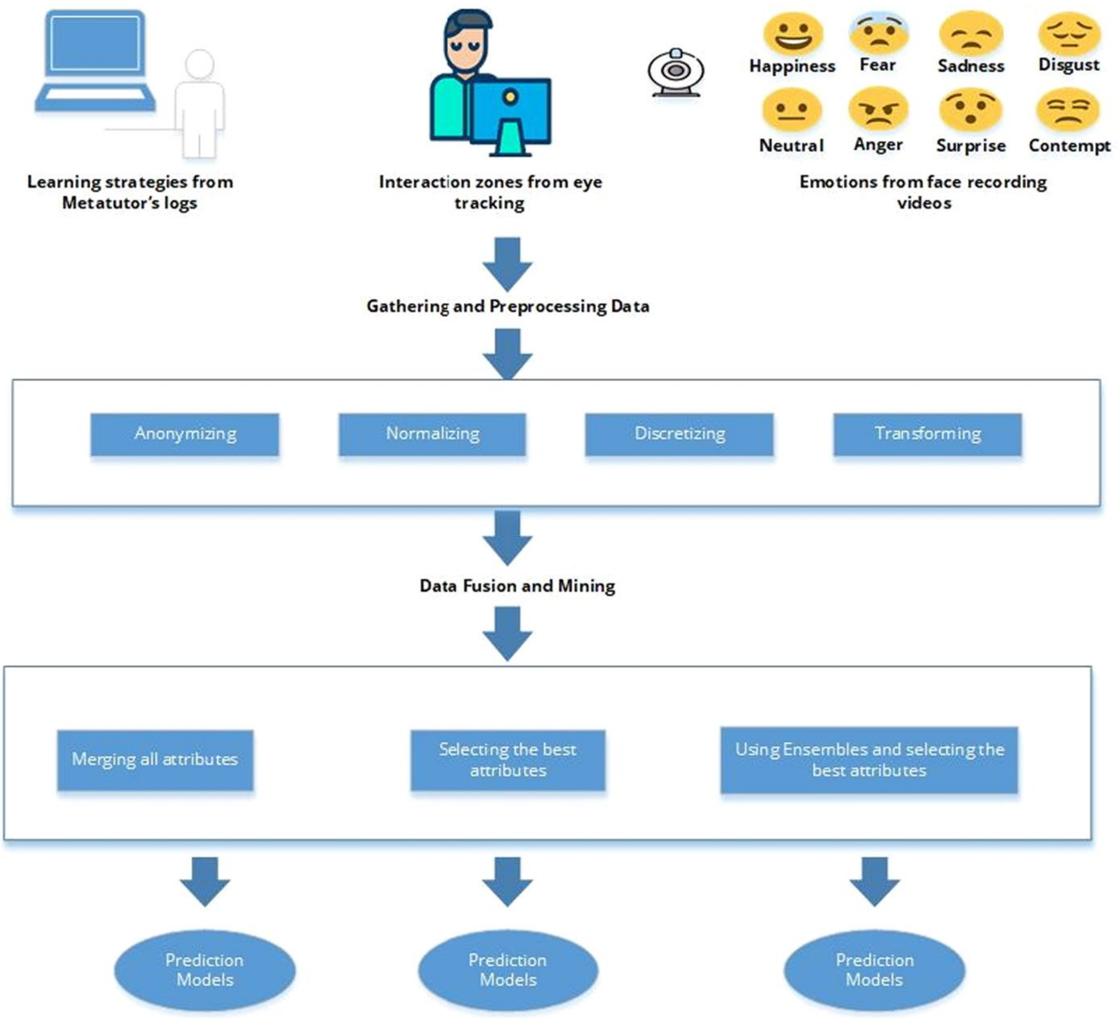


**Fig. 1** Proposed methodology for predicting students' performance from multiple data sources

- Second stage. Using different data fusion approaches: merge all attributes; selection of the best attributes, and ensembles of several white box classification algorithms. Finally, the predictions produced by the models are compared in order to find the best model and attributes to be used to predict the students' final performance.

## Data

Data were collected from 40 undergraduates (mean age = 23.58; SD = 8.18; 17 men and 23 women) enrolled at a public university in the north of Spain. The undergraduates participated in the study voluntarily and learned about a complex science topic (the circulatory system) while interacting with the MetaTutorES ITS (Cerezo et al., 2020a,b), a computerized learning environment. The students in the sample were studying in a variety of different knowledge areas: education, psychology, economics, law, philosophy, nursing, telecommunication, electrical engineering, geomatics, physics, and civil navy. Most students in the sample were first-year undergraduates, but there were also second-years, third-years and masters.

### Gathering data

We gathered information from four ITS data sources: learning strategies from MetatutorES logs, emotions from face videos, fixation from eye tracking, and performance from the content knowledge test. The data collected was produced spontaneously from interactions with the MetaTutorES ITS during a session lasting from two-and-a-half to three hours. The data collection for the study was developed and managed in line with the ethical research principles of the Declaration of Helsinki and the protocol was approved by the research ethics committee of the Principality of Asturias and the University of Oviedo.

### Learning strategies from MetaTutorES logs

Throughout each learning session, learner interaction with the ITS was logged in a log file unique to each learner. The learning environment is made up of information in text, charts, and images, through which students learn about the circulatory system. The system logs each user action and interaction with the learning environment and the study. Each line of a log represents an event or participant action in the learning environment and contains the timestamp of the event, the triggered event, the identifier of the theoretical content that the learner is studying and optional information related to that event.

For the present study, three variables were extracted from the log files: SummAll: The number of times that the learner wrote a *Summary* about the content they were studying, discarding the events in which they did not add any new information, e. g. After spending time reading the page about the role of the heart in the circulatory system, the user summarizes the reading; COIStotalFreq: *Coordinating Information*

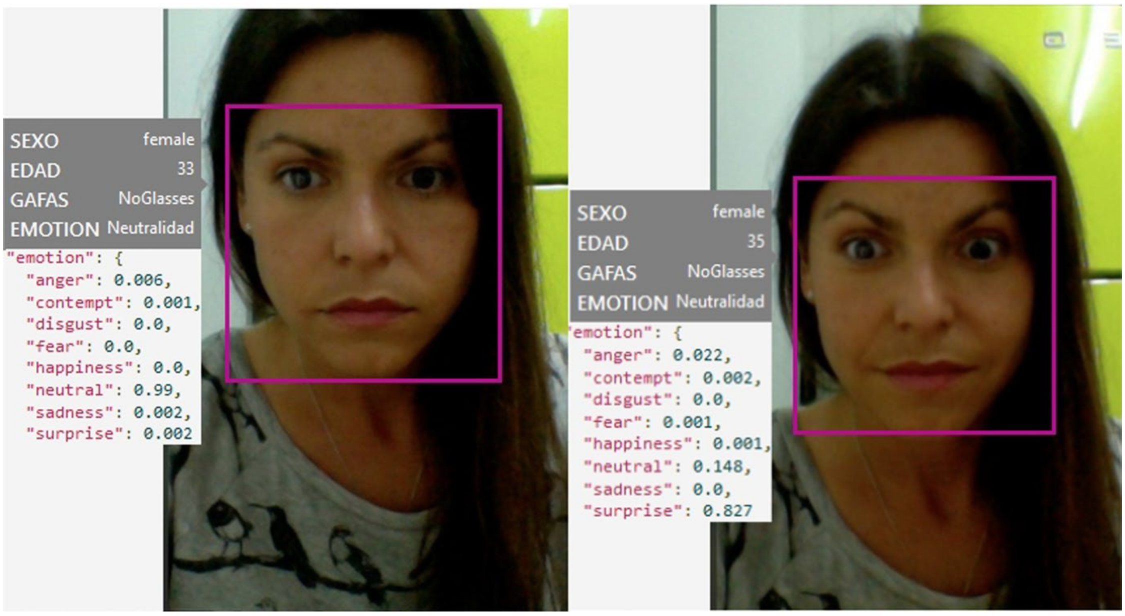


**Fig. 2** Examples of facial emotion recognition and classification (the left-hand column shows the emotion trend)

*Sources* (e.g. drawing and text) is the number of times the learner enlarged the image associated with the content being studied for at least fifteen seconds, e.g. Spend time studying about the heart and open the associated image. PKAtotalFreq: *Prior Knowledge Activation* is the number of times that the learner, after navigating to previously unvisited content, writes their prior knowledge about the new content. A correlate for when the student searches in their memory for relevant prior knowledge either before beginning task performance or during task performance., e.g. The student opens a page and, before reading, writes everything they already know about the topic on that page.

### Emotions from face recording video

During the learning session a video of the participants face was recorded using a web cam which was subsequently analyzed using a desktop app. Each participant's full session was recorded, the webcam on the computer was adjusted to the participant's position at the beginning and they were asked to sit facing forward and be as neutral as possible, although their facial expressions were expected to vary during the session. We asked participants to tie their hair back, make sure there was nothing around their neck, remove their glasses, and remove chewing gum if necessary to have the best conditions for the recording.

The learning session videos were analyzed using Microsoft Emotion API (2019 Automatic Facial Recognition Software). The API classifies facial expression in eight emotion classes: anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. These emotions are understood to be cross-culturally and universally communicated with specific facial expressions (Arora et al., 2018). We developed our specific application to use Microsoft Emotion API in local mode (see Fig. 2). Participants tended to experience all of emotions the system detects during the

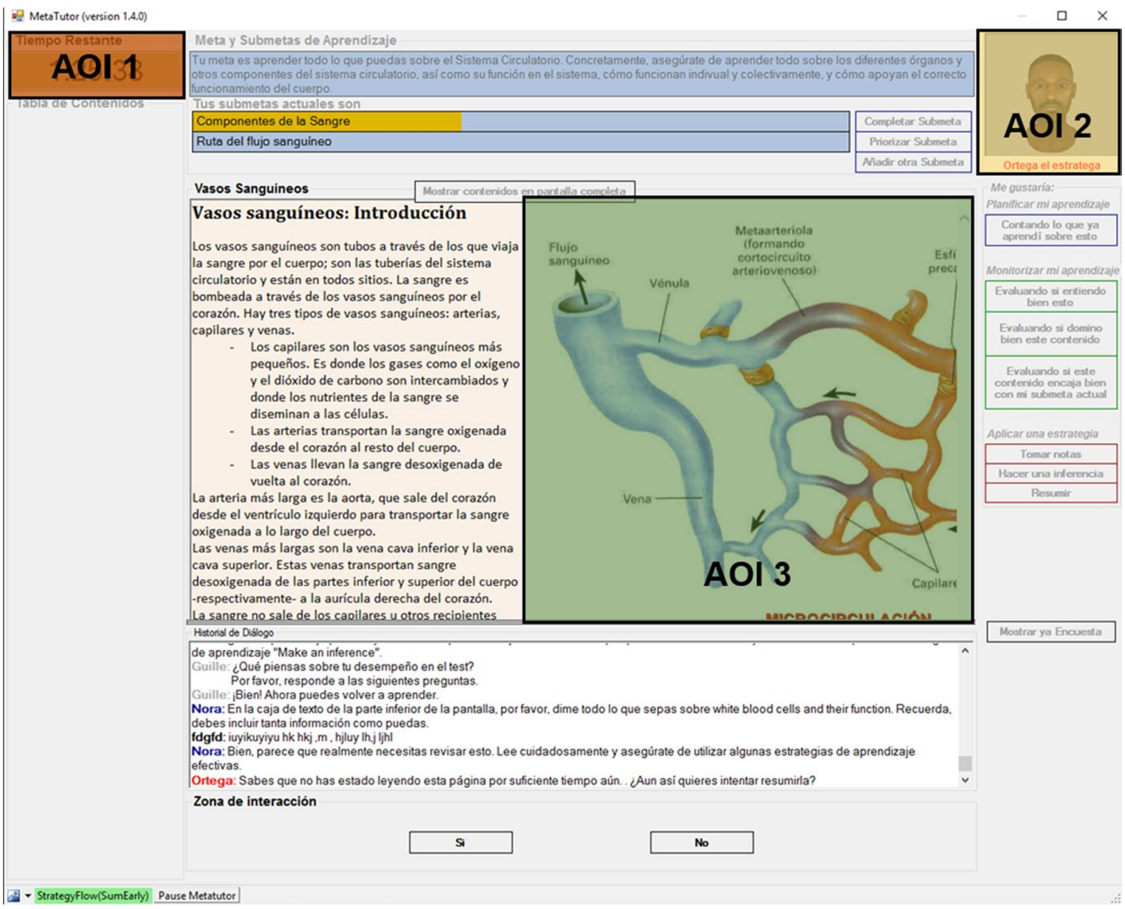


**Fig. 3** Map of areas of interest (AOIs) in the ITS

session, but we were able to produce a general index for each participant giving information about the general pattern. The analysis gave us at least one predominant emotion during the learning session from frame of student video, and there were a large number of frames (1 frame per second) for each student in every session. The confidence (values between 0 and 1) gives the likelihood for each class of emotion.

## Interaction zones from eye tracking

Data from each learner was collected throughout the session using the screen-based eye tracker RED500 (https://imotions.com/hardware/smi-red500/). We used SMI's BeGaze software in order to process the fixations on the learning environment AOIs. BeGaze performs the calculation automatically, identifying a fixation if a learner stares at an AOI for at least 80 ms with a maximum dispersion of 100px.

For the present study, we extracted three variables related to learner fixation on three AOIs (See Fig. 3). AOI1 The learning session timer (number of times the learner focused their attention on the area showing the time left in the learning session), which may denote time management or resource management strategies, while reduced or excessive fixations on AOI1 might indicate poor time management skills. AOI2 ITS agent/avatar (number of times the learner focused their attention on the area where the pedagogical agents appear). This variable may show that the participant is taking advantage of the prompts and feedback provided by the agents during the interaction in response to participants' goals, behaviors, self-evaluations, and progress. However, it must be considered carefully, because learners may not always need to look at an agent to process their audio prompts and feedback (Bondareva et al., 2013; Lallé et al., 2021). AOI3 Images/graphics supporting content (number of times the learner focused their attention on the area covered by the images related to the learning session contents). This variable may indicate integration contributing to information processing (Mason et al., 2013).

### Final grade from test/quiz

During the session and at the end of the session, each subject was tested about the learning content, giving a final performance value between 0 and 10, with 10 being the highest performance. There was a pretest about prior knowledge of the content at the beginning of the session, and a multiple-choice posttest of domain knowledge that was corrected based on pretest.

**Preprocessing data** We preprocessed all of the data in the aforementioned Excel files (Romero et al., 2014). Firstly, the data were anonymized, then the input attributes were normalized/rescaled, the output attributes and input attributes were discretized, and finally the format was transformed.

### Anonymizing

Student anonymity and privacy was maintained but the information in the four Excel files was linked to the same subject using anonymized coding. We implemented a basic solution, using a randomly generated number as a user ID rather than the users' names, and replaced the students' names with the ID in the four Excel files.

### Normalizing

We adjusted all of the input values, which used different scales, to a single common scale. This was necessary because the original values had a variety of ranges. Normalization is a data transformation where the attribute values are scaled so as to fall within a specified range, such as −1.0 to 1.0, or 0.0 to 1.0. Normalization helps to prevent attributes with large ranges from outweighing attributes with smaller ranges. In this case we rescaled/normalized all of the input attribute values to the same range [0–1] by using the well-known Min–Max method, which is a linear transformation of the original data using the formula: $Z_i = X_i – min(X)/max(X) – min(X)$, where $X = (x_1,\ldots,x_n)$ and $Z_i$ is now the *ith* normalized data.

### Discretizing

Discretization divides numerical data into categorical classes that are more user-friendly than precise magnitudes and ranges. It reduces the number of possible values of the continuous feature and provides a view of the data that is easier to understand. Generally, discretization smooths out the effect of noise and enables simpler models, which are less prone to overfitting. We discretized all the input attributes in order to have the same variables in both numerical and categorical formats. To do that, we used equal-width binning with the following 3 bins: *LOW*, *MEDIUM* and *HIGH*. Equal-width binning divides the range of possible values into *N* sub-ranges of the same size in which: *bin_width = (max value–min value)/N.*

We also discretized the output attribute or class to predict (the students' final performance or status). We used a manual discretization with the user directly specifying cut-off points. In our case, the class had the following 2 values and cut-off points:

- PASS: Students who scored 5 out of 10 or better in the performance tests. In our case, this was 21 out of 40 students (52.50%).
- FAIL: Students who scored less than 5 out of 10 in the performance tests. In our case, this was 19 out of 40 students (47.50%).

### Transforming

Finally, we converted the files from Excel to CSV (Comma-separated values) files. CSV is a delimited text file that uses a comma to separate values. Each line of the file is a data record. Each record consists of one or more fields, separated by commas. We transformed each of the two versions of the four Excel files (numerical and categorical values) into two CSV files because they can be directly opened and used by the WEKA data mining framework that we used in the experiments. We used the WEKA (Waikato Environment for Knowledge Analysis) data mining framework (Witten et al., 2011) to predict student performance. WEKA provides a collection of algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to these functions.

## Experiments

We carried out three different experiments using three different approaches and six classification algorithms with the preprocessed numerical and discretized data to predict student performance in the ITS (See Fig. 4).

We used two types of white box classification models: Rule induction algorithms and decision trees. The models produced by these algorithms (IF–THEN rules from decision trees) are simple and clear, and so are easy for humans to understand. IF–THEN classification rules provide a high-level knowledge representation that is used for decision making, while decision trees can also be converted into a set of IF–THEN classification rules. In our experiments, we selected six well-known classification algorithms integrated in the WEKA data mining tool (Witten et al., 2011): three decision tree algorithms (J48, REPTree and RandomTree) and three rule induction algorithms (JRip, Nnge and PART). We executed these algorithms using a k-fold cross-validation (k = 10) and Accuracy and Area under the ROC curve as evaluation metrics for classification:

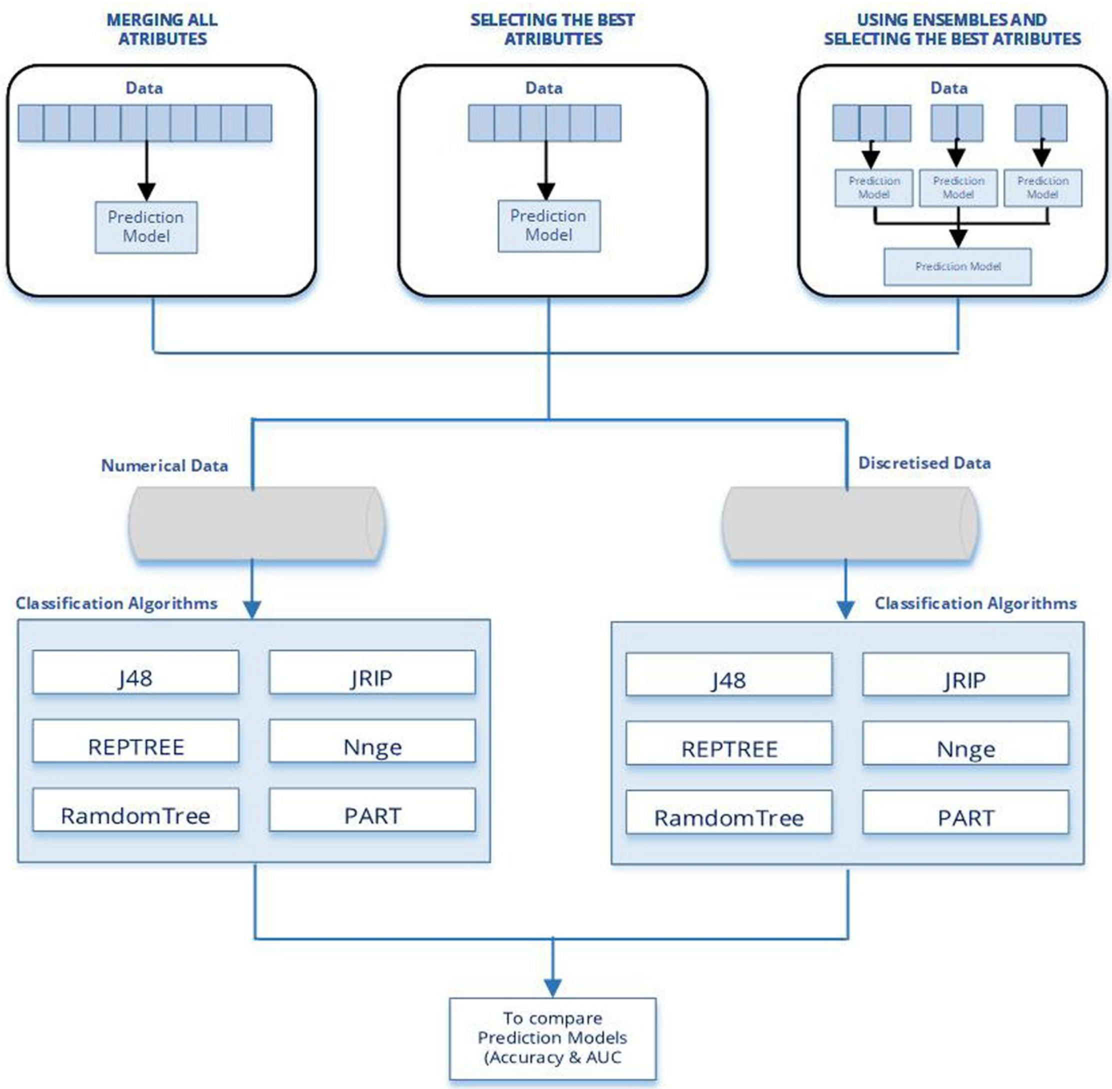


**Fig. 4** Visual description of the experiments

- **Accuracy (ACC)** is the most commonly-used traditional method for evaluating classification algorithms. It provides a single-number summary of performance. In our case, it is obtained by the equation: $\text{Acc} = \frac{\text{Number of students correctly classified}}{\text{Total number of students}}$. This metric shows the percentage of correctly classified students.
- **Area under the ROC curve (AUC)** measures the two-dimensional area underneath the entire Relative Operating Characteristic (ROC) curve. The ROC curve allows us to find possibly optimal models and discard suboptimal ones. AUC is often used when the goal of classification is to obtain a ranking because ROC curve construction requires a ranking to be produced.

## Experiment 1: merging all attributes

In experiment 1 we applied the classification algorithms to a single file with all the attributes of the three different data sources merged. We created two different numerical and discrete/categorical CSV files. Each dataset had fifteen input

**Table 1** Results produced by merging all attributes

| | Numerical data | | Discretized data | |
|---|---|---|---|---|
| | %Accuracy | AUC | %Accuracy | AUC |
| Jrip | 72.50 | 0.69 | 72.50 | 0.65 |
| Nnge | 62.50 | 0.61 | 62.50 | 0.62 |
| PART | 80.00 | 0.79 | 67.50 | 0.69 |
| J48 | 80.00 | 0.80 | 70.00 | 0.67 |
| REPTree | 72.50 | 0.74 | 67.50 | 0.61 |
| Randomtree | 70.00 | 0.70 | 72.50 | 0.69 |
| Avg | 73.33 | 0.72 | 68.75 | 0.66 |

**Table 2** Results of the attribute selection with CLASSIFIERSUBSETEVAL

| Dataset | # selected features | Name of selected features |
|---|---|---|
| Numerical | 2 | Metatutor.SummAll<br>Metatutor.COIStotalFreq |
| Discretized | 5 | Metatutor.SummAll<br>Interaction.AOI1FixCount<br>Interaction.AOI3FixCount<br>Emotion.anger<br>Emotion.happiness |

attributes (in numerical or discrete format) and only one output attribute or class. Finally, we executed six classification algorithms on the two summary datasets, producing the results (%Accuracy and ROC Area) shown in Table 1.

Table 1 shows that the best results (highest values) were produced by Part (80.0%Acc) and J48 (80.00%Acc and 0.80 AUC) algorithms with numerical data. In fact, on average, most of the algorithms exhibited slightly improved performance in both measures when using numerical data.

### Experiment 2: selecting the best attributes

In Experiment 2, we applied the classification algorithms to a single file with only the best attributes. Firstly, we applied attribute selection algorithms to the summary files from the Experiment 1 in order to eliminate redundant or irrelevant attributes. We used the well-known *CfsSubsetEval* (Correlation-based Featured Selection) method provided by the WEKA tool. This method selects the features that are more strongly correlated with the class. Starting from our initial 15 input attributes, we produced two sets of 2 optimal attributes for the numerical datasets and 5 optimal attributes (see Table 2) for the discretized datasets.

Following that, we executed the six classification algorithms with the two new summary datasets, producing the results (%Accuracy and ROC Area) shown in Table 3.

**Table 3** Results obtained when selecting the best attributes

| | Numerical data | | Discretized data | |
|---|---|---|---|---|
| | %Accuracy | AUC | %Accuracy | AUC |
| Jrip | 77.50 | 0.81 | 77.50 | 0.68 |
| Nnge | 80.00 | 0.80 | 75.00 | 0.75 |
| PART | 77.50 | 0.77 | 70.00 | 0.67 |
| J48 | 77.50 | 0.80 | 77.50 | 0.76 |
| REPTree | 80.00 | 0.78 | 70.00 | 0.63 |
| Randomtree | 82.50 | 0.82 | 75.00 | 0.77 |

**Table 4** Results of attribute selection with CFSSubsetEval

| Dataset | Type | # selected features | Name of selected features |
|---|---|---|---|
| Metatutor | Numerical | 1 | Metatutor.SummAll |
| | Discretized | 1 | Metatutor.SummAll |
| Interaction | Numerical | 1 | Interaction.AOI6FixCount |
| | Discretized | 2 | Interaction.AOI6FixCount<br>Interaction.AOI1FixCount |
| Emotion | Numerical | 1 | Emotion.surprise |
| | Discretized | 1 | Emotion.fear |

Table 3 shows that the best results (highest values) were produced by Randomtree (82.50% Acc and 0.82 AUC) algorithms. Again, on average most of the algorithms exhibited slightly improved performance in both measures when using numerical data.

### Experiment 3: using ensembles and selecting the best attributes

In Experiment 3 we applied an ensemble of classification algorithms to the best attributes from each different data source. Firstly, we selected the best attributes for each of the three different datasets, again using the well-known *CfsSubsetEval* attribute selection algorithm. This gave the list of attributes shown in Table 4.

Following that, we applied an ensemble or combination of multiple classification base models by using the well-known Vote (Kuncheva, 2014) for automatic combining several machine learning algorithms provided by WEKA. Vote combines the probability distributions of these base learners. It produces better results than individual classification models, if the set classifiers are accurate and diverse. It has demonstrated better results than homogeneous models for standard datasets.

We executed the six classification algorithms as base or individual classification models of our Vote method with the previously described numerical and discretized datasets. Table 5 shows the results (%Accuracy and ROC Area).

**Table 5** Results from using ensembles and selecting the best attributes

| | Numerical data | | Discretized data | |
|---|---|---|---|---|
| | %Accuracy | AUC | %Accuracy | AUC |
| Jrip | 82.50 | 0.88 | 82.50 | 0.86 |
| Nnge | 80.00 | 0.87 | 65.00 | 0.66 |
| PART | 80.00 | 0.84 | 75.00 | 0.78 |
| J48 | 82.50 | 0.86 | 80.00 | 0.84 |
| REPTree | 87.50 | 0.88 | 80.00 | 0.82 |
| Randomtree | 82.50 | 0.88 | 75.00 | 0.74 |

**Table 6** Average results from the three data fusion approaches

| Average | Numerical data | | Discretized data | |
|---|---|---|---|---|
| | %Accuracy | AUC | %Accuracy | AUC |
| Merging all attributes | 73.33 | 0.72 | 68.75 | 0.66 |
| Selecting the best attributes | 79.16 | 0.80 | 74.16 | 0.71 |
| Using ensembles and selection of the best attributes | 82.50 | 0.87 | 76.25 | 0.78 |

Table 5 shows that the best results (highest values) were produced by REPTree (87.50%Acc and 0.88 AUC). On average, most of the algorithms again exhibited slightly improved performance in both measures when using numerical data.

## Discussion

Below, we address the two initial research questions by discussing the results from our four experiments.

*Question 1*

*Can attribute selection and classification ensemble algorithms improve the prediction results of student final performance from our ITS data?*

We used three different data fusion approaches and six white-box classification algorithms to answer this question. Table 6 shows that the average prediction performance (Average of % Accuracy and AUC) of the classification algorithms increased with each new approach.

We first applied a traditional approach for merging all the attributes from the different data sources directly. This initial approach gave reasonable results (accuracy higher than 70% and AUC higher that 0.7) from numerical data. Our second approach selected the best attributes for each dataset. This was an improvement on the first approach (79% accuracy and 0.8 AUC). Finally, the third approach improved on the second approach and gave the best result by using ensembles

**Table 7** J48 decision tree produced when merging all attributes

| |
|---|
| If Metatutor.SummAll > 0.25 Then PASS |
| If Metatutor.SummAll < = 0.25 AND Emotions.surprise < = 0.061227 Then FAIL |
| If Emotions.surprise > 0.06 AND Interaction.AOI3FixCount < = 0.04 Then PASS |
| Else FAIL |
| Number of Rules: 4 |

and selection of the best attributes (82% accuracy and 0.87 AUC). In all the approaches the average values were higher when using numerical than discretized data.

However, we were unable to find a single best algorithm that would win in all cases in our experiments. This is logical and in line with the No-Free-Lunch theorem (Wolpert, 2002), in which it is generally accepted that no single supervised learning algorithm can beat another algorithm over all possible learning problems or different datasets. In the first experiment, the algorithm that produced the highest prediction values was J48 (80.00%Acc and 0.80 AUC), in the second experiment it was Randomtree (82.50%Acc and 0.82 AUC), and REPTREE produced the highest prediction values of %Acc (87.50) and AUC (0.88) when using an ensemble and selection of the best attributes from the discretized data in the fourth experiment.

*Question 2*

*How useful are the models produced and what are the best variables to help teachers understand how to predict students' final performance in the ITS?*

To answer this question, we will demonstrate and describe the meaning of the prediction model that produced the highest values of Accuracy and AUC in each of our 3 experiments.

In experiment 1, the prediction model producing the best prediction was produced by the J48 algorithm using discretized data (see Table 7).

This prediction model (see Table 7) has 4 rules. The first rule shows that the students who have scores higher than 0.25 in SummAll in MetaTutorES PASS the course. The second rule shows that if students have a score lower than 0.25 in SummAll in MetaTutorES and a surprise emotion lower than 0.06, then they FAIL the course. The third rule shows that if students have a surprise emotion higher than 0.06 and a value of AOI2FixCount lower than 0.04 in the pedagogical agent zone, then they PASS the course. Finally, the remaining students are classified as FAIL.

In experiment 2, the prediction model that produced the highest prediction values used the Randomtree algorithm with numerical data (see Table 8).

This prediction model (see Table 8) consists of 7 IF–THEN rules. In all these rules, the two most frequent attributes are the summary strategies (SummAll) and the frequency of use of the user coordination of information sources strategy (COIS-totalFreq). It is also important to note that in this model the predictions of students passing or failing was not influenced by any emotions or interaction zones.

**Table 8** Randomtree pruned tree produced when selecting the best attributes

If Metatutor.SummAll < 0.28
| Metatutor.COIStotalFreq < 0.04 Then Pass
| IF Metatutor.COIStotalFreq > = 0.04
| | IF Metatutor.SummAll < 0.03
| | | Metatutor.COIStotalFreq < 0.56 Then Fail
| | | IF COIStotalFreq > = 0.56
| | | | IF COIStotalFreq < 0.66
| | | | | Metatutor.COIStotalFreq < 0.59 Then Pass
| | | | | Metatutor.COIStotalFreq > = 0.59 Then Pass
| | | | Else Metatutor.COIStotalFreq > = 0.66 Then Fail
| | Else IF Metatutor.SummAll > = 0.03
| | | Metatutor.SummAll < 0.16 Then Pass
| | | Metatutor.SummAll > = 0.16 Then Fail
Else Metatutor.SummAll > = 0.28: Pass
Size of the tree: 15

**Table 9** RepTree decision trees produced using ensembles with selecting the best attributes

REPTree (Metatutor)
============
If Metatutor SummAll > = 0.03 Then Pass
Else Fail
Size of the tree: 3
REPTree (Interaction)
============
If Interaction.AOI3FixCount > = 0.29 Then Pass
Else Fail
Size of the tree: 3
REPTree (Emotion)
============
If Emotion.surprise > = 0.05 Then Pass
Else Fail
Size of the tree: 3

In experiment 3, the prediction model that produced the highest prediction values used the RepTree algorithm with numerical data (see Table 9).

This prediction model (see Table 9) is a combination of three different models showing that the behavior of students in relation to the frequency of the summary strategies, the proportion of fixations on AOI3 *Images/graphics supporting content* over the total session, and the surprise emotion are the most important attributes in predicting whether students PASS or FAIL. Students who interact with the ITS with a value higher than 0.03 in the SummAll variable, students who have a proportion of fixations on AOI3 over the total session higher than 0.29, and students who have an emotion of surprise higher than 0.05, are predicted to PASS the course, in other cases they are predicted to FAIL the course.

These results are not surprising considering that Summarizing and Content Coordination of Information Sources are classical strategies that contribute to students taking a strategic approach (Cerezo et al., 2020a,b), and positive emotions such as surprise, enjoyment and happiness are thought to promote motivation, facilitating use of flexible learning strategies, and supporting self-regulation of learning (Pekrun et al., 2011), all of which presumably promote better performance.

## Conclusions

This paper proposes the use of ensembles and attribute selection for improving the prediction of students' performance from multimodal data in an ITS. We collected and preprocessed data from 40 first-year university students from three different sources: learning strategies from MetaTutorES logs, emotions from face recording videos, and interaction zones from gaze data, along with marks from performance test about the learning content. We carried out 3 experiments in order to answer two research questions:

- Can attribute selection and classification ensemble algorithms improve the prediction of student final performance from our ITS data? Yes, the use of ensembles and selecting the best attributes approach from numerical data produced the best results in terms of Accuracy and AUC values. The REPTree classification algorithm produced the best results.
- How useful are the models produced and what are the best variables to help teachers understand how to predict students' final performance in the ITS? The white-box models we produced give teachers understandable explanations (IF–THEN rules) of how they arrived at their classifications of student performance. They showed that the attributes that appeared most in these rules were logs denoting use of *Summarizing* strategies and *Coordination of Information Sources* (SummAll and COIStotalFreq) from the ITS logs, paying attention to avatars and to images/graphics supporting text content (AOI2 and AOI3) from gaze data, and surprise from emotions.

The implications of the current study point to Web ITS and Web-based Adaptive Educational Systems. If data is captured from different data sources, the classifier ensemble methodology proposed in this study could make better, earlier performance predictions than the single data source models that are commonly used at present.

As the next step, we intend to investigate and perform new experiments with the aim of improving our results and in order to overcome some limitations:

- Adding additional different variables/attributes from the multimodal student interaction with the ITS such as think aloud data, self-report data, and/or physiological measures. In the context of multimodal data, classical self-report methodology remains valuable. Aspects such as achievement emotions experienced by students, students' learning goals and approaches, self-esteem, and epistemological beliefs may help to improve the prediction results. For instance, previous studies have shown that visual metrics (e.g., fixation rate, longest fixations) are significantly influenced by students' goals, so this could be applied to ITS design so that it adapts better to students' learning goals (Lallé et al., 2017). As well as this, using EEG (Electroencephalography), ECG (Electrocardiogram), EMG (Electromyography), EDA (Electrodermal Activity), sitting posture, etc. in order to produce more accurate values for predicting students' performance.

- Taking into account that there is recent evidence that emotions co-occur during learning in MetaTutor (Lallé et al., 2021), it should be considered for future research; the emotions in ITS are often studied as single affective state, like in the present work.
- We would also like to use additional classifier algorithms, particularly deep learning, which could perform significantly better than classic methods.
- Using raw data and other specific data fusion techniques. We used a basic fusion method that uses summary data. However, there are other data fusion theories and methods such as Probability-based methods (PBM) and Evidence reasoning methods (EBM) that we can use with raw data. We could also use semantic (abstract) level features in order to produce intelligent data aggregation.
- We are also aware of the limited generalizability of the results. The next step would be applying the current proposal in other learning systems such as Learning Management Systems (LMSs) or Personal Learning Environments (PLEs). This would allow us to compare results in different learning contexts and with a greater diversity of subjects.

**Acknowledgements** The authors acknowledge the financial subsidy provided by the Spanish Ministry of Science and Innovation "Deteccion Temprana e Intervencion en Dificultades del Aprendizaje Específicas desde el Modelo RtI" (PID2019-107201GB-I009), and the Spanish Ministry of Science and Innovation in the project "Improving Data Science User's Experience with Computational Intelligence (INTENSE)" (PID2020-115832GB-I00).

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

## References


Adejo, O. W., & Connolly, T. (2018). Predicting student academic performance using multi-model heterogeneous ensemble approach. *Applied Research in Higher Education, 10*(1), 61–75. https://doi.org/10.1108/JARHE-09-2017-0113

Antonietti, A., Colombo, B., & Di Nuzzo, C. (2015). Metacognition in self-regulated multimedia learning: Integrating behavioural, psychophysiological and introspective measures. *Learning, Media and Technology, 40*(2), 187–209. https://doi.org/10.1080/17439884.2014.933112

API de reconocimiento facial. (2019). https://azure.microsoft.com/es-es/services/cognitiveservices/face/#overview

Arora, R., Sharma, J., Mali, U., Sharma, A., & Raina, P. (2018). Microsoft cognitive services. *International Journal of Engineering Science and Computing, 8*(4), 17323. https://ijesc.org/upload/892ad2fdbc5ade1ebe57cefd6e2a9fff.Microsoft%20Cognitive%20Services.pdf.

Azevedo, R. (2009). Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: A discussion. *Metacognition and Learning, 4*(1), 87–95. https://doi.org/10.1007/s11409-009-9035-7

Azevedo, R., Bouchet, F., Harley, J. M., Feyzi-Behnagh, R., Trevors, G., Duffy, M., Taub, M., Pacampara, N., Agnew, L., & Griscom, S. (2011). MetaTutor: An intelligent multi-agent tutoring system designed to detect, track, model, and foster self-regulated learning. *Proceedings of the Fourth Workshop on Self-Regulated Learning in Educational Technologies*. SRL&ET. Doi: https://doi.org/10.13140/RG.2.1.1334.6640.

Azevedo, R., Taub, M., & Mudrick, N. V. (2018). Using multi-channel trace data to infer and foster self-regulated learning between humans and advanced learning technologies. In D. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 254–270). New York, NY: Routledge.

Blanchard, E. G., Volfson, B., Hong, Y. J., & Lajoie, S. P. (2009). Affective artificial intelligence in education: From detection to adaptation. *Frontiers in Artificial Intelligence and Applications, 200*(1), 81–88. https://doi.org/10.3233/978-1-60750-028-5-81

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics, 3*(2), 220–238. https://doi.org/10.18608/jla.2016.32.11

Bogarín, A., Cerezo, R., & Romero, C. (2018). Discovering learning processes using inductive miner: A case study with learning management systems (LMSs). *Psicothema, 30*(3), 322–329. https://doi.org/10.7334/psicothema2018.116

Bondareva, D., Conati, C., Feyzi-Behnagh, R., Harley, J. M., Azevedo, R., & Bouchet, F. (2013). Inferring learning from gaze data during interaction with an environment to support self-regulated learning. In *International Conference on Artificial Intelligence in Education* (pp. 229–238). Springer, Berlin, Heidelberg. Doi: https://doi.org/10.1007/978-3-642-39112-5_24.

Cerezo, R., Sánchez-Santillán, M., Paule-Ruiz, M. P., & Núñez, J. C. (2016). Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. *Computers and Education, 96*, 42–54. https://doi.org/10.1016/j.compedu.2016.02.006

Cerezo, R., Esteban, M., Vallejo, G., Sanchez-Santillan, M., & Nuñez, J. C. (2020a). Differential efficacy of an intelligent tutoring system for university students: A case study with learning disabilities. *Sustainability, 12*(21), 1–17. https://doi.org/10.3390/su12219184

Cerezo, R., Fernández, E., Gómez, C., Sánchez-Santillán, M., Taub, M., & Azevedo, R. (2020b). Multimodal protocol for assessing metacognition and self-regulation in adults with learning difficulties. *Journal of Visualized Experiments*. https://doi.org/10.3791/60331

Chan, M. C. E., Ochoa, X., & Clarke, D. (2020). Multimodal learning analytics in a laboratory classroom. *Intelligent Systems Reference Library, 158*, 131–156. https://doi.org/10.1007/978-3-030-13743-4_8

Chango, W., Cerezo, R., & Romero, C. (2021). Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses. *Computers and Electrical Engineering, 89*, 1–13. https://doi.org/10.1016/j.compeleceng.2020.106908

Chango, W., Sánchez-Santillán, M., Cerezo, R., & Romero, C. (2020). Predicting students' performance using emotion detection from face-recording video when interacting with an ITS [Conference presentation]. Educational data mining conference 2020. https://educationaldatamining.org/files/conferences/EDM2020/papers/paper_232.pdf.

Chen, C. M., & Wang, H. P. (2011). Using emotion recognition technology to assess the effects of different multimedia materials on learning emotion and performance. *Library and Information Science Research, 33*(3), 244–255. https://doi.org/10.1016/j.lisr.2010.09.010

Crescenzi-Lanna, L. (2020). Multimodal learning analytics research with young children: A systematic review. *British Journal of Educational Technology, 51*(5), 1485–1504. https://doi.org/10.1111/bjet.12959

D'Mello, S., & Kory, J. (2012). Consistent but modest: A meta-analysis on unimodal and multimodal affect detection accuracies from 30 studies. *ICMI'12—Proceedings of the ACM International Conference on Multimodal Interaction*. Association for Computing Machinery. Doi: https://doi.org/10.1145/2388676.2388686.

Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist, 45*(4), 203–209. https://doi.org/10.1080/00461520.2010.515935

Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior, 48*, 615–625. https://doi.org/10.1016/j.chb.2015.02.013

Jesus, J., Araújo, D., & Canuto, A. (2016). Fusion approaches of feature selection algorithms for classification problems [Conference presentation]. The 5th Brazilian conference on intelligent systems (BRACIS), Recife, Brazil. (pp. 379–384). IEEE. Doi: https://doi.org/10.1109/BRACIS.2016.075.

Kuncheva, L. I. (2014). *Combining pattern classifiers*. Wiley Online Library. https://doi.org/10.1002/9781118914564

Lallé, S., Taub, M., Mudrick, N. V., Conati, C., Azevedo, R. (2017). The impact of student individual differences and visual attention to pedagogical agents during learning with MetaTutor. In E. André, R. Baker, X. Hu, M. Rodrigo, B. du Boulay (Eds.), *Artificial Intelligence in Education. AIED 2017. Lecture Notes in Computer Science*. Springer, Cham. Doi: https://doi.org/10.1007/978-3-319-61425-0_13.

Lallé, S., Murali, R., Conati, C., Azevedo, R. (2021). Predicting co-occurring emotions from eye-Tracking and interaction data in MetaTutor. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, V. Dimitrova (Eds.), *Artificial Intelligence in Education. AIED 2021. Lecture Notes in Computer Science* (pp. 241–254). Springer, Cham. Doi: https://doi.org/10.1007/978-3-030-78292-4_20.

Lerche, T., & Kiel, E. (2018). Predicting student achievement in learning management systems by log data analysis. *Computers in Human Behavior, 89*, 367–372. https://doi.org/10.1016/j.chb.2018.06.015

Li, L. Y., & Tsai, C. C. (2017). Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance. *Computers and Education, 114*, 286–297. https://doi.org/10.1016/j.compedu.2017.07.007

Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education*, *60*(1), 95–109.

Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: When measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research, 60*(6), 723–735. https://doi.org/10.1080/00313831.2015.1066436

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The achievement emotions questionnaire (AEQ). *Contemporary Educational Psychology, 36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10*(3), e1355. https://doi.org/10.1002/widm.1355

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers and Education, 51*(1), 368–384. https://doi.org/10.1016/j.compedu.2007.05.016

Romero, C., Romero, J. R., & Ventura, S. (2014). *A survey on pre-processing educational data. Studies in computational intelligence* (pp. 29–64). Springer Verlag. https://doi.org/10.1007/978-3-319-02738-8_2

Taub, M., & Azevedo, R. (2019). How does prior knowledge influence fixations on and sequences of cognitive and metacognitive SRL processes during learning with an ITS? *International Journal of Artificial Intelligence in Education, 29*, 1–28.

Taub, M., Azevedo, R., Rajendran, R., Cloude, E., Biswas, G., & Price, M. (2021). How are students' emotions related to the accuracy of cognitive and metacognitive processes during learning with an intelligent tutoring system? *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2019.04.001

Winne, P. H., & Azevedo, R. (2021). Metacognition. In K. Sawyer (Ed.) Cambridge handbook of the learning sciences (3rd edn.). Cambridge, MA: Cambridge University Press.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (2nd ed.). Morgan Kaufmann.

Wolpert, D. H. (2002). The supervised learning no-free-lunch theorems. *Soft Computing and Industry*. https://doi.org/10.1007/978-1-4471-0123-9_3

Zhou, Q., Zheng, Y., & Mou, C. (2015). Predicting students' performance of an offline course from their online behaviors. *Proceedings of the Fifth International Conference on Digital Information and Communication Technology and its Applications (DICTAP)* (pp. 70–73). Doi: https://doi.org/10.1109/DICTAP.2015.7113173.



## Authors and Affiliations

**Wilson Chango[1] · Rebeca Cerezo[2] · Miguel Sanchez-Santillan[2] · Roger Azevedo[3] · Cristóbal Romero[4]**

Wilson Chango
wilson.chango@pucese.edu.ec

Miguel Sanchez-Santillan
sanchezsmiguel@uniovi.es

Roger Azevedo
roger.azevedo@ucf.edu

Cristóbal Romero
cromero@uco.es

1 Pontifical Catholic University of Ecuador, Quito, Ecuador

2 Developmental Psychology, University of Oviedo, Room 231 Plaza Feijoo s/n, Oviedo, Asturias, Spain

3 University of Central Florida, Orlando, USA

4 University of Córdoba, Córdoba, Spain