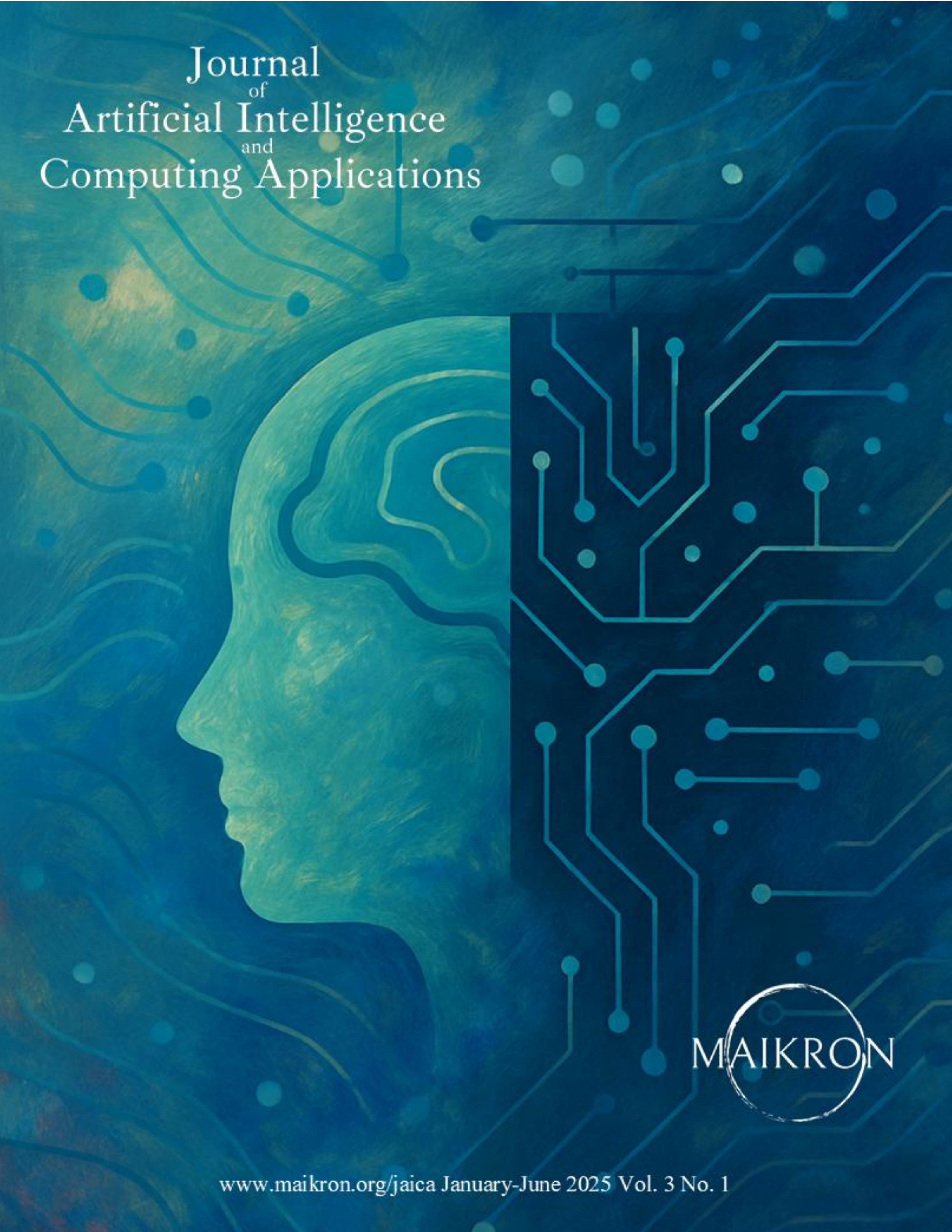
# Journal of Artificial Intelligence and Computing Applications



MAIKRON

www.maikron.org/jaica January-June 2025 Vol. 3 No. 1

**Journal** of **Artificial Intelligence** and **Computing Applications**

Journal of Artificial Intelligence and Computing Applications, Vol. 3, Núm. 1, enero – junio de 2025, es una publicación semestral editada por la Asociación para el Avance de las Aplicaciones Inteligentes y Tecnologías con Impacto Social, Calle 12 No. 103, Lotes No. 56 y 57, Ucú, Yucatán, C.P. 97357, Tel. (55) 28 55 70 20. URL: https://www.maikron.org/jaica, correo electrónico: contact@maikron.org. Editor responsable: Mauricio Gabriel Orozco del Castillo. Certificado de Reserva de Derechos al uso Exclusivo del Título: 04-2025-073112454300-102, eISSN: en trámite, por el Instituto Nacional del Derecho de Autor (INDAUTOR). Responsable de la última actualización de este número, Mauricio Gabriel Orozco del Castillo, Editor Responsable. Fecha de la última modificación: 30 de junio de 2025, Calle 12 No. 103, Lotes No. 56 y 57, C.P. 97357, Ucú, Yucatán.

**Journal of Artificial Intelligence and Computing Applications**
**Vol. 3, No. 1, January-June 2025**

# Contents

# Journal of Artificial Intelligence and Computing Applications

MAIKRON

# Foreword

**Dear Readers, Contributors, and Colleagues,**

It is with great pleasure that I introduce Volume 3, Issue 1 of the Journal of Artificial Intelligence and Computing Applications (JAICA). This regular issue marks the beginning of a new volume and continues our mission to publish high-quality research that explores the intersection of artificial intelligence with practical and socially meaningful domains.

A notable feature of this issue is the publication of several original research articles that originated from the International Conference on Artificial Intelligence for Mental Health (ICAIMH 2025). These papers reflect the growing role of AI in addressing pressing challenges in mental health and education—ranging from cognitive screening in dementia and the cognitive stimulation of elderly populations through serious games, to data-driven analyses of sleep quality, academic burnout, and the evolving scientific landscape of AI in mental health. Collectively, they demonstrate the diversity of methodologies and applications emerging from this rapidly developing area.

These articles are the result of a collaborative effort between JAICA and ICAIMH, and I would like to express my appreciation to both the journal's editorial reviewers and the ICAIMH 2025 program committee for their rigorous evaluations and insightful feedback during the review process. This partnership exemplifies the kind of interdisciplinary and interinstitutional collaboration we seek to foster—one that bridges communities and elevates the quality and reach of applied AI research.

In addition to the ICAIMH contributions, this issue also features independent submissions that expand our understanding of AI's transformative potential in a variety of contexts, further enriching the discourse within our pages.

I extend my sincere thanks to the authors for their contributions, to the reviewers for their thoughtful and meticulous evaluations, and to our sponsor Maikron for their continued support in advancing JAICA's vision.

As we look ahead, JAICA remains committed to providing a platform for researchers, practitioners, and innovators who are working to apply artificial intelligence to the benefit of society.

Warmest regards,

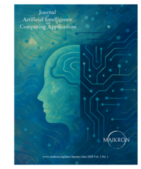
**Mauricio G. Orozco-del-Castillo**

Editor-in-Chief

Journal of Artificial Intelligence and Computing Applications (JAICA)

Journal of Artificial Intelligence and Computing Applications (2025) 3(1): 1–12

Journal of Artificial Intelligence and Computing Applications



*Research article*

# Analysis of the sleep quality of college students from different knowledge areas using a data mining approach

**Andrea Morales-Robles**[1,*], **Víctor Menéndez-Domínguez**[1], **and Héctor Rubio-Zapata**[1]

[1]Universidad Autónoma de Yucatan

**ABSTRACT**

Sleep is an essential physiological process involved in memory consolidation, metabolic and endocrine homeostasis, and immune system regulation. Therefore, good sleep quality is vital for maintaining physiological homeostasis. Poor sleep quality is prevalent among college students and affects both physical and mental health. Conventional statistical methods, such as logistic regression, are commonly used to generate predictive models of sleep quality and have been extensively applied to Health Sciences students, but their use has been less studied among students from other disciplines, such as Engineering and Exact Sciences. Data mining can help overcome certain limitations of these conventional methods, such as multicollinearity, by uncovering associations that might otherwise have gone unnoticed. In this study, we separately analyzed two samples of students from Health Sciences and Engineering and Exact Sciences. We found significant correlations between sleep quality and attributes such as perceived sleep quality, sleep latency, sleep duration, drug use, and the use of medication for depression and anxiety. Decision trees identified different predictive attributes between the two samples. These findings offer a novel insight into sleep quality among college students and may support informed decision-making and targeted interventions.

**Keywords:** sleep quality, college students, data mining

## 1. Introduction

Sleep is a physiological process that occupies nearly one-third of the human lifespan [1]; it is characterized by its reversibility and circadian periodicity, during which immobility and muscle relaxation are accompanied by a decrease in consciousness and responsiveness to external stimuli [2]. Given its crucial role in essential functions, such as memory consolidation, metabolic, endocrine and immune regulation [2], maintaining good sleep quality is vital.

College students are among the most extensively researched populations in relation to sleep, with Health Sciences students being the most studied subgroup. Poor sleep quality is highly prevalent in this population, affecting over 50% of students [3, 4, 5]. In contrast, students from other disciplines, such as Engineering and Exact Sciences, have been scarcely researched. However, literature reports a prevalence of poor sleep quality in this population between 48.7% and 67.8% [6, 7, 8].

Sleep quality is a multifactorial construct influenced by elements such as age, sex, the sleep environment, and comorbidities, among others [9]. Some of these factors may increase the risk of poor sleep quality, while others may serve as protective factors. Alghwiri et al., through logistic regression analysis, identified electronics usage hours, neck pain, headache, and other systemic diseases as risk factors for sleep quality [10]. Similarly, Li et

E-mail address: andreamoralesrobles@outlook.com
https://doi.org/10.5281/zenodo.16819159

al. found that college students in lower academic years (e.g., freshmen and sophomores), those who consumed alcohol or engaged in gambling, exercised less than once per week, or felt pressured during the academic year were at higher risk of experiencing poor sleep quality [11]. Although these approaches provide valuable information regarding students' sleep quality and may be the basis of therapeutic management, they may be limited by issues such as multicollinearity or imprecise estimates when some variable categories have a small number of cases. Data mining can help address these limitations, especially in the development of predictive models and in supporting clinical decision-making [12].

The aim of this study was to analyze the sleep quality of students from two campuses —Health Sciences and Engineering and Exact Sciences— using a data mining approach.

# 2. Method

## 2.1 Study design and participants

A cross-sectional study was conducted using probabilistic sampling. A total of 805 students from the Health Sciences Campus and 92 students from the Engineering and Exact Sciences Campus of the Universidad Autónoma de Yucatán (UADY) were recruited. The distribution of students by academic program on each campus was as follows:

- Health Sciences Campus: 32% medical students, 9.2% Nutrition students, 8.6% Rehabilitation students, 20.4% Odontology students, 18% Pharmaceutical Chemist Biologist (PCB) students, 8.1% Nursing students, and 3.7% Social Work Students.
- Engineering and Exact Sciences Campus: 19.6% Computer Sciences students, 2.8% Computer Engineering students, and 78.3% Software Engineering students.

All regular students of both sexes, aged 18 or older, enrolled in the first to sixth school semester, who agreed to participate voluntarily, were considered for inclusion. Students who were pregnant or working night shifts were excluded to minimize potential bias in sleep quality assessment.

## 2.2 Instruments

### *2.2.1 Sociodemographic data*

Data regarding sociodemographic information was collected through a self-administered 28-item survey. As with the other instruments, this questionnaire was applied using Microsoft Forms, which students could access by scanning a QR code.

### *2.2.2 Sleep quality*

The Pittsburgh Sleep Quality Index (PSQI) was administered to assess students' sleep quality; the Spanish version by Jimenez-Genchi et al. was used, with a Cronbach's alpha reliability coefficient of 0.78 [13]. PSQI is a self-administered 19-item questionnaire developed by Buysse in 1989, that evaluates seven components of sleep quality: subjective (perceived) sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction [14]. A global score, given by the sum of each component's score, of more than five points indicate poor sleep quality [13].

### *2.2.3 Daytime sleepiness*

Since sleep quality is a multifactorial phenomenon, we evaluated some confounding factors, daytime sleepiness being one of them, which was evaluated by the Epworth Sleepiness Scale (ESS). This self-administered 8-item questionnaire was applied to evaluate students' tendency to fall sleep in different situations; the Spanish version by Sandoval-Rincón et al. was used, with a Cronbach's alpha reliability coefficient of 0.89. A total score of 10 to 12 points is suggestive of marginal sleepiness, while more than 12 points indicate excessive sleepiness [15].

### *2.2.4 Internet addiction*

Other confounding factor was the presence of Internet addiction, assessed by the Internet Addiction Test (IAT); the Spanish version by Puerta-Cortés et al. was administered, with a Cronbach's alpha reliability coefficient of 0.91 [16]. This 20-item questionnaire distinguishes between four different levels of Internet addiction: no addiction (total score of 0 to 30 points), mild addiction (total score of 31 to 49 points), moderate addiction (total score of 50 to 79 points) and severe addiction (total score of 80 to 100 points) [17].

## 2.3 Data analysis

Data was analyzed through a data mining approach using the Waikato Environment for Knowledge Analysis (WEKA), version 3.9.6 (Figure 1).

A separate dataset was constructed for each campus, enabling independent analysis of the two populations. Initially, both datasets underwent preprocessing, which included removing inconsistent or incongruent information. A total of 55 attributes were selected for analysis in each dataset. The Health Sciences dataset included 805 instances, while the Engineering and Exact Sciences dataset comprised 92 instances. Then, a discretization process was employed, to transform continuous attributes into nominal ones, making data suitable for data mining algorithms [18].

Three data mining techniques were used to analyze data: tree-based classification, clustering, and association rule mining. For the classification analysis, the J48 decision tree algorithm was applied, with 10-fold cross-validation and the Sleep Quality variable as the reference attribute. For clustering, the SimpleKMeans algorithm was applied, using the Sleep Quality attribute

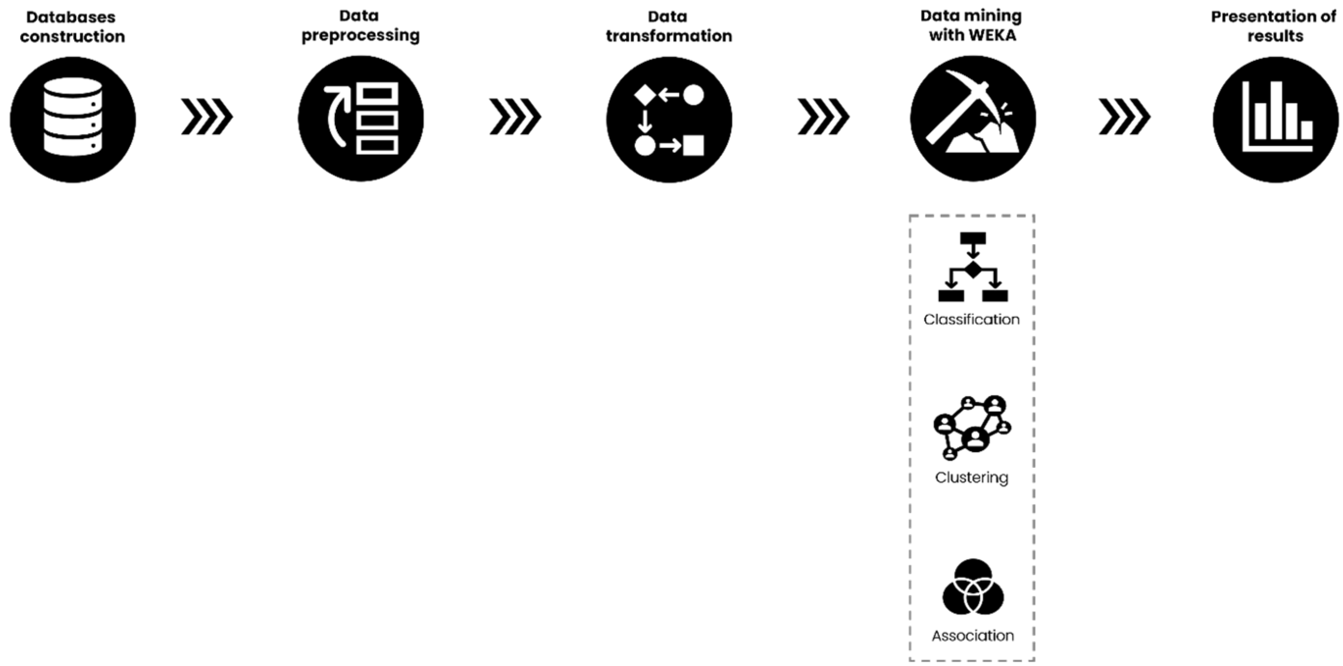



**Figure 1.** Study methodology. Data mining techniques used were tree-based classification (J48 algorithm), clustering (SimpleKMeans algorithm) and association rule mining (Apriori algorithm).

to evaluate cluster alignment. For association analysis, the Apriori algorithm was applied. All algorithms were unsupervised, except for the J48 algorithm.

## 3. Results

From the total sample of 897 students, 89.7% were from the Health Sciences Campus ($n = 805$), and 10.3% were from the Engineering and Exact Sciences Campus ($n = 92$). The prevalence of poor sleep quality among Health Sciences students —61.1% of whom were female— was 74.4%. Among Engineering and Exact Sciences students —18.5% of whom were female— the prevalence was 67.4%.

### 3.1 Data mining

All algorithms were applied separately for each campus dataset in the following order: tree-based classification to generate a decision tree for sleep quality prediction, clustering analysis to characterize students with good and poor sleep quality, and association rule mining to extract the most relevant rules for each campus.

#### *3.1.1 Tree-based classification algorithm*

As previously mentioned, the J48 algorithm was used to generate decision trees.

The algorithm was first applied to the Health Sciences dataset. Sleep Quality was used as the reference attribute, which has two possible values, according to the PSQI: good (equal to or less than five points) or poor (more than five points). A 10-fold cross-validation was performed. To optimize model performance, the MinNumObj parameter was set to 5, and to reduce overfitting, the confidence factor was set to 0.25. The accuracy of the resulting decision tree was 84.9689%, correctly classifying 684 out of 805 instances. Model performance metrics were as follows: Kappa statistic = 0.5903, mean absolute error = 0.2208, root mean squared error = 0.3463, relative absolute error = 57.9151%, and root relative squared error = 79.351% (Figure 2).

Then, the algorithm was applied to the Engineering and Exact Sciences dataset, again using Sleep Quality as the reference attribute and 10-fold cross-validation. For this analysis, the tree was left unpruned to maximize accuracy and minimize error. The MinNumObj parameter was set to 2. The resulting decision tree achieved an accuracy of 81.5217%, correctly classifying 75 out of 92 instances. Model performance metrics were as follows: Kappa statistic = 0.5969, mean absolute error = 0.1964, root mean squared error = 0.4033, relative absolute error = 44.5295%, and root relative squared error = 86.0252% (Figure 3).

#### *3.1.2 Clustering algorithm*

The SimpleKMeans algorithm was employed for this analysis. Using the classes to cluster evaluation mode, Sleep Quality served as the reference for evaluating cluster alignment. For each dataset, two clusters were generated: one predominantly associated with good sleep quality and the other with poor sleep quality. The most relevant attributes for each cluster are presented in Figures 4 and 5, respectively.

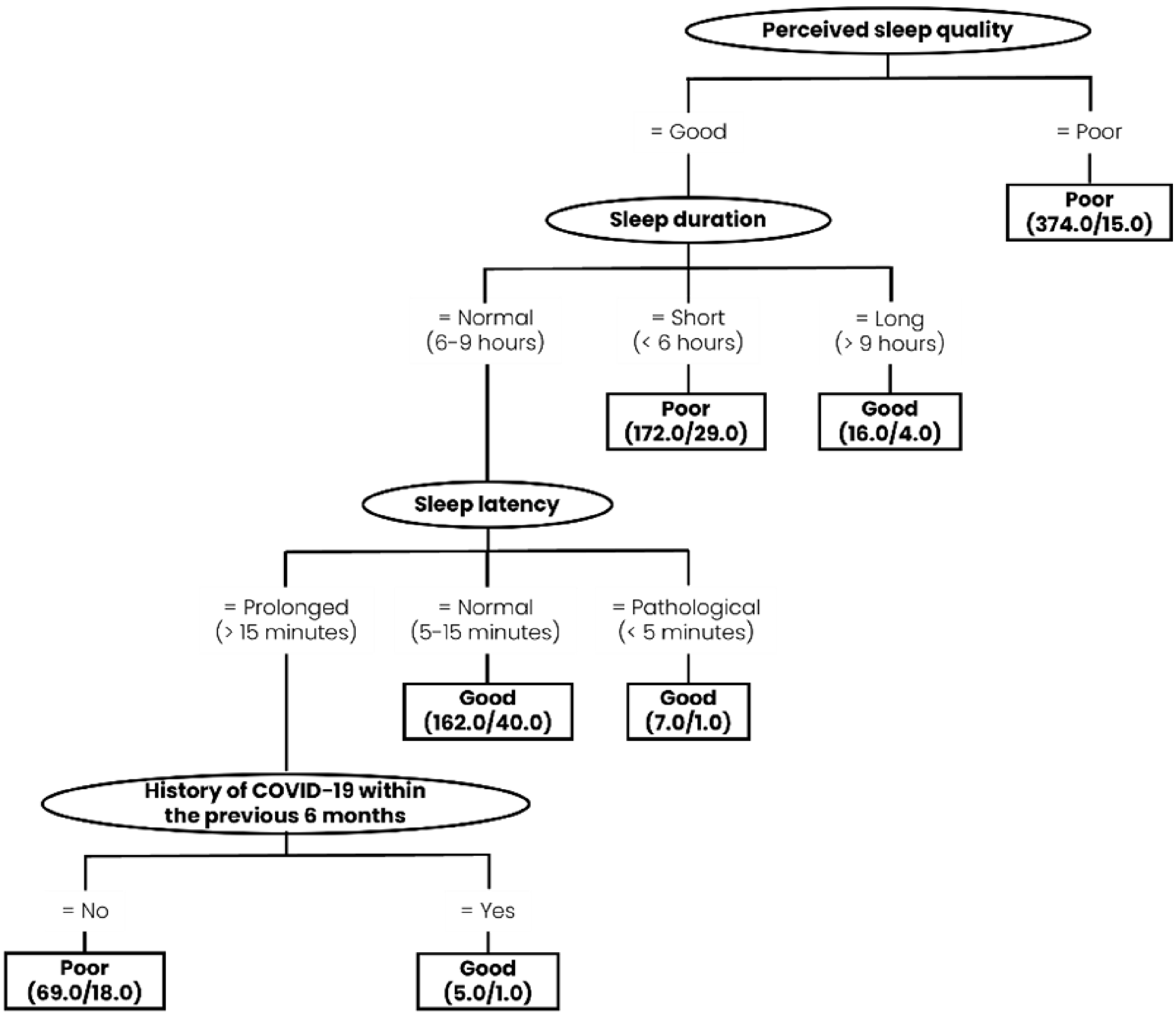



**Figure 2.** Decision tree (J48 algorithm) for sleep quality among Health Sciences students. To optimize model performance, the MinNumObj parameter was set to 5, with a confidence factor of 0.25.

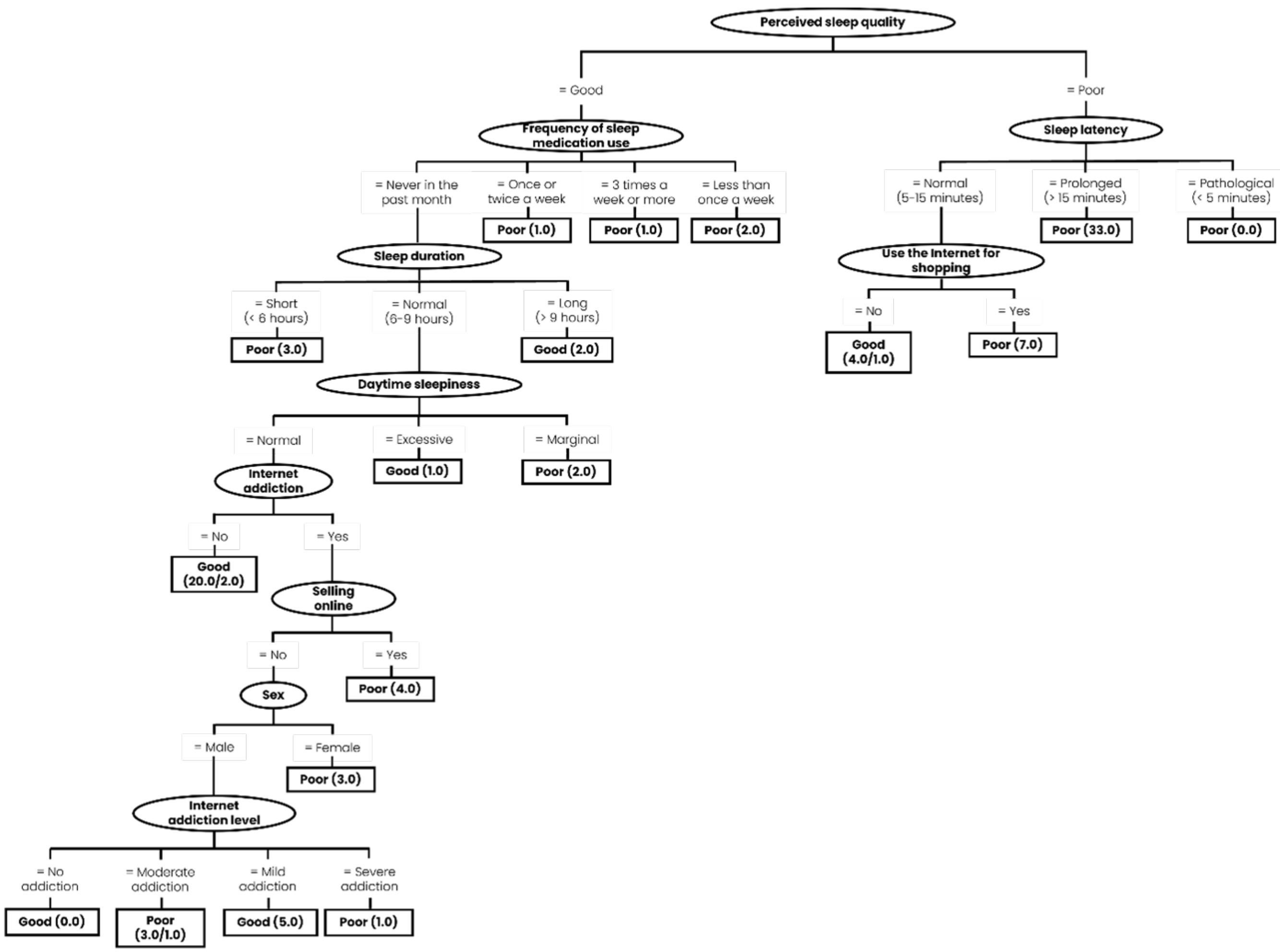



**Figure 3.** Decision tree (J48 algorithm) for sleep quality among Engineering and Exact Sciences students. The tree was left unpruned to maximize accuracy and minimize error; the MinNumObj parameter was set to 2.

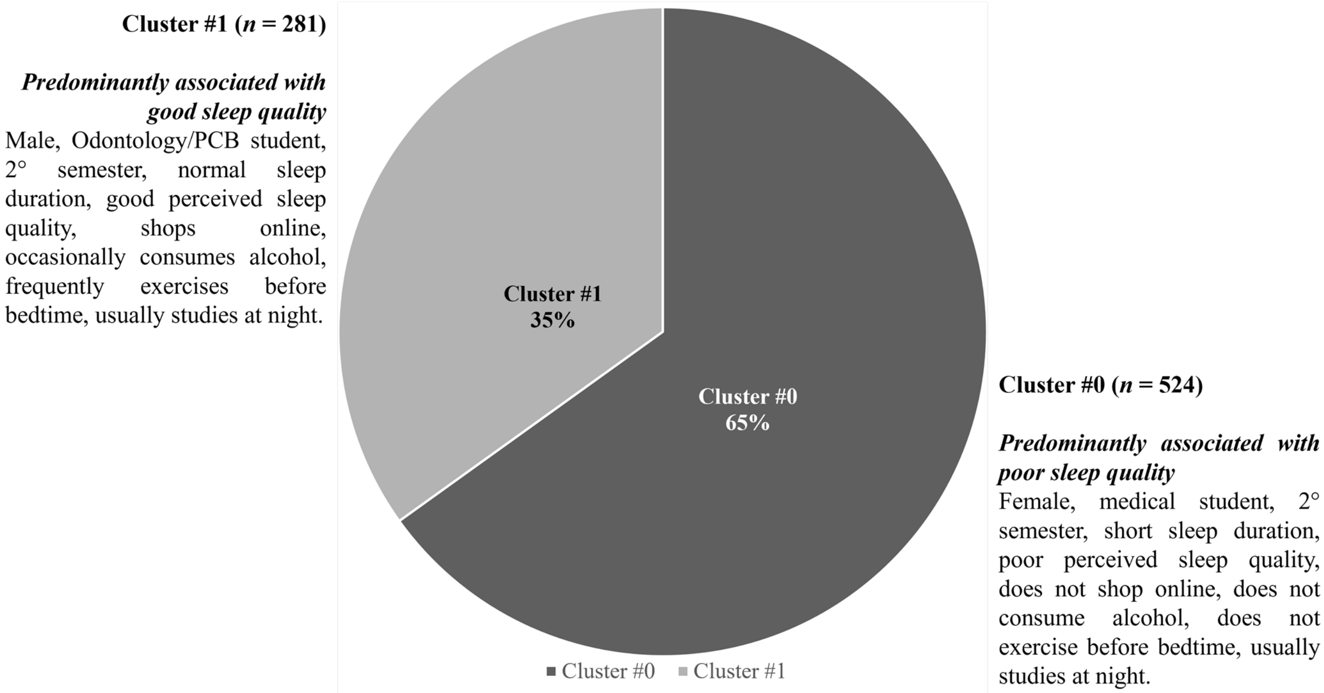



**Figure 4.** Generated clusters (SimpleKMeans algorithm) for the Health Sciences dataset, based on students' sleep quality. The most relevant attributes are presented.

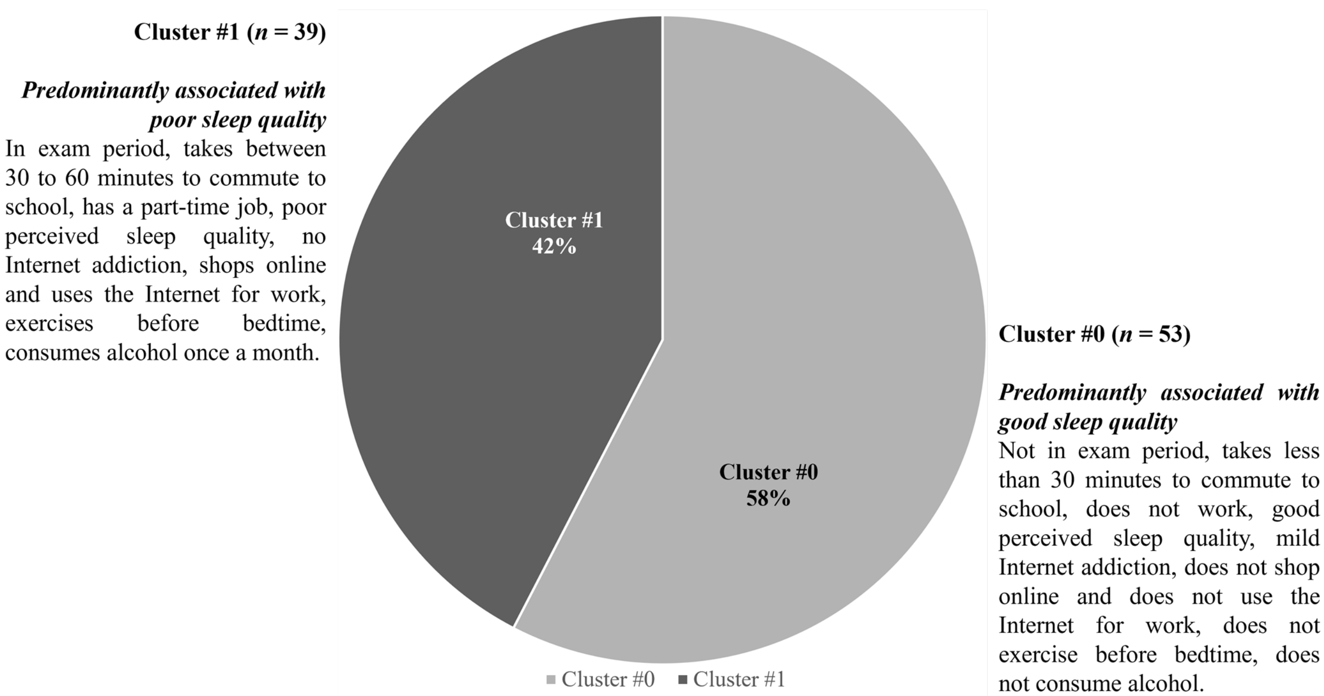



**Figure 5.** Generated clusters (SimpleKMeans algorithm) for the Engineering and Exact Sciences dataset, based on students' sleep quality. The most relevant attributes are presented.

In terms of clustering accuracy, 34.4099% of instances were incorrectly clustered in the Health Sciences dataset. In contrast, only 53.26% of instances were correctly clustered in the Engineering and Exact Sciences dataset. These results suggest that additional, unmeasured attributes may be influencing students' sleep quality and limiting the clustering model's discriminative power.

To optimize clustering model performance, several parameter configurations were tested. The most accurate results were obtained using the following settings: Euclidean Distance as the distance function, k-means++ as the initialization method, a maximum of 500 iterations, two clusters, preservation of instances order enabled, and a random seed of 100. This configuration was applied to both datasets.

#### *3.1.3 Association algorithm*

The Apriori algorithm was applied for this analysis. For the Health Sciences dataset, the algorithm performed 79 cycles, using a minimum support threshold of 0.92 (741 instances) and a minimum confidence threshold of 0.4. For the Engineering and Exact Sciences Campus dataset, 60 cycles were performed with a minimum support threshold of 0.94 (86 instances) and the same confidence threshold of 0.4.

In both datasets, the algorithms were run with the RemoveUseless filter enabled, and parameters were set as follows: number of rules = 50, delta = 0.001, upper bound of minimum support = 1.0, lower bound of minimum support = 0.3, and significance level = -1.0. The most meaningful rules based on relevance, confidence, lift, leverage and conviction metrics are presented in Tables 1 and 2. Conviction values for all rules are greater than 1, indicating reliable associations between attributes. Additionally, lift values above 1 and leverage values greater than 0 suggest positive correlations among the associated attributes.

## 4. Discussion

According to the tree-based classification algorithm, perceived sleep quality emerged as a key variable in evaluating a student's overall sleep quality, as it served as the root node in both decision trees. Since perceived sleep quality is a component of the PSQI, including this variable into a traditional logistic regression model may introduce multicollinearity among independent variables, limiting its value as a predictive factor. However, in a real-world clinical context, this attribute can provide useful information and help guide the assessment toward a more informed evaluation.

When analyzing the decision tree for the Health Sciences dataset, only three additional attributes —besides the root node— were identified as predictors of sleep quality: sleep duration, sleep latency and history of COVID-19 within the previous six months. As in the previous cases, the predictive value of sleep duration and sleep latency is limited. However, longer sleep duration and shorter sleep latency contribute to a lower PSQI overall score, which is interpreted as better sleep quality [13].

Regarding the history of COVID-19 within the previous six months, current evidence presents an ambiguous relationship with sleep quality, potentially linked to the role of sleep in modulating the immune response [19]. In their evaluation of the long-term effects of COVID-19, Mekhael et al. found out that patients who had recovered from the disease exhibited altered sleep architecture, shorter sleep duration and longer sleep latency [20], features commonly associated with poor sleep quality. Interestingly, the decision tree for Health Sciences students indicated that those who had COVID-19 within the previous six months were more likely to report good sleep quality. However, given the limited evidence regarding the association between COVID-19 and sleep quality, it cannot be concluded that a positive history of COVID-19 predicts good sleep quality. Yet, despite the evidence suggesting a negative impact of COVID-19 on sleep, our results indicate that students who had a positive history of COVID-19 within the previous six months were more likely to report good sleep quality, contradictory to the existing evidence.

The Engineering and Exact Sciences Campus decision tree also identified perceived sleep quality as the root node. In addition to sleep duration and sleep latency, other relevant predictors for sleep quality included the frequency of sleep medication use, daytime sleepiness, selling online and Internet addiction. Our results suggest that any frequency of sleep medication use is associated with poor sleep quality. This aligns with Alamir et al. findings, who reported that individuals consuming anxiolytics and antidepressants experienced nocturnal awakenings, daytime sleepiness and insomnia, factors associated with poor sleep quality [21].

Regarding daytime sleepiness, de Sousa et al., using an explanatory model, predicted a higher probability of daytime sleepiness among college students with poor sleep quality [22]. In contrast, our findings showed that those with excessive daytime sleepiness were more likely to report good sleep quality. This apparent contradiction may be explained by the possibility of a distorted perception of sleep as being restorative when the individual experiences greater fatigue.

In this study, female students showed a higher probability of experiencing poor sleep quality, consistent with the findings of Sánchez-Sánchez, who reported that college-aged women had a 13% higher risk of developing poor sleep quality [23]. According to Regal et al., this may be related to the influence of female hormones, a higher prevalence of stress and the onset of disorders such as depression or anxiety [24].

In relation to Internet addiction, our results indicate that students with no or mild addiction are more likely to experience good sleep quality, whereas those with moderate or severe addiction tend to experience poor sleep quality. These findings are consistent with a meta-analysis by Alimoradi et al., which concluded that individuals with Internet addiction have twice the risk of experiencing sleep disorders, daytime sleepiness,

**Table 1.** Generated association rules (Apriori algorithm) for Health Sciences students.

| Rule | Description | Confidence | Lift | Leverage | Conviction |
|---|---|---|---|---|---|
| 1 | If a student never consumes drugs, then they do not consume drugs. | 1.00 | 1.06 | 0.05 | 42.48 |
| 2 | If a student does not consume drugs, then they never consume drugs. | 1.00 | 1.06 | 0.05 | 42.48 |
| 3 | If a student is single and never consumes drugs, then they do not consume drugs. | 1.00 | 1.06 | 0.05 | 41.42 |
| 4 | If a student is single and does not consume drugs, then they never consume drugs. | 1.00 | 1.06 | 0.05 | 41.42 |
| 5 | If a student never consumes drugs and does not take medication for depression, then they do not consume drugs. | 1.00 | 1.06 | 0.05 | 41.42 |
| 6 | If a student does not consume drugs and does not take medication for depression, then they never consume drugs. | 1.00 | 1.06 | 0.05 | 41.42 |
| 7 | If a student never consumes drugs, then they are single and do not consume drugs. | 0.97 | 1.06 | 0.05 | 3.02 |
| 8 | If a student does not consume drugs, then they are single and never consume drugs. | 0.97 | 1.06 | 0.05 | 3.02 |
| 9 | If a student never consumes drugs, then they do not consume drugs and do not take medication for depression. | 0.97 | 1.06 | 0.05 | 3.02 |
| 10 | If a student does not consume drugs, then they never consume drugs and do not take medication for depression. | 0.97 | 1.06 | 0.05 | 3.02 |
| 11 | If a student uses the Internet for communication but does not take medication for anxiety, then they do not take medication for depression. | 0.99 | 1.02 | 0.02 | 2.78 |
| 12 | If a student has normal sleep efficiency and does not take medication for depression, then they do not take medication for anxiety. | 0.99 | 1.02 | 0.02 | 2.57 |
| 13 | If a student does not take medication for anxiety, then they do not take medication for depression. | 0.99 | 1.02 | 0.02 | 2.55 |
| 14 | If a student is single and does not take medication for anxiety, then they do not take medication for depression. | 0.99 | 1.02 | 0.02 | 2.48 |
| 15 | If a student has normal sleep efficiency and does not take medication for anxiety, then they do not take medication for depression. | 0.99 | 1.02 | 0.02 | 2.45 |
| 16 | If a student does not take medication for depression, then they do not take medication for anxiety. | 0.99 | 1.02 | 0.02 | 2.38 |
| 17 | If a student is single and does not take medication for depression, then they do not take medication for anxiety. | 0.99 | 1.02 | 0.02 | 2.31 |
| 18 | If a student uses the Internet for communication but does not take medication for depression, then they do not take medication for anxiety. | 0.99 | 1.02 | 0.02 | 2.27 |

**Table 2.** Generated association rules (Apriori algorithm) for Engineering and Exact Sciences students.

| Rule | Description | Confidence | Lift | Leverage | Conviction |
|---|---|---|---|---|---|
| 1 | If a student never consumes drugs, then they do not consume drugs. | 1.00 | 1.06 | 0.05 | 4.73 |
| 2 | If a student does not consume drugs, then they never consume drugs. | 1.00 | 1.06 | 0.05 | 4.73 |
| 3 | If a student is single and never consumes drugs, then they do not consume drugs. | 1.00 | 1.06 | 0.05 | 4.67 |
| 4 | If a student is single and does not consume drugs, then they never consume drugs. | 1.00 | 1.06 | 0.05 | 4.67 |
| 5 | If a student never consumes drugs and has no history of COVID-19 within the previous six months, then they do not consume drugs. | 1.00 | 1.06 | 0.05 | 4.67 |
| 6 | If a student does not consume drugs and has no history of COVID-19 within the previous six months, then they never consume drugs. | 1.00 | 1.06 | 0.05 | 4.67 |
| 7 | If a student never consumes drugs, then they are single and do not consume drugs. | 0.99 | 1.06 | 0.05 | 2.84 |
| 8 | If a student does not consume drugs, then they are single and never consume drugs. | 0.99 | 1.06 | 0.05 | 2.84 |
| 9 | If a student never consumes drugs, then they do not consume drugs and have no history of COVID-19 within the previous six months. | 0.99 | 1.06 | 0.05 | 2.84 |
| 10 | If a student does not consume drugs, then they never consume drugs and have no history of COVID-19 within the previous six months. | 0.99 | 1.06 | 0.05 | 2.84 |
| 11 | If a student does not take medication for anxiety, then they do not take medication for depression. | 1.00 | 1.02 | 0.02 | 1.93 |
| 12 | If a student does not take medication for anxiety and has no history of COVID-19 within the previous six months, then they do not take medication for depression. | 1.00 | 1.02 | 0.02 | 1.91 |
| 13 | If a student is single and does not take medication for anxiety, then they do not take medication for depression. | 1.00 | 1.02 | 0.02 | 1.89 |
| 14 | If a student uses the Internet for entertainment but does not take medication for anxiety, then they do not take medication for depression. | 1.00 | 1.02 | 0.02 | 1.87 |
| 15 | If a student is single, does not take medication for anxiety and has no history of COVID-19 within the previous six months, then they do not take medication for depression. | 1.00 | 1.02 | 0.02 | 1.87 |

and consequently, poor sleep quality [25]. It is likely that individuals with mild addiction share more characteristics with those who have no addiction than with those who have moderate or severe addiction, making them more likely to experience good rather than poor sleep quality, offering valuable insights not only for the clinical approach to these individuals but also for their management. Additionally, this attribute may have been particularly relevant among Engineering and Exact Sciences students, where the frequent use of electronic devices and constant Internet access is practically unavoidable due to the nature of their academic programs.

When examining the generated clusters, poor sleep quality appears to be prevalent across both datasets, which is consistent with current evidence indicating a prevalence rate above 60% [3, 4, 5]. Although specific evidence regarding the Engineering and Exact Sciences students is limited, since most existing studies focus on college students in general or specifically in healthcare students, the literature review revealed a prevalence of poor sleep quality of 48.7% among students from the fields of Informatics or Mechanical Engineering [8], and between 65.1% and 67.8% among students from Sciences and Engineering disciplines [6, 7].

Although the four generated clusters share some attributes, it is important to highlight certain distinguishing features. For example, in the Health Sciences campus, poor sleep quality was associated with female students in their second school semester. This finding aligns with the results reported by Li et al., who stated that college students in lower academic years are at a higher risk of experiencing poor sleep quality [11]. In contrast, good sleep quality among these students was associated with occasional consumption of alcohol, which has a sedative effect at low doses. However, with repeated and chronic consumption, this substance may lead to sleep disorders and poor sleep quality [26]. Conversely, unlike the cluster predominantly associated with poor sleep quality in the Health Sciences dataset, students from the Engineering and Exact Sciences campus showed an association between poor sleep quality and alcohol consumption once a month. This contradictory finding may be due to unexamined factors related to alcohol use itself, such as quantity consumed or type of alcohol.

Finally, according to the association algorithm, for students of both campuses, particular attention should be paid to the consumption of drugs, and medication for depression and anxiety as this attributes where strongly correlated with a confidence ranging between 0.97 and 1. As mentioned before, poor sleep quality is related to the consumption of anxiolytics and antidepressants [21], but also psychoactive substances, like alcohol, caffeine, or drugs, especially in those who consume them chronically [26].

### 4.1 Proposal of a Decision Support System based on AI models

Given the relevance of the identified predictors of sleep quality and the structure of the generated models, we propose the development of a simple, interpretable Decision Support System (DSS). This system would align with AI-assisted mental health frameworks and serve as an early identification tool for students at risk of poor sleep quality, thereby promoting preventive strategies.

The proposed DSS would integrate the three modeling approaches used in this study —decision trees, clustering, and association rule mining— and operate in the following sequence:

1. Input module: The system receives data from a brief student self-report instrument, including key variables identified in this study, such as perceived sleep quality, sleep latency, use of medications and Internet usage habits.

2. Risk classification: Using a trained decision tree model, the system classifies each student as likely to have good or poor sleep quality, supported by interpretable rules derived from the tree structure.

3. Profiling and feedback: The student is matched to one of the identified clusters (e.g., low-risk profile, high-risk profile with medication use, etc.), and a brief narrative is provided to describe the student's risk profile, including associated lifestyle or academic habits.

4. Recommendation engine: Based on the most relevant association rules, the system generates tailored recommendations, such as seeking psychological counseling, limiting electronic device usage before bedtime, or referral to a health professional.

5. Follow-up planning: The system stores the classification result for follow-up or intervention planning.

This DSS would emphasize interpretability, transparency and utility in non-specialized settings such as universities, making it a valuable tool for early intervention in students' mental health.

## 5. Conclusions

Poor sleep quality is a highly prevalent problem among students in both Health and Exact Sciences disciplines. Current evidence indicates that several attributes are associated with and can predict poor sleep quality. However, analysis using a data mining approach revealed additional, potentially unnoticed attributes, including perceived sleep quality, sleep latency or sleep duration. The decision trees and the profiles of students with poor sleep quality differed between campuses, identifying key factors specific to each population.

These findings underscore the importance of alternative analytical approaches that can overcome the limitations of conventional statistical methods, such as multicollinearity, and generate insights that may be more applicable in real-world clinical scenarios for predicting diagnoses and supporting informed therapeutic decision-making.

The proposed DSS may help address this problem by integrating the data mining techniques used in this study, offering a useful tool for early intervention in students' mental health in alignment with AI-assisted frameworks.

## Ethics Statement

This study was reviewed and approved by the Ethics Committee of the Faculty of Medicine at the Universidad Autónoma de Yucatán under protocol number 01-2023. All participants were informed about the study and provided voluntary consent prior to their involvement. This study used anonymized data collected via an online survey. No personally identifiable information was collected.

## CRediT authorship contribution statement

**Andrea Morales-Robles:** Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Software, Visualization, Writing – original draft, Writing – review & editing. **Víctor Menéndez-Domínguez:** Conceptualization, Methodology, Project administration, Supervision, Software, Writing – review & editing. **Héctor Rubio-Zapata:** Conceptualization, Methodology, Supervision, Writing – review & editing.

## Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Microsoft Copilot in order to improve readability. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

## Declaration of competing interest

The authors declare no competing interests.

## Acknowledgements

The authors would like to acknowledge the Health Sciences Campus and the Engineering and Exact Sciences Campus of the Universidad Autónoma de Yucatán for the facilities provided during the completion of this study. They also extend their gratitude to all the university students who voluntarily agreed to participate.

## References

[1] L. Fabres and P. Moya, "Sueño: conceptos generales y su relación con la calidad de vida," *Revista Médica Clínica Las Condes*, vol. 32, no. 5, pp. 527–534, 2021.

[2] P. Carrillo-Mora, J. Ramírez-Peris, and K. Magaña-Vázquez, "Neurobiología del sueño y su importancia: antología para el estudiante universitario," *Revista de la Facultad de Medicina de la UNAM*, vol. 56, no. 4, pp. 5–15, 2013.

[3] M. J. Tahir, N. I. Malik, I. Ullah, H. R. Khan, S. Perveen, R. Ramalho, A. R. Siddiqi, S. Waheed, M. M. Mohamed Shalaby, D. De Berardis, S. Jain, G. L. Vetrivendan, H. Chatterjee, W. X. Gopar Franco, M. A. Shafiq, N. T. Fatima, M. Abeysekera, Q. Sayyeda, S. F. Shamat, W. Aiman, Q. Akhtar, A. Devi, A. Aftab, S. Shoib, C.-Y. Lin, and A. H. Pakpour, "Internet addiction and sleep quality among medical students during the covid-19 pandemic: A multinational cross-sectional survey," *PLoS ONE*, vol. 16, no. 11, 2021.

[4] A. Mohammadbeigi, R. Absari, F. Valizadeh, M. Saadati, S. Sharifimoghadam, A. Ahmadi, M. Mokhtari, and H. Ansari, "Sleep quality in medical students; the impact of over-use of mobile cell-phone and social networks," *Journal of Research in Health Sciences*, vol. 16, no. 1, pp. 46–50, 2016.

[5] A. Morales-Robles, "Asociación entre la calidad de sueño y la adicción al internet en estudiantes de diferentes licenciaturas del área de la salud," Master's thesis, Universidad Autónoma de Yucatán, Mérida, México, 2024.

[6] N. Hasim, J. Yusof, F. A. Ibrahim, M. S. Bungsu, N. A. Yakub, and N. I. Shari, "Stress and sleep quality among engineering students," *Sains Humanika*, vol. 15, no. 3, pp. 29–40, 2023.

[7] T. Kuhn, S. Karsan, J. J. Heisz, and L. E. Middleton, "The differing relationships between academic discipline, sleep hygiene, and dysfunctional sleep attitudes on sleep quality and duration in canadian university students," *Frontiers in Psychology*, vol. 15, 2024.

[8] M. J. Schmickler, S. Blaschke, R. Robbins, and F. Mess, "Determinants of sleep quality: A cross-sectional study in university students," *International Journal of Environmental Research and Public Health*, vol. 20, no. 3, 2023.

[9] J. C. Sierra, C. Jiménez-Navarro, and J. D. Martín-Ortiz, "Calidad del sueño en estudiantes universitarios: importancia de la higiene del sueño," *Salud Mental*, vol. 25, no. 6, pp. 35–43, 2002.

[10] A. A. Alghwiri, F. Almomani, A. A. Alghwiri, and S. L. Whitney, "Predictors of sleep quality among university students: the use of advanced machine learning techniques," *Sleep and Breathing*, vol. 25, no. 2, pp. 1119–1126, 2021.

[11] Y. Li, W. Bai, B. Zhu, R. Duan, X. Yu, W. Xu, M. Wang, H. Wanqing, W. Yu, W. Li, and C. Kou, "Prevalence and correlates of poor sleep quality among college students: a cross-sectional survey," *Health and Quality of Life Outcomes*, vol. 18, no. 210, 2020.

[12] W.-T. Wu, Y.-J. Li, A.-Z. Feng, L. Li, T. Huang, A.-D. Xu, and J. Lyu, "Data mining in clinical big data: the frequently used databases, steps, and methodological models," *Military Medical Research*, vol. 8, no. 44, 2021.

[13] A. Jiménez-Genchi, E. Monteverde-Maldonado, A. Nenclares-Portocarrero, G. Esquivel-Adame, and A. de la Vega-Pacheco, "Confiabilidad y análisis factorial de la versión en español del Índice de calidad de sueño de pittsburgh en pacientes psiquiátricos," *Gaceta Médica de México*, vol. 144, no. 6, pp. 491–496, 2008.

[14] D. J. Buysse, C. F. Reynolds III, T. H. Monk, S. R. Berman, and D. J. Kupfer, "The pittsburgh sleep quality index: A new instrument for psychiatric practice and research," *Psychiatry Research*, vol. 28, pp. 193–213, 1989.

[15] M. Sandoval-Rincón, R. Alcalá-Lozano, I. Herrera-Jiménez, and A. Jiménez-Genchi, "Validation of the epworth sleepiness scale in mexican population," *Gaceta Médica de México*, vol. 149, no. 4, pp. 409–416, 2013.

[16] D. X. Puerta-Cortés, X. Carbonell, and A. Chamarro, "Análisis de las propiedades psicométricas de la versión en español del internet addiction test," *Trastornos Adictivos*, vol. 14, no. 4, pp. 99–104, 2012.

[17] K. S. Young and C. N. de Abreu, *Internet Addiction: A handbook and guide to evaluation and treatment.* Hoboken, New Jersey: John Wiley & Sons, Inc., 2011.

[18] A. Rajalakshmi, R. Vinodhini, and K. Fathima Bibi, "Data discretization technique using weka tool," *International Journal of Computer Science Engineering and Technology*, vol. 6, no. 8, pp. 293–298, 2016.

[19] L. Shaik, S. Boike, K. Ramar, S. Subramanian, and S. Surani, "Covid-19 and sleep disturbances: A literature review of clinical evidence," *Medicina*, vol. 59, no. 5, 2023.

[20] M. Mekhael, C. H. Lim, A. H. El Hajjar, C. Noujaim, C. Pottle, N. Makan, L. Dagher, Y. Zhang, N. Chouman, D. L. Li, T. Ayoub, and N. Marrouche, "Studying the effect of long covid-19 infection on sleep quality using wearable health devices: Observational study," *Journal of Medical Internet Research*, vol. 24, no. 7, 2022.

[21] Y. A. Alamir, K. J. Zullig, S. Wen, H. Montgomery-Downs, A. L. Kristjansson, R. Misra, and J. Zhang, "Association between nonmedical use of prescription drugs and sleep quality in a large college student sample," *Behavioral Sleep Medicine*, vol. 17, no. 4, pp. 470–480, 2017.

[22] E. A. de Sousa, L. A. Freires, G. D. Lopes Loureto, and J. C. Albuquerque da Costa, "Sleep quality and daytime sleepiness among college students: testing an explanatory model," *Ciencias Psicológicas*, vol. 17, no. 2, 2023.

[23] Z. Sánchez Sánchez, "Relación del estrés y la calidad de sueño en estudiantes universitarios: Revisión bibliográfica," *FACSA - Terapia Física*, vol. 1, no. 1, 2020.

[24] A. R. Regal, M. C. Amigo, and E. Cebrián, "Sueño y mujer," *Revista de Neurología*, vol. 49, no. 7, pp. 376–382, 2009.

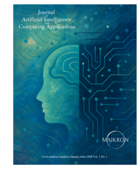
[25] Z. Alimoradi, C.-Y. Lin, A. Broström, P. H. Bülow, Z. Bajalan, M. D. Griffiths, M. M. Ohayon, and A. H. Pakpour, "Internet addiction and sleep problems: A systematic review and a meta-analysis," *Sleep Medicine Reviews*, vol. 47, pp. 51–61, 2019.

[26] M. Sejbuk, I. Mironczuk-Chodakowska, and A. M. Witkowska, "Sleep quality: A narrative review on nutrition, stimulants, and physical activity as important factors," *Nutrients*, vol. 14, no. 9, 2022.

Journal of Artificial Intelligence and Computing Applications (2025) 3(1): 13–23

MAIKRON

Journal of Artificial Intelligence and Computing Applications



*Research article*

# Appraising cognitive status in dementia via touch-based reaction time: a preliminary machine learning study

**Marco Esquer-Rochin[1,*], Luis-Felipe Rodriguez[1], and J. Octavio Gutierrez-Garcia[2]**

[1]Instituto Tecnologico de Sonora (ITSON)
[2]Instituto Tecnológico Autónomo de México (ITAM)

**ABSTRACT**

People with dementia (PwD) perform cognitive-based therapeutic activities. Literature reports a variety of studies exploring relationships between the cognitive status of PwD as determined by the Mini-Mental State Examination (MMSE) and their reaction times from a myriad of stimuli incorporated into cognitive activities. Nevertheless, these technology-supported activities usually include distracting elements, complex instructions, and unfamiliar devices for older adults, introducing bias into reaction times. The objective of this work is to appraise the cognitive status of people with dementia using reaction times from touch interaction tasks. For this purpose, a relatively simple cognitive activity (involving the intuitive tap gesture) and a 32-inch wide touchscreen were designed and implemented. Afterward, 21 PwD from a day center located in Sonora, Mexico were recruited. The participants were instructed to carry out a cognitive activity consisting of five consecutive taps and their reaction times were recorded. The collected data was analyzed using (i) a correlation analysis, (ii) a bootstrap evaluation of machine learning classification models, and (iii) a logistic regression analysis. From the empirical results, it can be concluded that there is a negative relationship between the MMSE score of PwD and the reaction times from taps. In addition, the bootstrapped mean accuracy results of the classifiers suggest that it may be feasible to automatically classify PwD.

**Keywords:** dementia, cognitive tasks, machine learning

## 1. Introduction

Dementia is a general term for a group of progressive symptoms that affect cognitive functions such as memory, thinking, orientation, comprehension, calculations, learning capacity, language, and judgment [1, 2]. According to the World Health Organization, it is projected that there will be 78 million people suffering from dementia by 2030 and 139 million by 2050 [3]. One of the most common risk factors for developing dementia is longevity. Most people who develop dementia are 65 years of age or older [2].

The cognitive function of people with dementia (PwD) deteriorates affecting their quality of life and requiring assistance from caregivers to carry out daily activities [4]. In this regard, there are two types of caregivers: *informal* and *formal*. The former refers to a role commonly played by PwD's family members who do not have caregiving experience [5], whereas the latter refers to professionals formally trained to assist PwD in day centers [4].

In day centers, PwD perform cognitive activities that help improve their cognitive functions and emotional state [6]. The cognitive activities carried out by

E-mail address: octavio.gutierrez@itam.mx
https://doi.org/10.5281/zenodo.16860215

PwD are selected based on their current cognitive status (which is formally assessed using neurocognitive tests such as the Mini-Mental State Examination (MMSE) [1]) and based on formal caregivers' subjective observations about the previous performance of PwD on cognitive activities [7].

Technology-supported cognitive activities provide the opportunity to automatically collect data from PwD, including reaction times [8], attained overall scores [9], or the number of activities completed [10]. For this reason, previous studies have designed and implemented cognitive activities using devices such as tablets [11, 12], computers [13, 14], and tangible devices [1, 8]. Then, the data collected are analyzed to assess the cognitive status of PwD.

The *reaction time*, defined as the time between a stimulus and the motor response to that stimulus [13], has been investigated to determine the cognitive status of PwD [11, 1, 13, 8, 12, 14, 15]. Those studies measure the reaction time of PwD while performing relatively complex cognitive activities on a screen including distractors such as butterflies or moles with hats [16]. In addition, in order to carry out cognitive activities, PwD frequently had to follow elaborate instructions that may have been difficult to understand, e.g., pressing a left or right button according to a green, red or blue light with no apparent clues [8, 11] or even use unfamiliar devices for older people, e.g., a relatively small keyboard [14]. However, it is known that traditional input devices such as keyboards can be an adoption barrier for the elderly [17]; moreover, detailed and complex instructions may be inappropriate for PwD [18]; and creative design features may distract PwD from a given cognitive activity and therefore increase their reaction times [19, 20]. Thus, due to the aforementioned challenges, some research efforts exploring potential relationships between the cognitive status of PwD and their reaction times may be biased by possible difficulties experienced by PwD while performing relatively complex cognitive activities. Furthermore, some research efforts have used supervised machine learning techniques such as support vector machines [14] and random forest [12, 15] to classify the cognitive status of PwD based on their reaction times. For example, random forest models have been trained to screen individuals for cognitive impairment [12], while other approaches have built machine learning models for automated cognitive assessment [8].

The objective of this work is to appraise the cognitive status of people with dementia using reaction times from touch interaction tasks. For this purpose, a relatively simple cognitive activity (involving the intuitive tap gesture) and a 32-inch wide touchscreen were designed and implemented. Afterward, 21 PwD from day center *Dorita de Ojeda* (located in Sonora, Mexico) were recruited. The participants were instructed to carry out a cognitive activity consisting of five consecutive taps and their reaction times were recorded. The data collected was analyzed using (i) a correlation analysis, (ii) a bootstrap evaluation of machine learning classification models, and (iii) a logistic regression analysis.

This work contributes by providing empirical evidence of a negative relationship between the MMSE score of PwD and the reaction times from taps. In addition, bootstrapped mean accuracy results of machine learning models suggest that it may be feasible to automatically classify PwD as individuals having a relatively *low* or *high* MMSE score based on reaction times from taps.

The rest of this article is as follows. Section 2 presents related work; Section 3 describes the participants, instruments, and methods used to explore relationships between the MMSE score of PwD and the reaction time from touch interaction tasks; Sections 4 and 5 present and discuss results, respectively; and Section 6 presents concluding remarks.

## 2. Related work

Literature reports a variety of studies exploring the potential relationship between the cognitive status of PwD (as determined by the MMSE) and their reaction times from a myriad of stimuli incorporated into cognitive activities, which are carried out on technological devices such as tablets, computers, and tangible devices, see Table 1.

In the context of dementia, cognitive activities (designed to measure and collect reaction times) often involve relatively complex instructions, a large number of stimuli, and multiple difficulty levels, see [8, 11, 1, 12, 14]. However, such characteristics may complicate cognitive activities unnecessarily. In addition, cognitive activities are frequently implemented on devices that can be overwhelming for older adults.

In this vein, the study presented in [8], instructed participants to press one out of two buttons depending on a color displayed on a screen while ignoring distractors, e.g., a multicolor LED. The authors collected and analyzed the reaction time from pressing the correct button. In a more elaborate scenario, the authors of [11] implemented a serious game in which participants had to place their hands in a particular position to free their index fingers in order to tap and hold objects for 15 seconds. Similarly, the authors of [1] implemented the traditional Whack-a-Mole game and instructed participants to locate and tap a mole as soon as it appeared on the screen, recording their reaction times. Likewise, in [12], participants were instructed to tap an object that appeared at a random location within a circular area divided into four sections. In [14], the authors evaluated a relatively more complex cognitive activity, in which participants had to press a key on the computer in response to a stimulus displayed on a screen. Unlike the aforementioned studies, this present work makes use of a relatively simple cognitive activity (based on the intuitive tap gesture) involving simple stimulus and using a wide touchscreen.

It should be mentioned that other studies have explored the feasibility of building machine learning models to screen individuals for (mild) cognitive impairment based on their reaction times. For example, random forest and support vector machine models were built in [12, 15] and [14], respectively. In this regard, this present work explores the feasibility of automatically classifying PwD as individuals having a relatively low or high MMSE score based on reaction times from taps using three machine learning algorithms: logistic regressions, decision trees, and support vector machines.

## 3. Material and methods

This section describes the participants involved in the study (Section 3.1), the instruments used (Section 3.2), and the procedure to appraise the cognitive status of PwD using reaction times from touch interaction tasks (Section 3.3).

### 3.1 Participants

Twenty-one individuals formally diagnosed with dementia were recruited from day center *Dorita de Ojeda* to participate in the present study. All participants have an MMSE score of 24 or less (indicating cognitive impairment) with a mean of 14.04 and a standard deviation of 7.36. The age of the participants ranged from 57 to 91 years with a mean of 78.28 and a standard deviation of 7.64.

### 3.2 Instruments

A 32-inch screen with an infrared detection frame was used to enable touch interaction functionalities (see Figure 1 for details). The participants' interactions were recorded using a 150-degree wide-angle camera with a focal distance of 2.1 mm, which was perpendicular to the screen base.

An activity was designed to collect the reaction time of participants from taps. The activity involved a patient tapping five circles that appear one after the other at random locations on a touchscreen. The activity begins once an instructor presses a software-integrated button on the touchscreen. As a result, the first circle appears at a random location, and if and only if the patient taps on the circle, the circle disappears and another one appears at another random location on the touchscreen (see Figure 2). Upon completion, the system displays a flashing visual cue indicating the end of the activity. The activity was implemented using pygame (a Python library). In addition, the design of the graphical user interface was based on best design practices for PwD [21].

### 3.3 Procedure

The study took place at day center *Dorita de Ojeda*. Formal caregivers (working at the day center) recommended conducting the study in several sessions due to the number of participants. They also recommended conducting the study at 11 am, since that is the time PwD normally carry out cognitive activities. The participants and formal caregivers were informed about the objective of the study and their participation was voluntarily. In addition, they were informed that they were free to leave the activity at any time for any reason. Participants performed the activity individually (see Figure 3). It should be mentioned that participants' ages and MMSE scores were requested from the day center's staff.

Before officially beginning the activity, each participant was given a test round to guarantee that the participant understood the instructions. The activity started when the instructor asked the participant to begin. Once the participant completed the activity, a message indicating the end of the activity was displayed at the center of the touchscreen. The activity was recorded using a wide-angle camera and OpenCV in Python.

The reaction time from each tap was computed via video analysis using LINCE PLUS [22], a software for qualitative video analysis. LINCE PLUS allows for video annotations to identify meaningful events, in this case, reaction times from taps. Reaction times were stored in a file with a total of twenty-one (n=21) records each consisting of 9 fields: identifier, MMSE score, MMSE category, five reaction times, and total time (i.e., the sum of reaction times). The MMSE category was set to *low* if the MMSE score was less than 16; otherwise, it was set *high*. This cut-off value was defined to achieve a relatively balanced number of patients in each category: 10 participants categorized as *low* and 11 participants categorized as *high*.

Table 2 reports the dataset, Figure 4 presents histograms and pairwise relationships of reaction times and MMSE scores, and Figure 5 presents box plots highlighting outliers for reaction times for each tap.

The data analysis was guided by the objective of this present study, which is to appraise the cognitive status of PwD using reaction times from touch interaction tasks. The analysis involved three methods: (i) a correlation analysis, (ii) a bootstrap evaluation of machine learning classification models, and (iii) a logistic regression analysis.

The correlation analysis was conducted to explore relationships between the MMSE score of PwD and the reaction times from taps. Due to the relatively small sample and the presence of outliers (Figure 5), Spearman's rank correlation coefficient was used.

To explore the feasibility of machine learning models for classifying PwD as having a *low* or *high* MMSE score based on reaction times, three commonly used machine learning algorithms were used: decision trees, support vector machines, and logistic regression. Using the *MMSE category* as the target variable, models were built for two feature sets, one set including all reaction times from tap 1 to tap 5 in addition to the total time, and another set including only the reaction time from tap 1 (under the hypothesis that the initial reaction time may be the most informative). Due to the small number of instances available for training and evaluating the models, a bootstrap evaluation of machine learning models was used, see Algorithm 1. Bootstrapping is a statistical method for estimating the distribution of an estimator, which, in this article, corresponds to a machine learning model. For each feature set and each

**Table 1.** Related work comparison

| Research effort | Participant type | Activity type | Delivery tool | Machine learning models |
|---|---|---|---|---|
| [8] | Mild cognitive impairment | A choice task | A tangible device | Not specified |
| [11] | Mild cognitive impairment & Alzheimer's | A serious game | A tablet | No |
| [12] | Healthy & Mild cognitive impairment | A tap-based task | A tablet | Random forest |
| [1] | Healthy & Mild cognitive impairment | A serious game | A tangible device | No |
| [14] | Healthy & PwD | A tap-based task | A computer | Support vector machine |
| **Authors' present work** | PwD | A tap-based task | A wide touchscreen | Decision trees, logistic regressions, and support vector machines |

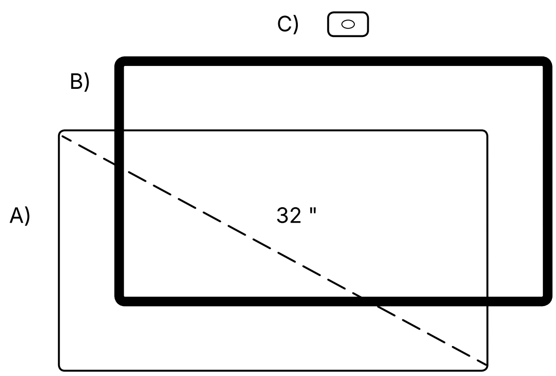



**Figure 1.** Hardware for touch interaction: A) a 32-inch wide touchscreen, B) an infrared detection frame, and C) a wide-angle camera

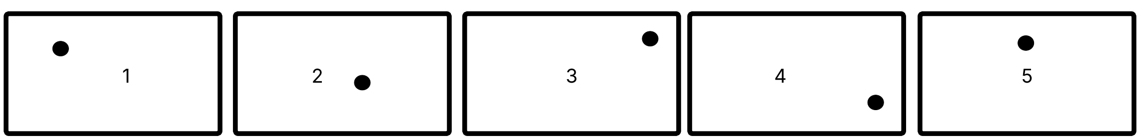



**Figure 2.** Touch activity designed for this study (numbers represent the sequence of events on the screen)

**Figure 3.** A patient from day center *Dorita de Ojeda* performing the activity

**Table 2.** Dataset of 21 People with dementia including reaction times from taps and MMSE scores

| **ID** | **P1** | **P2** | **P3** | **P4** | **P5** | **P6** | **P7** | **P8** | **P9** | **P10** | **P11** | **P12** | **P13** | **P14** | **P15** | **P16** | **P17** | **P18** | **P19** | **P20** | **P21** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MMSE score | 23 | 23 | 17 | 17 | 22 | 19 | 21 | 21 | 15 | 16 | 16 | 10 | 15 | 5 | 9 | 8 | 15 | 3 | 20 | 0 | 0 |
| MMSE category | H | H | H | H | H | H | H | H | L | H | H | L | L | L | L | L | L | L | H | L | L |
| Tap 1 | 2 | 4 | 7 | 4 | 2 | 5 | 2 | 2 | 10 | 4 | 8 | 5 | 7 | 6 | 31 | 11 | 5 | 8 | 3 | 3 | 4 |
| Tap 2 | 2 | 1 | 7 | 4 | 9 | 5 | 5 | 6 | 17 | 6 | 14 | 9 | 6 | 5 | 10 | 9 | 25 | 5 | 24 | 14 | 9 |
| Tap 3 | 2 | 2 | 9 | 3 | 2 | 2 | 2 | 4 | 21 | 14 | 4 | 7 | 5 | 5 | 9 | 35 | 12 | 38 | 6 | 9 | 3 |
| Tap 4 | 2 | 1 | 4 | 5 | 3 | 5 | 1 | 3 | 7 | 26 | 10 | 10 | 5 | 2 | 7 | 13 | 19 | 4 | 4 | 12 | 2 |
| Tap 5 | 2 | 1 | 6 | 4 | 2 | 3 | 6 | 4 | 10 | 15 | 35 | 4 | 11 | 3 | 19 | 30 | 19 | 7 | 11 | 5 | 14 |
| Total time | 10 | 9 | 33 | 20 | 18 | 20 | 16 | 19 | 65 | 65 | 71 | 35 | 34 | 21 | 76 | 98 | 80 | 62 | 48 | 43 | 32 |

*Note:* H and L correspond to high and low MMSE categories, respectively.

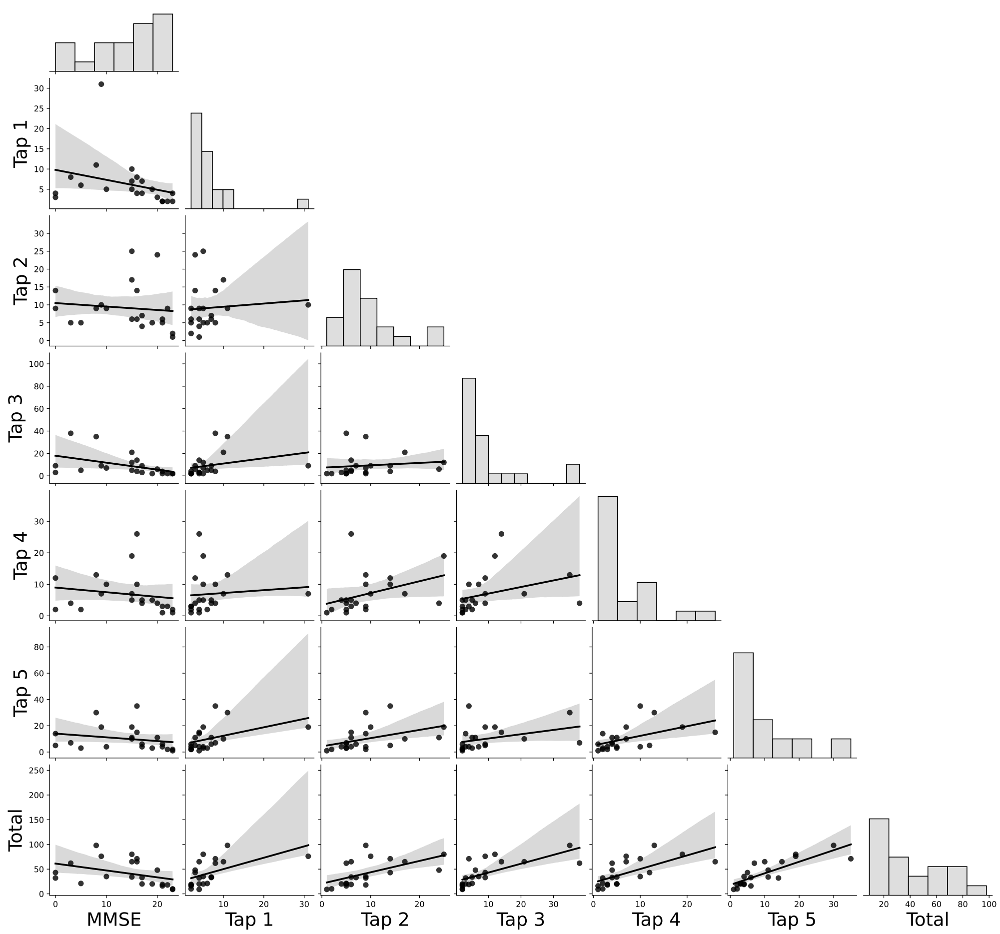



**Figure 4.** Histograms and pairwise relationships of reaction times and MMSE scores

machine learning algorithm, 10,000 bootstrap iterations were executed. Each iteration consisted of (i) sampling the dataset with replacement to create a bootstrap sample (with an approximate size of 64% of the instances) for model training; (ii) creating an out-of-bag set for model evaluation with an approximate size of 36% of the instances; (iii) building a classifier model; and (iv) evaluating the model in terms of accuracy. For each machine learning model, the mean and median accuracy as well as 95% confidence intervals were computed. The machine learning models were built using statsmodels (for logistic regression) and scikit-learn (for decision trees and support vector machines) both Python libraries. It should be noted that default model parameter values were used.

The logistic regression analysis was conducted using statsmodels to explore whether there is a significant effect of the reaction time from tap 1 on the cognitive status of PwD. It should be noted that only the effect related to the reaction time from tap 1 was explored because the logistic regression models trained with the reaction time from tap 1 yielded the best median accuracy (see Section 4 for details).

Except for the correlation analysis, the machine learning models and the regression analysis were conducted using a scaled dataset due to the presence of outliers (see Figure 5). The dataset was scaled using scikit-learn's robust scaler.

## 4. Results

To assess statistical significance, hypothesis test results with a p-value less than 0.05 were considered significant. Statistical analysis was conducted using scipy, a Python library.

### 4.1 Correlation between MMSE score and reaction time

According to Spearman's rank correlation coefficients (see Table 3), there is a relatively moderate and negative monotonic association between the (i) MMSE score of PwD and (ii) the reaction times from taps (except for tap 2) and the total time. In general, as the reaction times from taps increase, the MMSE scores decrease or vice versa. However, it should be noted that causality should not be inferred from these results.

### 4.2 Feasibility of machine learning models for classifying people with dementia based on reaction times

The bootstrapped accuracy results reported in Table 4 show that the median accuracy attained by all the machine learning models was greater than 0.500 (i.e., better than random guessing) regardless of the machine learning algorithm used or whether the model was built using all features (namely, reaction times from tap 1 to tap 5 and total time) or only the reaction time from tap 1. However, for each model, the lower bound of its 95% confidence interval was less than 0.5, see Figs. 6, 7, and 8.

It is worth mentioning that the decision tree models and the support vectors machine models achieved the same median accuracy regardless of whether the models were built with all features or only the reaction time from tap 1 (see Figs. 6 and 7). However, the logistic regression model built with only the reaction time from tap 1 achieved a higher accuracy than the logistic regression model built with all features (see Figure 8). In fact, the logistic regression model trained with only the reaction time from tap 1 yielded the best accuracy results with a mean accuracy of 0.665 and a median accuracy of 0.667 (see Figure 8b). Nevertheless, while the upper bound of its 95% confidence interval is 1.000, its lower bound is less than 0.5.

The median accuracies achieved by the machine learning models suggest that it may be feasible to automatically classify PwD as individuals with a relatively low or high MMSE score based on the reaction times from taps. However, the lower bounds of the 95% confidence intervals do not allow drawing definitive conclusions. Hence, data from more participants must be collected to confirm or refute this finding.

### 4.3 Logistic regression analysis of reaction times to predict MMSE categories

As indicated by the result of the likelihood-ratio test (reported in Table 5), the logistic regression model provides a better fit to the data than the intercept-only model, i.e., the logistic regression model is statistically significant (p-value $< 0.05$). Also, the result of a Wald test rejects the null hypothesis (at a significance level of 0.05) for *tap 1*'s coefficient, then the alternative hypothesis that there is a significant effect of the reaction time from *tap 1* on the cognitive status of PwD is accepted. In addition, since the coefficient for *tap 1* is negative (-1.9624), it can be concluded that as the reaction time of an individual increases, the probability of being categorized as a PwD with a relatively high MMSE score decreases.

## 5. Discussion

This study explored the use of reaction times from a relatively simple activity performed on a wide touch screen to appraise the cognitive status of PwD. The present study involved a representative group of PwD corresponding to the majority of patients attending a day center in Sonora, Mexico. Although this sample is relatively small, the logistic regression analysis concluded that there is a significant effect (p-value $< 0.05$) of the reaction time from *tap 1* on the cognitive status of PwD. In addition, statistically significant Spearman's rank correlation coefficients were found between the MMSE scores and the reaction times from 4 (out of 5) taps. However, with respect to the feasibility of machine learning models for classifying PwD, the results should be interpreted in the context of the relatively small dataset used to train the models. This is because

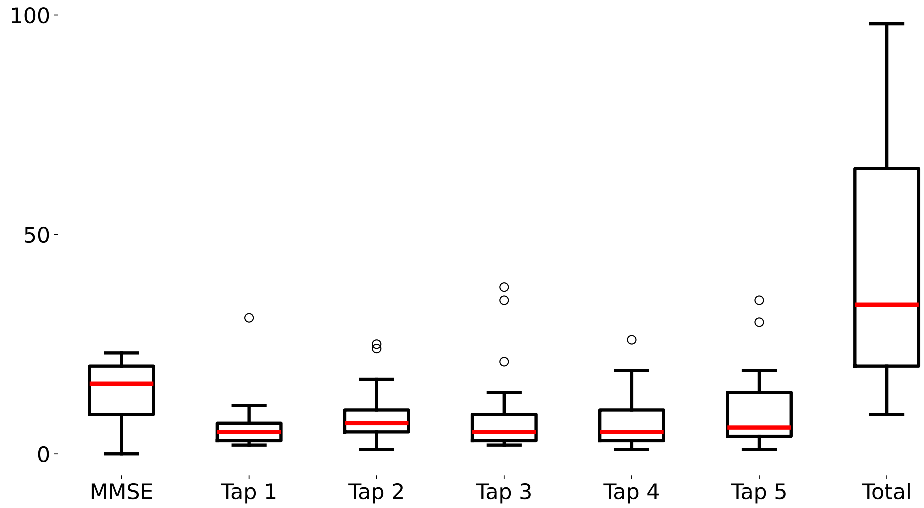



**Figure 5.** Data distribution (box plots) of reaction times and MMSE scores

**Algorithm 1** Bootstrap Evaluation of a Machine Learning Model

**Require:** A *dataset* of people with dementia including reaction times and MMSE scores

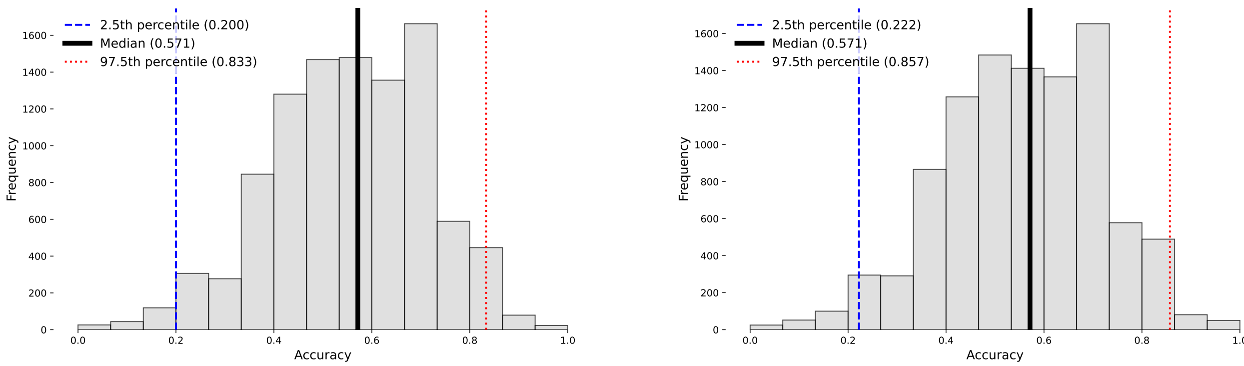
Scale dataset
$n \leftarrow 10000$ // bootstrap iterations
$accuracyScores \leftarrow [\ \ ]$
**for** $i = 1$ to $n$ **do**
    Create a bootstrap sample from the *dataset* for model training
    Create an out-of-bag set for model evaluation
    Train a model using the bootstrap sample
    Evaluate the model using the out-of-bag set
    Append the accuracy score to $accuracyScores$
**end for**
Using $accuracyScores$, compute mean and median accuracy as well as its 95% confidence interval

**Table 3.** Spearman correlation results

| | **Coefficient** | **p-value** |
|---|---|---|
| Tap 1 | -0.567 | 0.0073 |
| Tap 2 | -0.379 | 0.089 |
| Tap 3 | -0.645 | 0.0015 |
| Tap 4 | -0.437 | 0.047 |
| Tap 5 | -0.492 | 0.023 |
| Total time | -0.6282 | 0.0023 |

**Table 4.** Bootstrapped accuracy results of the machine learning models

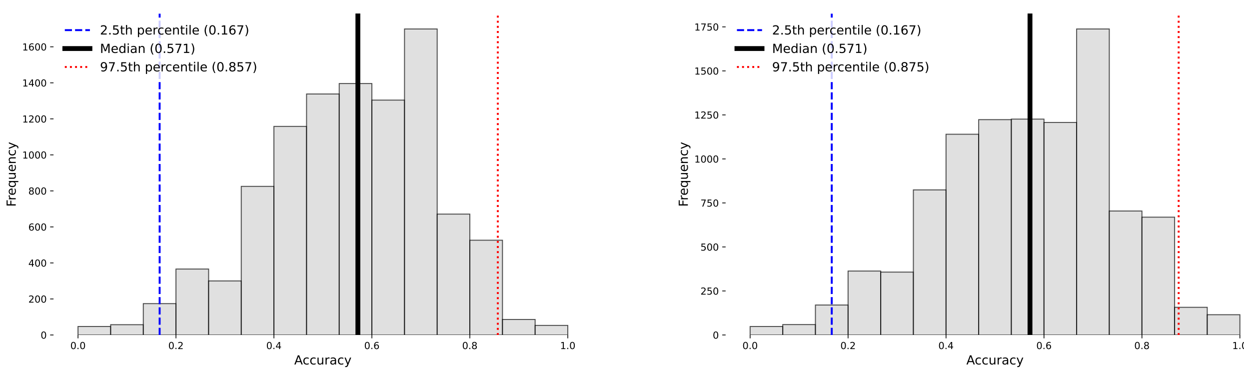
| **Model** | **Features** | ***Mean accuracy*** | ***Median accuracy*** | **95% CI** |
|---|---|---|---|---|
| Decision tree | Reaction times from taps 1-5 and total time | 0.547 | 0.571 | (0.200, 0.833) |
| Decision tree | Reaction time from tap 1 | 0.549 | 0.571 | (0.222, 0.857) |
| Support vector machine | Reaction times from taps 1-5 and total time | 0.549 | 0.571 | (0.167, 0.857) |
| Support vector machine | Reaction time from tap 1 | 0.558 | 0.571 | (0.167, 0.875) |
| Logistic regression | Reaction times from taps 1-5 and total time | 0.591 | 0.600 | (0.250, 0.875) |
| Logistic regression | Reaction time from tap 1 | **0.665** | **0.667** | **(0.333, 1.000)** |




**(a)** Using taps 1-5 and total time

**(b)** Using only tap 1

**Figure 6.** Bootstrap sampling distributions of the accuracy of the decision tree models.




**(a)** Using taps 1-5 and total time

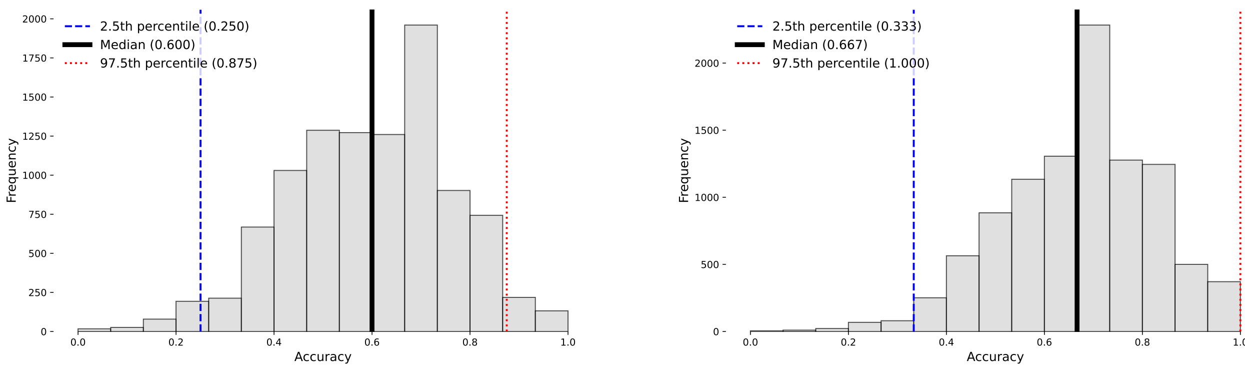
**(b)** Using only tap 1

**Figure 7.** Bootstrap sampling distributions of the accuracy of the support vector machine models.

**Table 5.** Results of the logistic regression analysis

**Likelihood-ratio test (vs. null model) p-value:** 0.007311

| | **Coefficient ($\beta$)** | **Std error** | **z** | **P-value** | **95% confidence interval** |
|---|---|---|---|---|---|
| **Constant** | 0.2378 | 0.522 | 0.455 | 0.649 | (-0.786, 1.261) |
| **Tap 1** | -1.9624 | 0.987 | -1.988 | 0.047 | (-3.897, -0.028) |




**(a)** Using taps 1-5 and total time

**(b)** Using only tap 1

**Figure 8.** Bootstrap sampling distributions of the accuracy of the regression logistic models.

whereas the median accuracy attained by the best performing model (namely, the logistic regression model) was 0.667 with an upper bound of its 95% confidence interval of 1.000, its lower bound was 0.333.

In addition, it is acknowledged that the sample of participants may have included patients with common comorbidities such as arthritis, tremors, or visual difficulties, which may have influenced their reaction times. However, it should be noted that people of advanced age commonly have those comorbidities and that all participants (involved in this study) were able to complete the touch interaction task. Nonetheless, a study involving a relatively large sample may allow grouping participants, for instance, by comorbidity (if any) and conducting separate analysis for each group.

This preliminary study presented models built from three commonly used machine learning algorithms: decision trees, support vector machines, and logistic regression. Each algorithm has a set of hyperparameters that can be tuned to improve their performance, for instance, in decision trees, the quality of a split can be measured using criteria such as gini or entropy. Then, there is, in fact, the possibility of improving the performance of the models presented. In addition, exploring other (deep) machine learning algorithms may help improve model performance. However, this preliminary study focused on exploring the use of reaction times from a relatively simple activity to appraise the cognitive status of PwD.

## 6. Conclusions

The significance of the present work is that it is among the first studies (to the best of authors' knowledge) to (i) appraise the cognitive status of people with dementia using reaction times from touch interaction tasks and (ii) identify significant relationships between the MMSE score of people with dementia (PwD) and the reaction time from taps.

From the empirical results (obtained from the present study involving 21 PwD), it can be concluded that there is a negative relationship between the MMSE score of PwD and the reaction times from taps, i.e., as the MMSE score decreases, the reaction times from taps increase or vice versa. Nevertheless, no causality should be inferred. In addition, the bootstrapped mean accuracy results of the classifiers suggest that it may be feasible to automatically classify PwD as individuals having a relatively *low* or *high* MMSE score based on reaction times from taps. However, the lower bounds of the 95% confidence intervals do not allow drawing definitive conclusions and more experiments should be conducted to confirm or refute this finding.

Future work will focus on designing and implementing cognitive activities for the wide touchscreen to explore the performance of PwD and its relationship to the MMSE score. Another direction for future research is to explore whether the performance of PwD in other touch interaction tasks, for example, tasks involving drag & drop gestures, is related to their MMSE scores. Also, future work will focus on incorporating more sensors (such as microphones and cameras) into the hardware for touch interaction so as to collect and analyze sounds and gestural movements of PwD.

## Ethics Statement

This study was reviewed and approved by the Ethics Committee of Instituto Tecnologico de Sonora under protocol number 393. All participants provided written informed consent in accordance with the Declaration of Helsinki.

## CRediT authorship contribution statement

**Marco Esquer-Rochin:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review editing, Visualization. **Luis-Felipe Rodriguez:** Conceptualization, Project administration, Supervision. **J. Octavio Gutierrez-Garcia:** Conceptualization, Project administration, Supervision.

## Declaration of Generative AI and AI-assisted technologies in the writing process

This manuscript was written without the assistance of generative AI tools. All content, including figures and text, was produced by the authors.

## Declaration of competing interest

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

## Acknowledgements

The authors thank day center *Dorita de Ojeda* for their support. This work was supported by PROFAPI 2024. J. O. Gutierrez-Garcia gratefully acknowledges the financial support from the Asociación Mexicana de Cultura, A.C.

## References

[1] Y.-T. Chen, C.-J. Hou, N. Derek, S.-B. Huang, M.-W. Huang, and Y.-Y. Wang, "Evaluation of the reaction time and accuracy rate in normal subjects, mci, and dementia using serious games," *Applied Sciences*, vol. 11, no. 2, p. 628, 2021.

[2] X.-X. Zhang, Y. Tian, Z.-T. Wang, Y.-H. Ma, L. Tan, and J.-T. Yu, "The epidemiology of alzheimer's disease modifiable risk factors and prevention," *The journal of prevention of Alzheimer's disease*, vol. 8, no. 3, pp. 313–321, 2021.

[3] W. H. Organization *et al.*, "Global status report on the public health response to dementia," 2021.

[4] T.-A. Barreras Diaz, M. Esquer-Rochin, J. O. Gutierrez-Garcia, and L.-F. Rodriguez, "Design and evaluation of a technological platform for monitoring patients with dementia: Unifying requirements from mexican day centers," *Journal of Medical Systems*, vol. 47, no. 1, p. 30, 2023.

[5] P. Lindeza, M. Rodrigues, J. Costa, M. Guerreiro, and M. M. Rosa, "Impact of dementia on informal care: a systematic review of family caregivers' perceptions," *BMJ supportive & palliative care*, vol. 14, no. e1, pp. e38–e49, 2020.

[6] I. D. Saragih, S. I. Tonapa, I. S. Saragih, and B.-O. Lee, "Effects of cognitive stimulation therapy for people with dementia: A systematic review and meta-analysis of randomized controlled studies," *International Journal of Nursing Studies*, vol. 128, p. 104181, 2022.

[7] A. Chávez, G. Borrego, J. O. Gutierrez-Garcia, and L.-F. Rodríguez, "Design and evaluation of a mobile application for monitoring patients with alzheimer's disease: a day center case study," *International journal of medical informatics*, vol. 131, p. 103972, 2019.

[8] C. Vizitiu, V. Stara, L. Antognoli, A. Dinculescu, A. Mosoi, D. M. Kristaly, A. Nistorescu, M. Rampioni, K. Dominey, M. Marin *et al.*, "An iot-based cognitive impairment detection device: A newly proposed method in older adults care—choice reaction time-device development and data-driven validation," *Digital Health*, vol. 10, p. 20552076241293597, 2024.

[9] R. Hamaguchi, S. Hongo, H. Ide, R. Saito, J. Kishimoto, N. Handa, S. Horie *et al.*, "Prospective observational study to evaluate the feasibility of the mobile app for mild cognitive impairment detection and screening," *Frontiers in Digital Health*, vol. 7, p. 1535900, 2025.

[10] L.-Y. Liu, Y. Xing, Z.-H. Zhang, Q.-G. Zhang, M. Dong, H. Wang, L. Cai, X. Wang, and Y. Tang, "Validation of a computerized cognitive training tool to assess cognitive impairment and enable differentiation between mild cognitive impairment and dementia," *Journal of Alzheimer's Disease*, vol. 96, no. 1, pp. 93–101, 2023.

[11] K. L. Sloane, R. Fabian, A. Wright, S. Saxena, K. Kim, C. M. Stein, Z. Keser, S. Glenn, and A. E. Hillis, "Supervised, self-administered tablet-based cognitive assessment in neurodegenerative disorders and stroke," *Dementia and geriatric cognitive disorders*, vol. 52, no. 2, pp. 74–82, 2023.

[12] C. Mychajliw, H. Holz, N. Minuth, K. Dawidowsky, G. W. Eschweiler, F. G. Metzger, F. Wortha *et al.*, "Performance differences of a touch-based serial reaction time task in healthy older participants and older participants with cognitive impairment on a tablet: Experimental study," *JMIR aging*, vol. 7, no. 1, p. e48265, 2024.

[13] S. Ferreira, A. Raimundo, J. del Pozo-Cruz, and J. Marmeleira, "Psychometric properties of a computerized and hand-reaction time tests in older adults using long-term facilities with and without mild cognitive impairment," *Experimental gerontology*, vol. 147, p. 111271, 2021.

[14] J. Wallert, E. Westman, J. Ulinder, M. Annerstedt, B. Terzis, and U. Ekman, "Differentiating patients at the memory clinic with simple reaction time variables: a predictive modeling approach using support vector machines and bayesian optimization," *Frontiers in aging neuroscience*, vol. 10, p. 144, 2018.

[15] Y. Hong, R. L. Alvarado, A. Jog, D. N. Greve, and D. H. Salat, "Serial reaction time task performance in older adults with neuropsychologically defined mild cognitive impairment," *Journal of Alzheimer's Disease*, vol. 74, no. 2, pp. 491–500, 2020.

[16] T. Tong, M. Chignell, and C. A. DeGuzman, "Using a serious game to measure executive functioning: response inhibition ability," *Applied Neuropsychology: Adult*, vol. 28, no. 6, pp. 673–684, 2021.

[17] R. Gündogdu, A. Bejan, C. Kunze, and M. Wölfel, "Activating people with dementia using natural user interface interaction on a surface computer," in *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 2017, pp. 386–394.

[18] L. J. Bird and Y. Y. Lim, "Considerations for the use and design of technology for people living with dementia," *International Psychogeriatrics*, vol. 34, no. 2, pp. 91–94, 2022.

[19] B. Zhang and M. Chignell, "A framework for using cognitive assessment games for people living with dementia," in *2020 IEEE 8th International Conference on Serious Games and Applications for Health (SeGAH)*. IEEE, 2020, pp. 1–8.

[20] C. Krebs, M. Falkner, J. Niklaus, L. Persello, S. Klöppel, T. Nef, P. Urwyler *et al.*, "Application of eye tracking in puzzle games for adjunct cognitive markers: pilot observational study in older adults," *JMIR serious games*, vol. 9, no. 1, p. e24151, 2021.

[21] N. A. Mohd Hassan, A. Baharum, Z. H. Abdullah Sani, K. Chau, and N. A. Mat Noor, "Reducing cognitive impairment among dementia users through mobile application." *Pertanika Journal of Science & Technology*, vol. 29, no. 2, 2021.

[22] A. Soto-Fernández, O. Camerino, X. Iglesias, M. T. Anguera, and M. Castañer, "Lince plus software for systematic observational studies in sports and health," *Behavior research methods*, pp. 1–9, 2022.

Journal of Artificial Intelligence and Computing Applications (2025) 3(1): 24–37

MAIKRON
Journal of Artificial Intelligence and Computing Applications

*Review article*

# Mapping the scientific landscape of artificial intelligence in mental health

**Ivanna Shubina**[1,*] **and Adrian Jarema Dzido**[2]

[1]Liberal Arts Department, American University of the Middle East, 54200, Egaila, Kuwait
[2]Computing Science Department, Radboud University, Nijmegen, the Netherlands

**ABSTRACT**

Artificial Intelligence (AI) has gained increasing popularity in contemporary scientific research; however, its application in mental health still requires a consolidated understanding of existing findings regarding effectiveness. This bibliometric study aims to synthesize current knowledge and explore research trends related to AI's role in mental health. It investigates how advancements in modern technologies—including machine learning, chatbots, and robotics—are used to predict, prevent, and treat mental disorders, and evaluates their effectiveness. A literature search was conducted using Lens software to retrieve peer-reviewed empirical studies in English from highly ranked databases, covering the period from 2005 to 2025. A total of 97 relevant publications were identified and analyzed for patterns, trends, and associations using the Bibliometrix package in R. Results reveal a sharp increase in publications after 2020. Clinical and applied psychology emerged as dominant fields. *Eating and Weight Disorders* is the leading journal ($n$=22), followed by the *Journal of Psychopathology and Behavioral Assessment* ($n$=19) and *Cognitive Therapy and Research* ($n$=17). The United States is both the most productive ($n$=149) and most cited country ($n$=8,896). AI has demonstrated promise in detecting symptoms of depression and suicidal behavior, preventing mental health disorders, and enhancing traditional psychological interventions. Nonetheless, several gaps remain, including the underrepresentation of diverse populations and a limited understanding of factors influencing user acceptance of AI-based tools. This study provides researchers with a comprehensive overview of publication trends, collaboration networks, keyword analysis, and future research directions. It also supports practitioners in selecting appropriate AI-based interventions to improve mental health outcomes and overall well-being within healthcare systems.

**Keywords:** artificial intelligence, mental health, depression

## 1. Introduction

The use of artificial intelligence (AI) in mental health shows potential in diagnosing, preventing, and treating various psychological disorders. AI serves as a valuable supportive tool in psychological interventions for improving mental health, psychological well-being, and overall quality of life among patients. While numerous studies exist on AI's role in clinical psychology, there is no extensive research that brings together findings about its effectiveness, benefits, and future research paths. Some researchers have explored the broader role of technology in mental health [1, 2], while others have specifically examined AI's use in detecting depression

E-mail address: ivanna.shubina@aum.edu.kw
https://doi.org/10.5281/zenodo.16860499

[3, 4], preventing it [5, 6], or identifying suicidal behaviors [7, 8]. Other studies have also looked at AI's effectiveness in dealing with eating disorders [9], anorexia [10], and obesity [11]. Additionally, some efforts have been made to use AI to predict generalized anxiety disorder [12]. However, a comprehensive investigation into the link between AI and mental health remains limited. Therefore, more empirical research is needed to better understand how AI can be effectively applied in the diagnosis, prevention, and treatment of mental disorders. This bibliometric study aims to collect, assess, and combine findings from existing literature to create a full understanding of the correlation between mental health and AI. The study also intends to outline publication patterns and trends in the field, identify knowledge gaps, improve detection of mental disorders, relieve symptoms, and enhance psychological and overall wellbeing.

## 2. Theoretical framework

One major direction in current research focuses on exploring the relationship between AI and mental health. For instance, Higgins et al. [1] examined AI's supportive role in the mental health field and identified several practical challenges in its implementation. Similarly, Ramshaw et al. [2] conducted a comparative analysis of traditional and AI-assisted diagnostic methods, highlighting key barriers to accurate mental health diagnosis. Shan et al. [13] conducted a systematic review of interventions targeting student mental health and emphasized the need for further investigation in this area.

Some studies have specifically assessed the effectiveness of AI in psychological interventions. For example, research suggests that AI can contribute to more personalized and efficient treatment approaches [5]. Evidence highlights the integration of an internet-based cognitive behavioral therapy (CBT) and computer-assisted therapy, although limitations such as the inability to convey body language were noted. A study on AI's role in predicting GAD symptoms found that AI is effective in collecting data critical to risk management for the disorder [12]. In contrast, a study by Maslej et al. [14] indicated that psychiatrists tend to prefer human-based support for patients with depression. Meanwhile, Blease et al. [15] explored psychology students' familiarity with AI, revealing a wide range of opinions and highlighting the importance of integrating AI-related content into academic curricula.

Research has also extensively examined the utilization of AI in the prevention, detection, and treatment of various mental illnesses, including depression [3], suicidal behavior [7, 16, 17], anxiety [12], stress [18], eating disorders [9], and generalized anxiety disorder (GAD) [19], among others. Research has shown that not only CBT and mindfulness-based interventions can be effective in treating mental health conditions, but also social support—whether in person or delivered through AI—can be effective [19].

Another prominent area of research involves the use of AI-based tools to detect various mental health disorders, particularly depression and suicidal behavior. Joshi and Kanoongo [3] analyzed the role of images, facial expressions, the text from social media, and emotional chatbots in detecting depression and emotions, concluding that AI can effectively identify emotional states through multiple modalities, including speech, text, and video. Barros et al. [7] examined factors influencing suicide risk, such as gender, age, psychiatric diagnosis, and type of therapeutic intervention, suggesting that Bayesian networks can either support or hinder psychotherapy outcomes. Similarly, Fonseka, Bhat, and Kennedy [20] evaluated AI's predictive capabilities for suicide risk and stressed the importance of both individual and social dimensions, acknowledging both the strengths and limitations of AI tools [16, 17]. Buddhitha and Inkpen [21] also investigated how AI can detect suicidal ideation among social media users.

Liu et al. [6] argued that AI and related technologies will likely play a larger role in predicting and preventing depression in the future, due to their impartiality in decision-making. Their study analyzed how nonpathological variables such as gender, academic year, lifestyle, and social media use influence depression risk. Inkster, Kadaba, and Sybramanian [22] reported that AI tools helped reduce depressive symptoms among mothers. Kelley et al. [4] also examined the role of social media in recognizing signs of depression and other mental disorders.

The application of AI in identifying and treating eating disorders has gained attention as well. Studies have confirmed AI's efficacy in the identification, prevention, and treatment of such disorders [9]. Noguero et al. [10] focused on gender bias in AI tools used for patients with anorexia, emphasizing the importance of equitable application. Fang et al. [11] evaluated AI-enhanced CBT in treating obesity and reported significant improvements in body weight and body fat percentage.

A growing body of literature has also begun to explore the intersection of AI and psychological wellbeing. Chen [23], for instance, studied the effectiveness of AI-based interventions in improving students' well-being during the COVID-19 pandemic. Sabour et al. [24] assessed the impact of chatbots on mental health, showing that they can play a practical role in reducing psychological distress. Similarly, Yorita et al. [25] investigated how chatbots and robots support occupational stress self-management. Tanaka et al. [26] demonstrated AI's ability to reduce anxiety and traumatic memory in children undergoing distressing medical experiences, thereby enhancing their overall wellbeing. He et al. [27] found that AI could mitigate depression among elderly individuals by reducing social isolation through the analysis of so-called "information cocoons". Another study explored the option of recognizing depression on Twitter [28] or using a smartphone application for the prevention of suicidal episodes among adolescents [29].

Li et al. [30] conducted a meta-analysis on AI's role in promoting well-being, concluding that relationship quality, effective communication, and meaningful content engagement are key factors in users' experiences. Another investigation emphasized the effectiveness of a digital CBT-based resilience intervention for children, recommending the implementation of the CUES program in schools [31]. Jeong et al. [32] showed that AI can enhance students' psychological well-being by increasing motivation, and a similar study [33] found that AI can effectively monitor emotional states, academic performance, university transition, and interpersonal relationships [34, 35, 36, 37, 38, 39].

## 3. Methodology

### 3.1 Methodology of bibliometric analysis

A bibliometric study utilizes statistical methods to analyze research outputs published in peer-reviewed journals, conference proceedings, reviews, and other scholarly sources. This approach enables a comprehensive examination of both theoretical and empirical data, categorized by subject areas, publication years, citation metrics, keywords, and contributions from leading authors, institutions, sources, and countries, and evaluates the impact of these studies. Bibliometric analysis facilitates the identification of prevailing trends and patterns within a specific research domain and related fields over a defined time period. To carry out such an analysis, access to a well-structured and comprehensive research database is essential [40, 41, 42].

### 3.2 Data collection and analysis

Using Lens to access prestigious databases (e.g., PubMed, Crossref, Microsoft Academic, and CORE), we collected recently published papers. Various types of sources—including scholarly works and patents—were statistically analyzed to present datasets with unique content. We adopted a bibliometric method to extract descriptive data on publications, author and co-author metrics, source evaluations, and keyword, thematic map, and network analyzes. Data visualization was performed using Bibliometrix, an R package designed for in-depth bibliometric analysis.

The following parameters were used for the search:

- **Search query:** `Scholarly Works (97) = Artificial AND intelligence OR (digital AND (media AND (impact AND (mental AND health))))`
- **Filters:** Year Published = 2005–2025
- **Field of Study:** Clinical Psychology
- **Subject:** Clinical Psychology

The search for the current bibliometric paper was conducted to retrieve data in plain text (.txt) and Microsoft Excel (.csv) files for analysis and presentation in this paper. Descriptive and bibliometric data analysis was conducted using RStudio in combination with the Bibliometrix package.

### 3.3 Research questions and objective

The purpose of a bibliometric study is to examine publication patterns in the field concerning the use of AI in mental health. It intends to evaluate how the current body of literature helps uncover research gaps and informs potential future research directions. In line with the purpose and scope of this study, the following research questions have been designed:

- RQ1: What are the descriptive characteristics of the publications in this field, including the number of studies published and cited between 2005 and 2025?
- RQ2: What trends can be observed in terms of the most productive authors and co-authors, dominant research themes, key journals, and contributing countries? How do citation metrics reflect the impact of this body of work?
- RQ3: Which keywords are most commonly associated with the published studies, and what do they reveal about the thematic focus of the field?
- RQ4: What patterns of collaboration and research networks exist among the authors of these publications?

## 4. Results

### 4.1 Publication descriptives

A total of 97 relevant studies published between 2005 and 2025 within the explored research domain were identified using Lens (Figure 1). The data sources contributing to this dataset included Microsoft Academic (61 results), PubMed (57), OpenAlex (37), PubMed Central (32), and CORE (14). Due to overlapping entries across databases, duplicates were removed, resulting in a final set of 97 unique publications. The included papers were published with an annual growth rate of 4.33%. The majority of these works were authored in English by 410 authors and co-authors publishing in 10 different sources. Only three papers were written by single authors, and no international co-authorship was identified. Authors used 186 keywords; their papers received 3,871 citations (an average of 39.64 each) and used 6,559 references overall. The average age of each published study is 6.71 years.

### 4.2 Published papers trends over time

The publication records by year, starting from 2005 to 2025, are presented in Figure 2, demonstrating changing trends in interest in conducting studies on the use of AI in mental health. According to Figure 2, two phases can be distinguished. The first phase was between 2005 and 2019, when the publication record ranged from 0 to

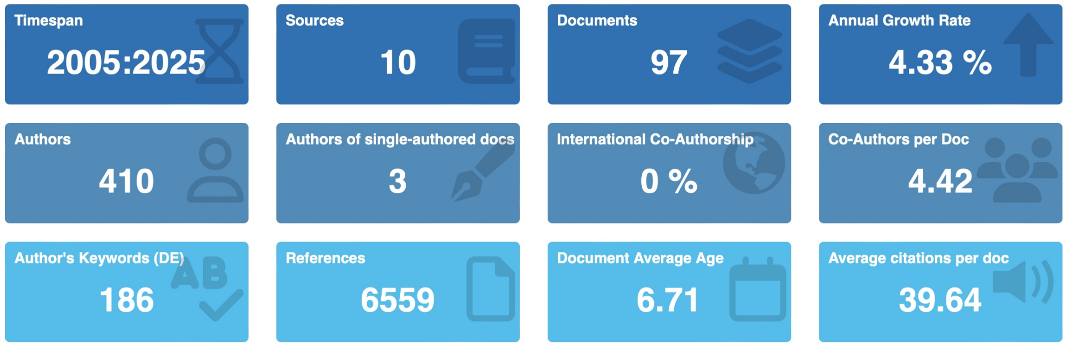



**Figure 1.** Overview of publications included into the analysis.

6 per year, demonstrating comparatively low interest in research on AI in mental health. The second phase, between 2020 and 2025, shows a sharp increase (in 2020, $n = 12$; in 2024, $n = 13$), indicating growing scholarly interest among researchers within the field. Figure 2 illustrates the increasing popularity of AI in mental health as a research subject.

Figure 3 visualizes the publication trends on AI in mental health in selected journals over the past two decades, underlining the dynamics of scholarly output. A significant increase after 2019 reflects growing interest in exploring AI in mental health, likely influenced by the COVID-19 pandemic. The Journal of Psychopathology and behavioral Assessment consistently ranks among the top sources, serving as a key outlet for research on AI in mental health. Cognitive Therapy and Research began contributing after 2014, reaching 16 papers published by 2025. Eating and Weight Disorders published its first relevant paper in 2012 and has emerged as the leading journal over the past three years.

### 4.3 Distribution of productive authors, countries, and collaboration network

Figure 4 highlights the most productive authors over the past two decades and their collaboration networks in the research field of utilizing AI for mental health. It illustrates the publication output of the top 10 authors between 2005 and 2025. Most of them published one or two papers and received low to modest citation counts, reflecting an emerging research dynamic. All authors were active during varying periods, either briefly or over a few years. Boris Birmaher was particularly productive between 2009 and 2016, contributing foundational work that helped shape the direction of future studies. Authors like Barnaby D. Dunn and Anna Brytek-Matera were active primarily between 2015 and 2019, while others have emerged more recently, contributing between 2020 and 2025, indicating increasing interest in this research field. Figure 4 also illustrates temporal trends in research and author involvement, showing the appearance of recent contributors after 2020. The total citation count remains low, likely suggesting that the domain is still evolving.

Figure 5 visualizes the geographic distribution of corresponding authors, along with their overall productivity and collaboration types—single-country publications (SCP) and multiple-country publications (MCP). Authors from the USA contributed the highest number of papers ($n = 149$), most of which were written by American authors (SCP $n = 94$), followed by China ($n = 85$) with a similar distribution between SCP ($n = 40$) and MCP ($n = 45$), Spain ($n = 70$) with a prevalence of SCP ($n = 20$), and Brazil ($n = 65$) with a strong prevalence of SCP ($n = 48$), emphasizing their prominent role in exploring links between AI and mental health. Notably, Poland, Sweden, and Turkey had fewer papers (13, 14, and 15, respectively). The results suggest that SCP publications play key roles in research collaboration on AI and mental health, with the exception of China, the UK, and Chile, where MCP is more prominent. These findings reveal not only where research is being produced, but also how globally integrated each country is in advancing the AI and mental health domain.

Figure 6 illustrates the data on co-authorship collaboration networks within the studied research area, where each node represents an author, and the relationships between them show co-author connections.

The majority of nodes are fragmented or have weak connections, suggesting a growing but still limited collaboration landscape, where a few co-author nodes play a key role in developing thematic clusters, while most research is conducted by isolated author groups. For instance, in research on AI and mental health, the authors' collaboration network is largely concentrated in a few clusters established by Jahrami H and Ricca V, suggesting foundational influence. Clusters with authors like Porter A.L., Merigó J.M., and Kumar S. demonstrate both higher productivity and broader collaborative reach, indicating their foundational impact on shaping this research area. Overall, while collaboration networks are emerging, the landscape remains relatively fragmented, highlighting opportunities for greater international and interdisciplinary collaboration.

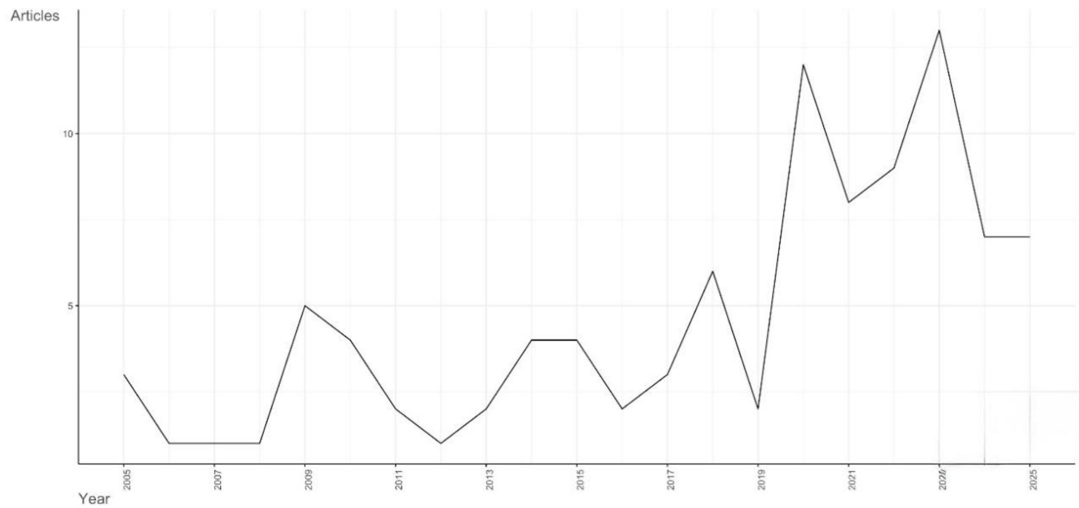



**Figure 2.** Changes in document counts by publication year within the top 10 subjects' areas (2005-2025).

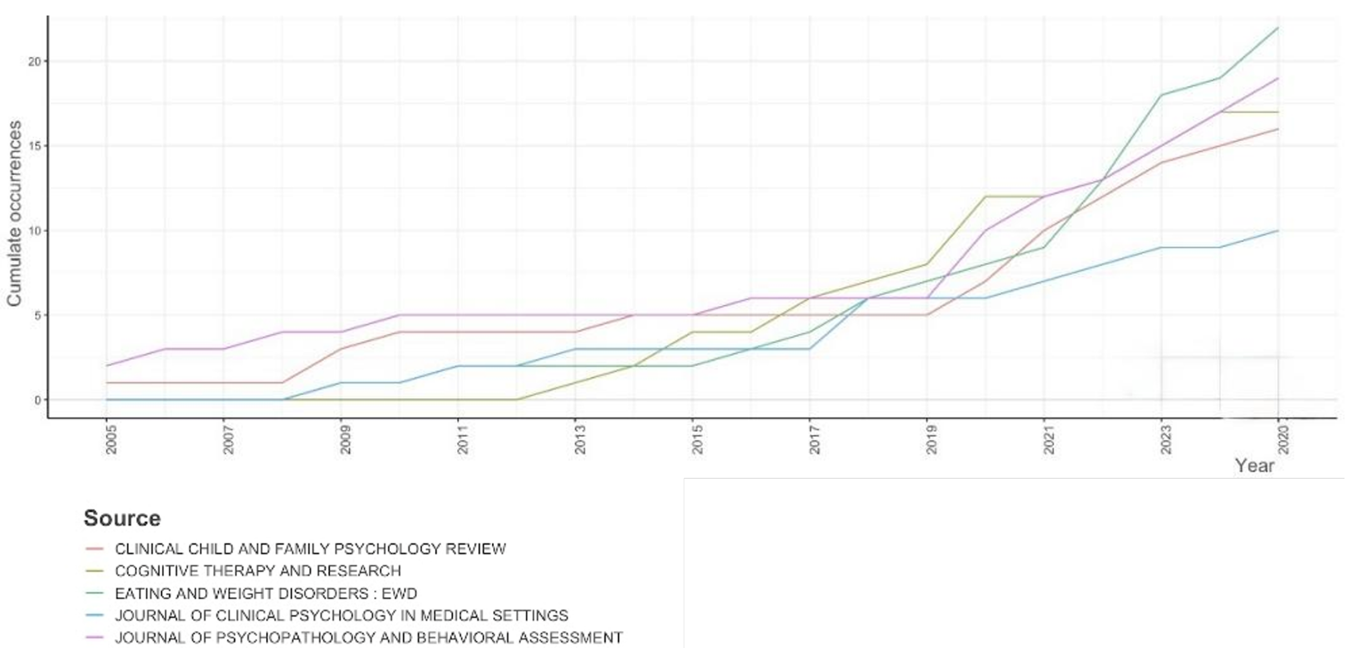



**Figure 3.** Annual publication output of key journals on AI in mental health (2005–2025).

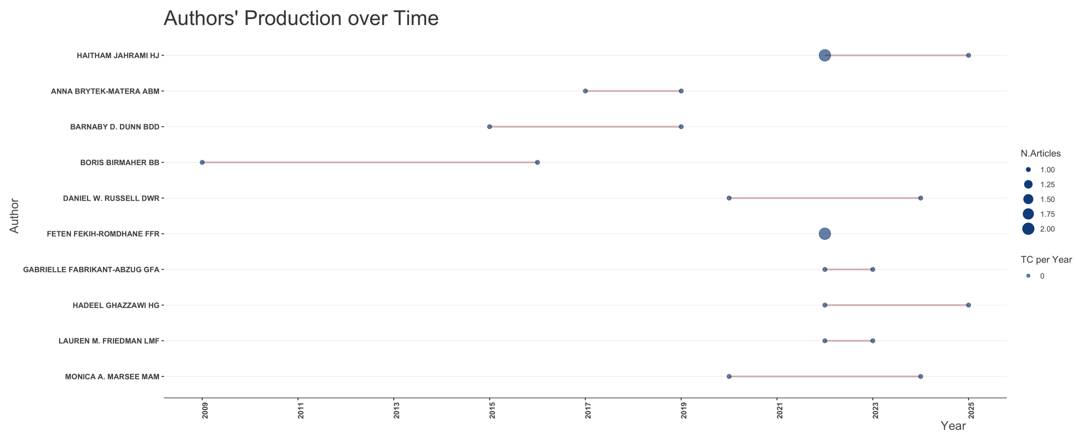



**Figure 4.** Authors' publication output over time (2005-2025).

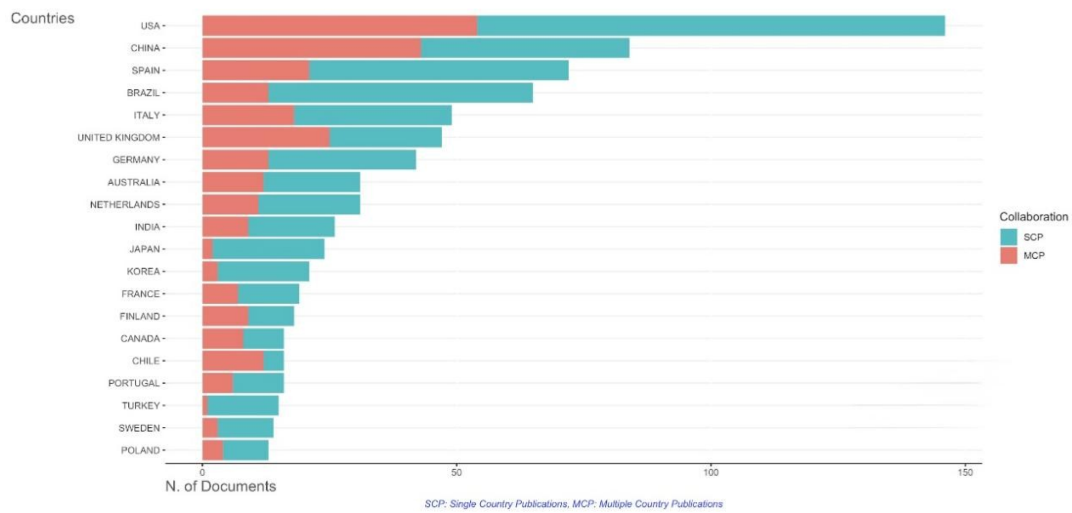



**Figure 5.** Geographic distribution of corresponding authors' contributions and collaboration types.

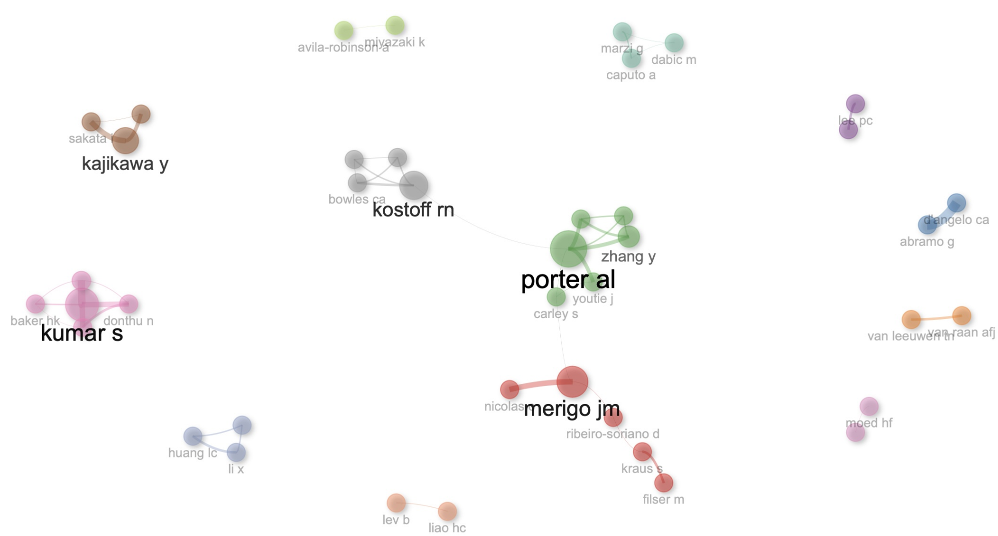



**Figure 6.** Author collaboration network in AI in mental health research.

### 4.4 Keyword analysis and trend topics

Figure 7 visualizes the keyword usage dynamics in studies on the use of AI in the mental health field from 2005 to 2025, highlighting shifts in research interests over time. Starting in 2016, keywords such as "adolescent," "orthorexia nervosa," and "depression" appeared in the domain. After 2022, "eating disorders," "orthorexia nervosa," "depression," and "systematic review" gained prominence, suggesting rising interest in these topics. The increase in the keyword "orthorexia nervosa" may reflect its emerging niche within the domain, potentially driven by novel diagnostic awareness and its correlation with digital cognitive and behavioral patterns. Overall, the findings point to a chronological transition in keyword usage—from more general concepts to a research focus increasingly centered on eating disorders and depression over the past three years—triggering the development of AI-driven diagnostic, preventive, and supportive tools.

Figure 8 illustrates the frequency with which keywords co-occur in research on AI and mental health, presented as nodes. The size of each cluster reflects keyword frequency, and the thickness of the connections indicates the frequency of co-occurrence.

Based on our findings (Figure 8), the keyword co-occurrence network emphasizes the thematic structure of studies on AI applications in mental health. The core concept is "eating disorders," connected with several other keywords, such as "anorexia nervosa" through the "application of AI" and "treatment approaches," and "orthorexia nervosa" via "diagnostic criteria," reflecting a significant research focus on the detection and classification of mental disorders. The keyword "eating disorders" is also connected to "depression," "body image," "self-monitoring," and "obesity," indicating comorbidity and the use of self-tracking technologies in treatment or assessment. Another thematic cluster connects "depression" to "children" through "maltreatment," extending to "CBT" and "unemotional traits," pointing to a developmental and clinical treatment focus among youth. A significant but separate node on "systematic review" indicates the growing number of evidence studies synthesizing data for conceptual consolidation within the field. Moreover, a few weakly connected and peripheral clusters—such as "attitudes toward psychological help-seeking," "childhood trauma," "parents' distress," and "autism spectrum disorder"—represent underexplored or emerging areas within the domain. The structure of the nodes suggests an evolving network in the research domain on AI-driven interventions for mental health (mainly eating and mood disorders), with growing interest in diagnostic, monitoring, and treatment-oriented directions.

### 4.5 Thematic clustering

Studied themes were categorised as motor, basic, niche, and emerging (Figure 9) using two criteria: centrality and density, offering significant insights into the structure of scholarly work on AI and mental health. According to our analysis, the categories focused on "depression," "adolescents," and "anxiety" belong to motor themes, meaning they are central and well-developed. Their location underlines their foundational role in the domain, especially regarding the use of AI tools in diagnosing and treating adolescent mood and eating disorders. However, their positioning also highlights the need for further exploration and development. Moreover, the positioning of the categories "unemotional traits" and the cluster joining "anorexia nervosa," "bulimia," and "AI" on the border of the motor and basic themes suggests their developing relevance in clinical profiling and predictive modelling. A category with "eating disorders" and "orthorexia nervosa" belongs purely to the basic themes, reflecting central but still developing concepts, especially in relation to the diagnostic area. In contrast, the niche themes are well-developed but less central—specialized and methodologically mature. This category includes "body image," "feeding and eating," and "body mass index." A second cluster, "eating disorder," is located near the emerging/declining themes and illustrates the rare development of this topic in the research area. Lastly, the emerging/declining themes include two categories: "obesity" and "systematic reviews." These may indicate either underdeveloped or outdated areas of research.

### 4.6 Citations and co-citations, influential scores and countries

Figure 10 illustrates the top 10 most productive journals that published relevant papers on the use of AI in mental health between 2005 and 2025. *Eating and Weight Disorders* is the leading journal, contributing 22 publications, underlining a significant record of research interest in AI applications for identifying, monitoring, and treating eating disorders. *Journal of Psychopathology and behavioral Assessment* published 19 relevant studies, and *Cognitive Therapy and Research* published 17 papers, suggesting an influential focus on cognitive, behavioral, and clinical appraisal approaches enhanced by AI tools. *Clinical Child and Family Psychology Review* published 16 relevant studies, and *Journal of Clinical Psychology in Medical Settings* contributed 10, indicating ongoing exploration of AI applications in child and family mental health contexts and integrated interventions. Other journals contributed six or fewer studies each, showing a less frequent record of research interest across related psychological, psychiatric, and interdisciplinary topics. Overall, the prevalence of clinically focused and disorder-specific journals indicates that the area is rooted in applied mental health research, with a strong focus on cognitive assessment, eating disorders, and adolescent psychopathology.

Figure 11 shows the top 10 countries by total citation count in the research field of applying AI in mental health.

The United States is leading with 8,896 citations, reflecting its strong position in publishing highly influential studies. This may be a result of the country's

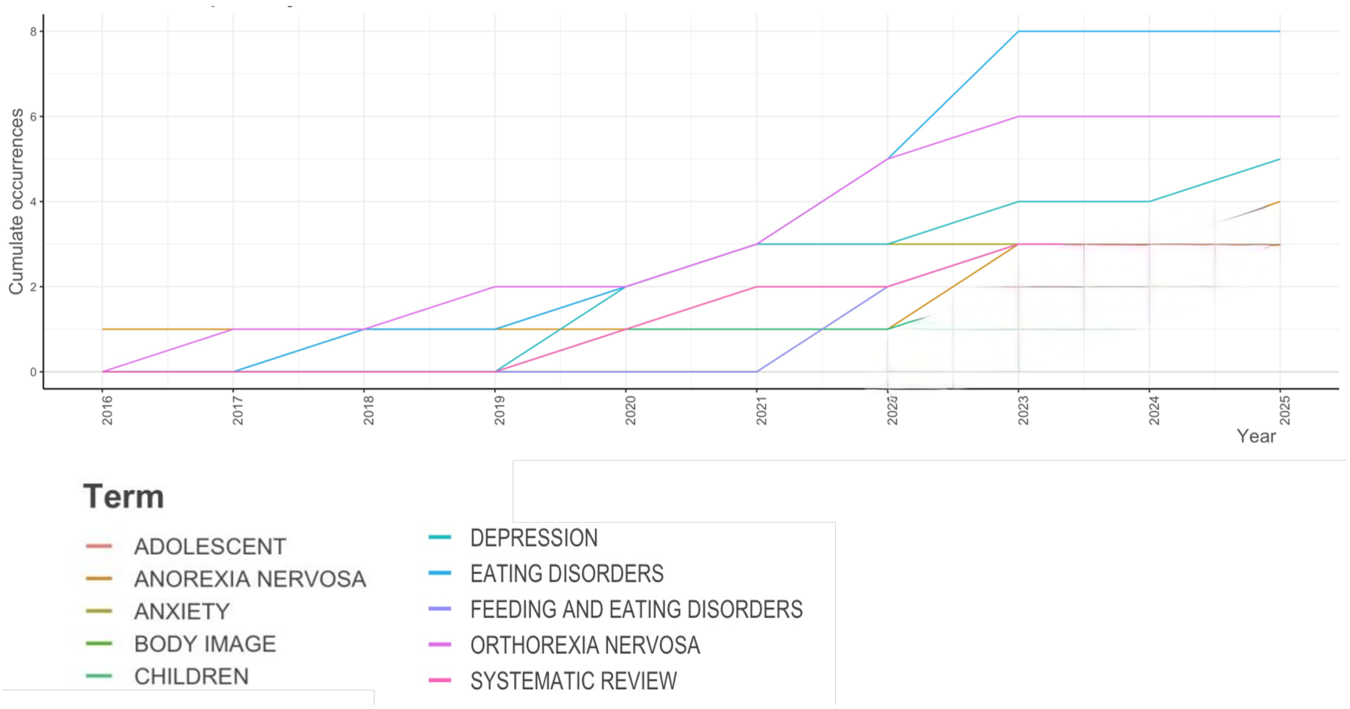



**Figure 7.** Evolution of keywords in publications on AI and Mental Health (2005-2025).

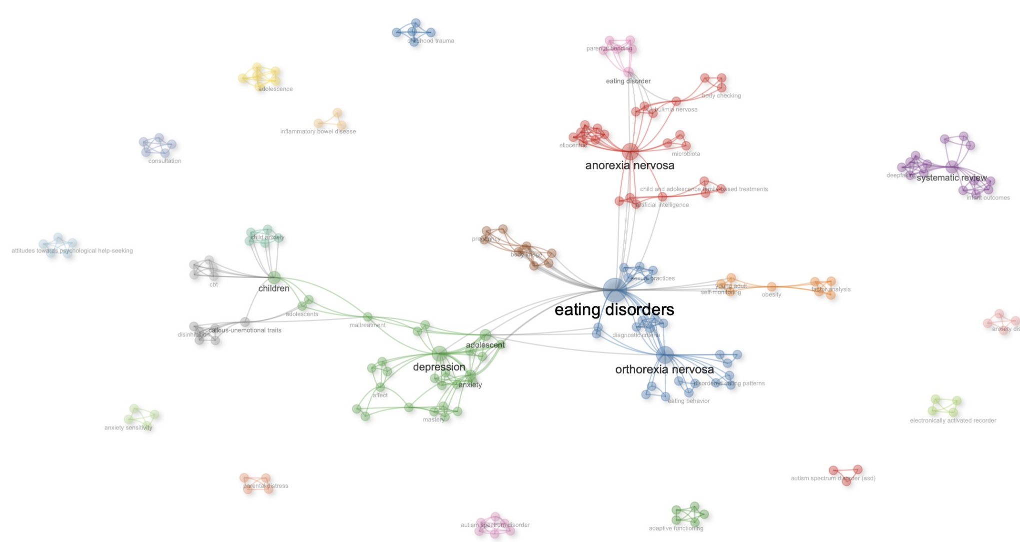



**Figure 8.** Keywords co-occurrence network.

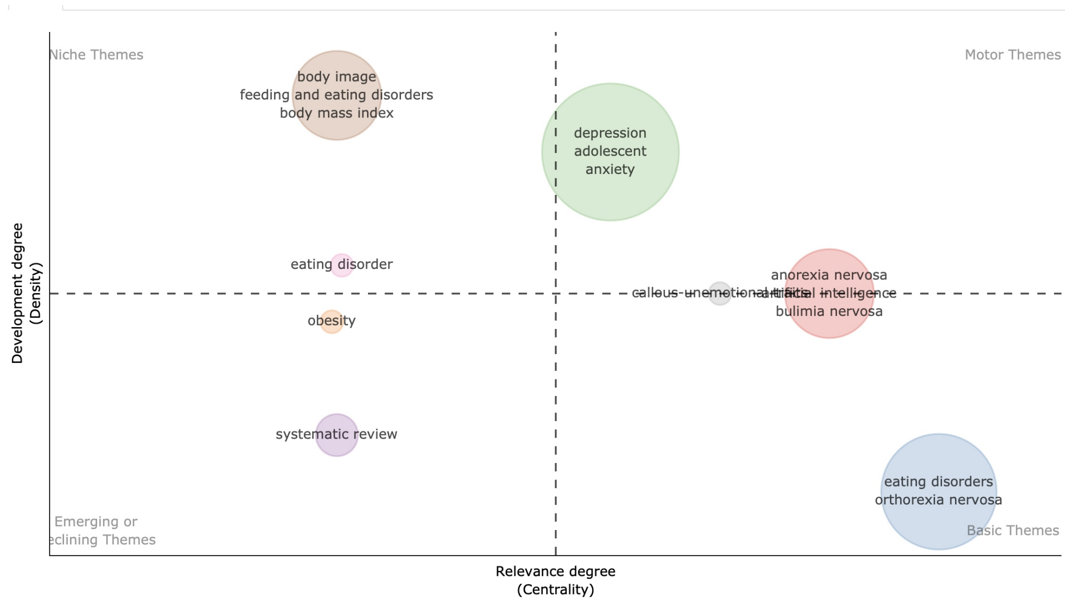



**Figure 9.** Thematic map of research themes in AI and mental health.

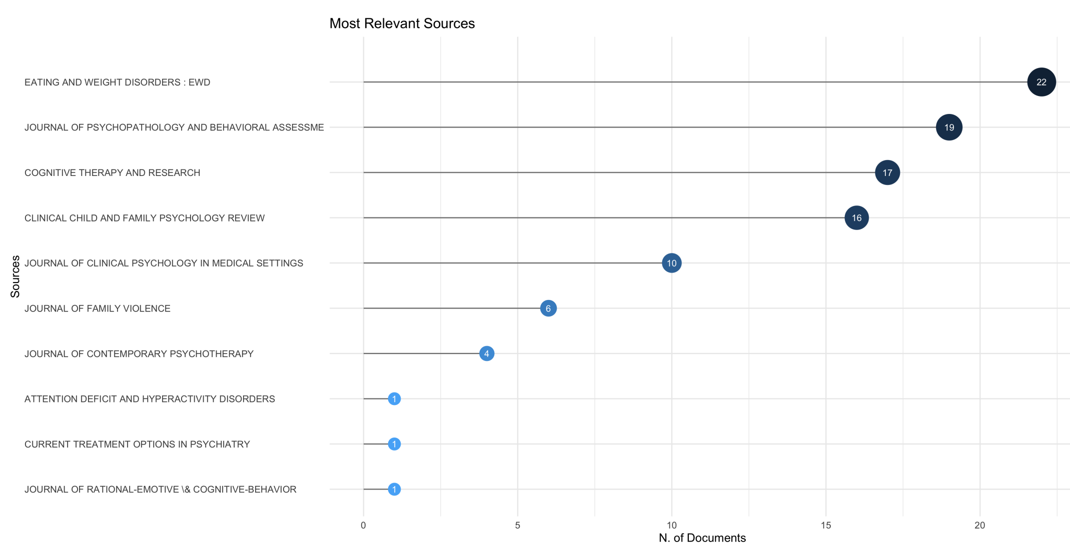



**Figure 10.** Most relevant journals publishing on AI and mental health (2005–2025).

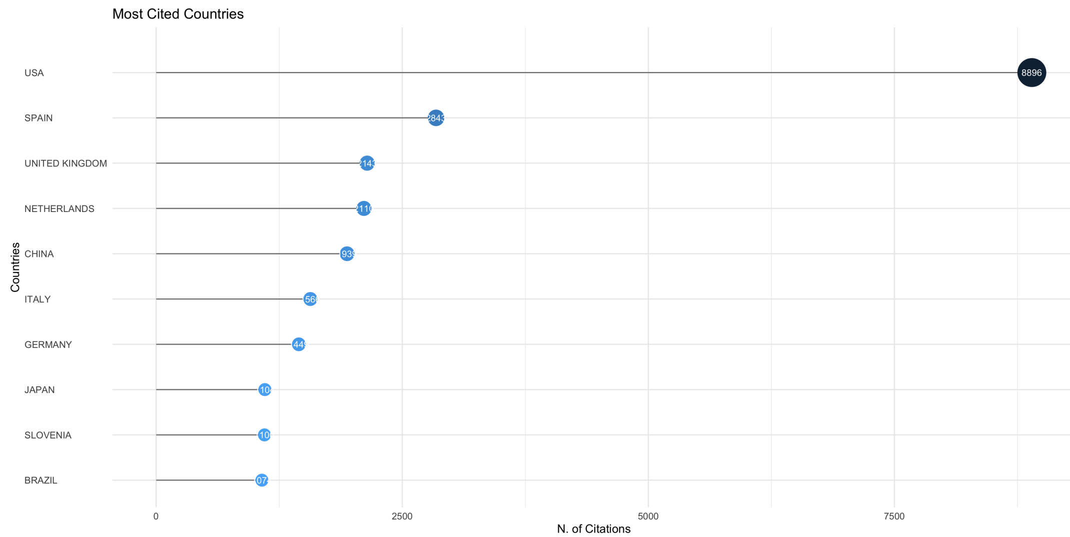



**Figure 11.** Top 10 most cited countries in AI and mental health research (2005–2025).

large research output, funding capacity, and concentration of top academic and scientific institutions. Spain ($n$ = 2,843), the United Kingdom ($n$ = 2,143), and the Netherlands ($n$ = 2,110) have demonstrated strong academic impact, showing their prominent contribution to establishing scholarly discourse in this area. China, while being one of the major contributors, demonstrates a comparatively lower citation count ($n$ = 1,939), suggesting a potential gap in international visibility or publication impact. Other countries, including Italy ($n$ = 1,564), Germany ($n$ = 1,447), Japan and Slovenia ($n$ = 1,104 each), and Brazil ($n$ = 1,070), round out the list, emphasizing the global significance and international collaborative nature of studies on AI in mental health. Overall, the citation data illustrates that the U.S. maintains a leadership position; however, the presence of several European and Asian countries in the top 10 reflects increasingly influential contributions to the developing domain of AI in mental health research.

## 5. Discussions and conclusions

This bibliometric study reveals a growing upward tendency in the number of publications on the application of AI in mental health from 2005 to 2025, with a notable increase in publication volume after 2020. This trend is consistent with global tendencies toward using digital health solutions after the COVID-19 pandemic, as reflected in recent reviews [13, 23]. The rise in scholarly output reflects the growing role of AI in diagnostic, preventive, and therapeutic mental health applications, as highlighted in multiple empirical studies and systematic reviews [3, 5, 9].

The prevalence of clinical topics such as eating disorders, depression, and anxiety is a key finding. Depression—its diagnosis and treatment—emerges as the most frequently addressed condition, with numerous studies leveraging AI's utility for early detection, suicide risk assessment, and digital intervention delivery [6, 12, 16, 17, 27]. The thematic map further validates the centrality and advancement of these areas, with "depression" and "anxiety" appearing in the motor themes category, indicating both centrality and maturity in this research area. Conversely, the increasing applicability of AI in addressing "eating disorders" and "orthorexia nervosa," identified as basic yet emerging themes, supports recent findings highlighting its diagnostic, preventive, monitoring, and therapeutic potential in addressing mental disorders [9, 10].

Keyword co-occurrence analysis emphasizes the following findings. The keyword "eating disorders" frequently co-occurs with "anorexia nervosa," "orthorexia nervosa," and "AI," mediated by co-words like "treatment" and "diagnostic criteria." This aligns with studies by Fang et al. [11], who demonstrate the use of AI-enhanced CBT and identify gender biases in AI tools, respectively. The frequent appearance of "systematic review" as an independent, high-frequency keyword highlights the evolving community's effort to consolidate fragmented results and evaluate the clinical reliability of AI tools [13, 28].

Despite the increasing volume of publications in the domain of AI in mental health, the co-authorship network remains notably fragmented, with limited cross-national and inter-institutional collaboration. Only a few authors, such as Jahrami and Ricca, occupy central positions in the network, indicating localized clusters of expertise rather than a unified global research effort. This reflects a concentration of expertise within isolated clusters, mirroring previous concerns about limited interdisciplinary and international integration in AI-based psychological research [14, 15]. The observed network structure suggests that although digital tools are globally applicable, the research remains largely restricted by national or institutional boundaries, potentially decreasing the generalizability and applicability of AI interventions in mental health.

Geographically, authors from the United States continue to dominate the field in terms of both productivity and citation impact. This trend aligns with other studies emphasizing the role of strong leadership, funding opportunities, developed research infrastructure, and early implementation of AI technologies [4, 20]. Nevertheless, countries such as the UK, Spain, the Netherlands, and China have also demonstrated impact within the field, supporting observations of international research efforts in AI-related suicide risk prediction [7, 16, 17, 21].

Finally, the growing interest in digital mental health diagnosis and prevention—such as chatbots, digital CBT, mood monitoring apps, and well-being platforms—is reflected in the keyword network and source analysis. Published papers have emphasized the essential role of AI in promoting psychological well-being, resilience, and stress management among children, students, and adolescents [24, 26, 22, 31]. Studies increasingly report that psychological interventions supported by AI are not only more accessible and cost-effective but also improve emotional resilience, reduce anxiety, enhance self-regulation, and mitigate social isolation [28, 29, 30].

In conclusion, the findings of this bibliometric study indicate both progress and limitations within the developing area of AI in mental health. While significant advancements have been made in terms of research output and thematic categorization, the domain would benefit greatly from more extensive and harmonized research methodologies, increased international collaboration, and the development of shared theoretical frameworks and technological standards. These steps are essential to fully harness the transformative potential of AI in advancing mental health care on a global scale.

### 5.1 Study limitations

This bibliometric review delivered significant findings as a synthesis of available data on publication patterns, productive authors, countries, sources, and collaboration networks. However, several limitations should be acknowledged. First, the database and query scope were limited to a selection of prestigious databases, including Crossref, Microsoft Academic, PubMed, OpenAlex, and CORE, with data extracted through Lens. Databases such as Scopus and Web of Science were not included in the search; therefore, non-indexed preprints and publications outside the selected databases may have been missed. Second, while the search query was carefully constructed to capture the most relevant studies, it relied on specific terminology. This approach may have excluded papers using alternative but related terms, thereby omitting some relevant studies.

Third, the analysis was limited to peer-reviewed publications written in English, leading to the exclusion of research published in other languages. Finally, a common concern in bibliometric studies is citation lag, in which newly published high-quality research may not have accumulated enough citations to fairly reflect its scholarly impact. This can result in the underrepresentation of recent contributions in the analysis.

### 5.2 Future research directions

The present bibliometric study highlights the necessity for further research into the effectiveness of utilizing AI to enhance the accuracy and efficiency of mental health detection, prevention, and professional treatment of various mental disorders. Such efforts may lead to improvements in psychological and overall well-being. The analysis of published research papers enabled us to formulate several significant recommendations for future studies within this domain.

Barros et al. [7] recommended conducting additional studies in varied environmental contexts, including educational institutions and primary health facilities, to assess the effectiveness of AI models with particular attention to the variable of age. Moreover, future research should explore the role of AI in detecting mental disorders by measuring responses across various categories of patients experiencing a wide range of emotions [3].

According to Liu et al. [6], personal preferences and individual differences should be considered when designing AI-supported prevention or treatment strategies for depression among college students [38, 39]. Da Fonseca, Bhat, and Kennedy [8] suggested integrating AI with electronic medical records (EMRs) to improve suicide prevention, while also emphasizing the need for further research on its safety and effectiveness.

Higgins et al. [1] noted that more studies are required to explore the impact of AI on treatment outcomes and the identification of missed care, as well as to better understand attitudes and beliefs related to the use of AI for promoting well-being [35, 37].

To conclude, further research is needed to understand the factors that increase acceptance of and engagement with AI in mental health contexts. This will require studies with larger and more diverse samples to assess whether AI applications can be generalized to broader populations.

Current research represents a significant step in the comprehensive exploration of publication trends related to the use of artificial intelligence in clinical psychology for the detection, prevention, and treatment of mental disorders from 2005 to 2025. Through bibliometric analysis, significant and relevant studies within the explored domains were systematically examined and synthesized, resulting in the establishment of a complex and comprehensive overview of the relationship between AI and mental health.

The identified research patterns and trends in the analyzed sources offer valuable insights for both clinical psychology practitioners and researchers, enabling them to be well-informed about existing knowledge, evidence, gaps, and emerging needs in the field. These findings can guide the development of more efficient interventions aimed at providing help to individuals struggling with mental health issues. Additionally, the analysis underscores the importance of further studies to estimate the practical impact of AI-based approaches in promoting psychological and overall well-being.

### 5.3 Conclusion

This bibliometric study highlights the rising scholarly interest in the use of artificial intelligence in mental health, particularly since 2020, indicating a significant shift likely influenced by the COVID-19 pandemic. The special focus on eating disorders, including emerging topics such as orthorexia nervosa, along with depression and adolescent mental health, suggests that these areas are well-established but still require greater clarity in research priorities to enable more targeted and effective development.

However, the fragmented author collaboration networks and the dominance of single-country publications reflect structural limitations in global research cooperation. While countries like the U.S. and Spain demonstrate high impact in terms of productivity and citation records, others such as China and Brazil face visibility challenges despite high productivity. These findings underscore the need for more integrative, international, and interdisciplinary approaches to strengthen the field.

Future research should also expand into underrepresented areas, such as digital diagnostics for autism, parental distress, and trauma-related disorders, to fully leverage AI's transformative potential in mental health care.

This study makes three major contributions. First, it provides a systematic mapping of the field's intellectual structure by identifying leading authors, institutions, sources, and countries, facilitating deeper insights into global collaboration patterns. Second, it uncovers thematic transitions from early-stage technological experimentation to the integration of AI into mental health diagnosis and interventions, demonstrating the theoretical advancement of the domain. Third, it serves as a reference point for academics, practitioners, researchers, and technology developers aiming to identify emerging areas for development, address knowledge gaps, and foster interdisciplinary partnerships.

## Ethics statement

This study did not involve human participants or animals and therefore did not require ethical approval.

## CRediT authorship contribution statement

**Ivanna Shubina**: conceptualization, methodology, software, validation, formal analysis, writing – original draft and final version, review & editing, and visualization. **Adrian Jarema Dzido**: software support, data processing, and assistance with data interpretation.

## Declaration of generative AI and AI-assisted technologies in the writing process

The authors declare that no generative AI or AI-assisted technologies were used in the writing of this manuscript. All text and figures were produced with the help of the Bibliometrix software. ChatGPT was used only for proofreading purposes.

## Declaration of competing interest

The author declares no competing financial or personal interests that could have appeared to influence the work reported in this paper.

## References


[1] O. Higgins, B. L. Short, S. K. Chalup, and R. L. Wilson, "Artificial intelligence (ai) and machine learning (ml) based decision support systems in mental health: An integrative review," *International Journal of Mental Health Nursing*, vol. 32, no. 4, pp. 966–978, 2023.

[2] G. Ramshaw, A. McKeown, R. Lee, A. Conlon, D. Brown, and P. J. Kennedy, "Introduction of technology to support young people's care and mental health—a rapid evidence reviews," *Child and Youth Care Forum*, vol. 52, no. 3, pp. 509–531, Jun 2023.

[3] M. L. Joshi and N. Kanoongo, "Depression detection using emotional artificial intelligence and machine learning: A closer review," *Materials Today: Proceedings*, vol. 58, pp. 217–226, 2022.

[4] S. W. Kelley, C. N. Mhaonaigh, L. Burke, R. Whelan, and C. M. Gillan, "Machine learning of language use on twitter reveals weak and non-specific predictions," *NPJ Digital Medicine*, vol. 5, no. 1, p. 35, 2022.

[5] S. Zhou, J. Zhao, and L. Zhang, "Application of artificial intelligence on psychological interventions and diagnosis: an overview," *Frontiers in Psychiatry*, vol. 13, p. 811665, 2022.

[6] X. Q. Liu, Y. X. Guo, W. J. Zhang, and W. J. Gao, "Influencing factors, prediction and prevention of depression in college students: a literature review," *World Journal of Psychiatry*, vol. 12, no. 7, pp. 860–873, 2022.

[7] J. Barros, S. Morales, A. García, O. Echávarri, R. Fischman, M. Szmulewicz, others, and A. Tomicic, "Recognizing states of psychological vulnerability to suicidal behavior: a bayesian network of artificial intelligence applied to a clinical sample," *BMC Psychiatry*, vol. 20, pp. 1–20, 2020.

[8] M. da Fonseca, G. Maffei, R. Moreno-Bote, and A. Hyafil, "Mood and implicit confidence independently fluctuate at different time scales," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 23, no. 1, pp. 142–161, 2023.

[9] J. Fardouly, R. D. Crosby, and S. Sukunesan, "Potential benefits and limitations of machine learning in the field of eating disorders: current research and future directions," *Journal of Eating Disorders*, vol. 10, no. 1, p. 66, 2022.

[10] D. S. Noguero, D. Ramírez-Cifuentes, E. A. Ríssola, and A. Freire, "Gender bias when using artificial intelligence to assess anorexia nervosa on social media: data-driven study," *Journal of Medical Internet Research*, vol. 25, p. e45184, 2023.

[11] F. X. Fang, D. T. Ng, J. K. Leung, and S. K. Chu, "A systematic review of artificial intelligence technologies used for story writing," *Education and Information Technologies*, vol. 28, no. 11, pp. 14 361–14 397, Nov 2023.

[12] N. C. Jacobson and B. Feng, "Digital phenotyping of generalized anxiety disorder: using artificial intelligence to accurately predict symptom severity using wearable sensors in daily life," *Translational Psychiatry*, vol. 12, no. 1, p. 336, 2022.

[13] Y. Shan, M. Ji, W. Xie, R. Li, X. Qian, X. Zhang, and T. Hao, "Interventions in chinese undergraduate students' mental health: Systematic review," *Interactive Journal of Medical Research*, vol. 11, no. 1, p. e38249, 2022.

[14] M. M. Maslej, S. Kloiber, M. Ghassemi, J. Yu, and S. L. Hill, "Out with ai, in with the psychiatrist: a preference for human-derived clinical decision support in depression care," *Translational Psychiatry*, vol. 13, no. 1, p. 210, 2023.

[15] C. Blease, A. Kharko, M. Annoni, J. Gaab, and C. Locher, "Machine learning in clinical psychology and psychotherapy education: a mixed methods pilot survey of postgraduate students at a swiss university," *Frontiers in Public Health*, vol. 9, p. 623088, 2021.

[16] N. Kirlic, E. Akeman, D. C. DeVille, H. W. Yeh, K. T. Cosgrove, T. J. McDermott, others, and R. L. Aupperle, "A machine learning analysis of risk and protective factors of suicidal thoughts and behaviors in college students," *Journal of American College Health*, vol. 71, no. 6, pp. 1863–1872, 2023.

[17] M. Macalli, M. Navarro, M. Orri, M. Tournier, R. Thiébaut, S. M. Côté, and C. Tzourio, "A machine learning approach for predicting suicidal thoughts and behaviours among college students," *Scientific Reports*, vol. 11, no. 1, p. 11363, 2021.

[18] R. Rois, M. Ray, A. Rahman, and S. K. Roy, "Prevalence and predicting factors of perceived stress among bangladeshi university students using machine learning algorithms," *Journal of Health, Population and Nutrition*, vol. 40, pp. 1–12, 2021.

[19] M. H. Huang and R. T. Rust, "Artificial intelligence in service," *Journal of Service Research*, vol. 21, no. 2, pp. 155–172, 2018.

[20] T. M. Fonseka, V. Bhat, and S. H. Kennedy, "The utility of artificial intelligence in suicide risk prediction and the management of suicidal behaviors," *Australian & New Zealand Journal of Psychiatry*, vol. 53, no. 10, pp. 954–964, 2019.

[21] P. Buddhitha and D. Inkpen, "Multi-task learning to detect suicide ideation and mental disorders among social media users," *Frontiers in Research Metrics and Analytics*, vol. 8, p. 1152535, 2023.

[22] B. Inkster, M. Kadaba, and V. Subramanian, "Understanding the impact of an ai-enabled conversational agent mobile app on users' mental health and wellbeing with a self-reported maternal event: a mixed method real-world data mhealth study," *Frontiers in Global Women's Health*, vol. 4, p. 1084302, 2023.

[23] T. Chen, "Investigating the mental health of university students during the covid-19 pandemic in a uk university: a machine learning approach using feature permutation importance," *Brain Informatics*, vol. 10, no. 1, p. 27, 2023.

[24] S. Sabour, W. Zhang, X. Xiao, Y. Zhang, Y. Zheng, J. Wen, J. Zhao, and M. Huang, "A chatbot for mental health support: exploring the impact of emohaa on reducing mental distress in china," *Frontiers in Digital Health*, vol. 5, p. 1133987, 2023.

[25] A. Yorita, S. Egerton, C. Chan, and N. Kubota, "Chatbots and robots: a framework for the self-management of occupational stress," *ROBOMECH Journal*, vol. 10, no. 1, p. 24, 2023.

[26] K. Tanaka, M. Hayakawa, C. Noda, A. Nakamura, and C. Akiyama, "Effects of artificial intelligence intervention on alleviating distress and fear in children," *Child and Adolescent Psychiatry and Mental Health*, vol. 16, no. 1, p. 87, 2022.

[27] Y. He, D. Liu, R. Guo, and S. Guo, "Information cocoons on short video platforms and its influence on depression among the elderly: a moderated mediation model," *Psychology Research and Behavior Management*, pp. 2469–2480, 2023.

[28] S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, "Recognizing depression from twitter activity," in *Proceedings 33rd Annual ACM Conference on Human Factors in Computating Systems*, 2015, pp. 3187–3196.

[29] S. Mouchabac, P. Leray, V. Adrien, F. Gollier-Briant, and O. Bonnot, "Prevention of suicidal relapses in adolescents with a smartphone application: Bayesian network analysis of a preclinical trial using in silico patient simulations," *Journal of Medical Internet Research*, vol. 23, no. 9, p. e24560, 2021.

[30] H. Li, R. Zhang, Y. C. Lee, R. E. Kraut, and D. C. Mohr, "Systematic review and meta-analysis of ai-based conversational agents for promoting mental health and well-being," *NPJ Digital Medicine*, vol. 6, no. 1, p. 236, 2023.

[31] I. Shubina, "Scientific publication patterns of interactive mobile technologies for psychological, social, medical and business interventions for mental and physical health: bibliometric analysis," *International Journal of Interactive Mobile Technologies*, vol. 15, no. 21, p. 4, 2021.

[32] S. Jeong, L. Aymerich-Franch, K. Arias, S. Alghowinem, A. Lapedriza, R. Picard, others, and C. Breazeal, "Deploying a robotic positive psychology coach to improve college students' psychological well-being," *User Modeling and User-Adapted Interaction*, vol. 33, no. 2, pp. 571–615, 2023.

[33] S. Jolley, J. Lord, D. Plant, E. Wood, K. Bracegirdle, S. Browning, others, and K. James, "A cluster randomised, 16-week, parallel-group multicentre trial to compare the effectiveness of a digital school-based cognitive behavioural resilience/wellbeing-building intervention ... : a study protocol to the 'cues for schools' trial," *Trials*, vol. 24, no. 1, p. 253, 2023.

[34] C. Ehrlich, S. E. Hennelly, N. Wilde, O. Lennon, A. Beck, H. Messenger, others, and E. L. Davies, "Evaluation of an artificial intelligence enhanced application for student wellbeing: pilot randomised trial of the mind tutor," *International Journal of Applied Positive Psychology*, vol. 9, no. 1, pp. 435–454, 2024.

[35] I. Shubina, O. Plakhotnik, and O. Plakhotnik, "Professional education and technology usage for establishing methodological competence among future professors: bibliometric analysis," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 19, pp. 235–250, 2021.

[36] I. Shubina, "Scientific publication patterns of cognitive-behavioral therapy and interactive mobile technologies for treatment of eating disorders: a bibliometric analysis," *Journal of Positive Psychology and Wellbeing*, vol. 6, no. 2, pp. 570–583, 2022.

[37] ——, "Scientific publication patterns of systematic reviews on psycho-social interventions improving well-being: a bibliometric analysis," *Interactive Journal of Medical Research*, vol. 11, no. 2, p. e41456, 2022.

[38] ——, "Scientific publication patterns of psychological capital for organizational change during utilizing positive psychology interventions," *Journal of Positive School Psychology*, pp. 4676–4683, 2022. [Online]. Available: https://www.journalppw.com/index.php/jpsp/article/view/7309/4778

[39] B. C. de Caux and I. Shubina, "A longitudinal analysis of mental health and wellbeing among graduate students around the world: The intra-individual and societal impact of a pandemic," *International Journal of Well-being*, vol. 13, no. 3, 2023. [Online]. Available: https://doi.org/10.5502/ijw.v13i3.2775

[40] M. S. Jalali, S. Razak, W. Gordon, E. Perakslis, and S. Madnick, "Health care and cybersecurity: bibliometric analysis of the literature," *Journal of Medical Internet Research*, vol. 21, no. 2, p. e12644, 2019.

[41] L. Cui, "Rating health web sites using the principles of citation analysis: A bibliometric approach," *Journal of Medical Internet Research*, vol. 1, no. 1, p. e652, 1999.

[42] H. Esfahani, K. Tavasoli, and A. Jabbarzadeh, "Big data and social media: A scientometrics analysis," *International Journal of Data and Network Science*, vol. 3, no. 3, pp. 145–164, 2019.

Journal of Artificial Intelligence and Computing Applications (2025) 3(1): 38–48

Journal of Artificial Intelligence and Computing Applications

*Research article*

# A methodology for the development of serious games for the cognitive stimulation of elderly people with mild cognitive impairment

**Luisa Andrea Morales-García**[1,*], **Luis Enrique Sucar**[1], **Delia Irazú Hernández Farias**[2], **and Alberto L. Moran Y Solares**[2]

[1]Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, Mexico.
[2]Universidad Autónoma de Baja California, Ensenada, Mexico.

**ABSTRACT**

Mild cognitive impairment affects many older adults and can lead to severe dementia. Early detection and intervention are key to slowing its progression. Serious games offer a promising way to stimulate cognitive function. However, there is a lack of clear methodologies for developing effective serious games for cognitive stimulation. In this paper, we introduce a methodology for developing serious games to help people with mild cognitive impairment. We also include a case study of this methodology through the development of a serious game that underwent usability testing with older adults. The obtained results provide evidence that, by following the proposed methodology, it is possible to develop serious games that are well received by the target population.

**Keywords:** serious games, cognitive impairment, cognitive stimulation

## 1. Introduction

Mild Cognitive Impairment (also denoted as MCI) is characterized by a cognitive impairment that causes minimal impact on instrumental activities of daily living (e.g., language, visuospatial skills, executive functions, etc.) [1]. Its early identification could serve to initiate treatment and take into account the appropriate measures to avoid its further progression [2]. Rehabilitation or cognitive stimulation can help improve the cognitive abilities of people with this condition [3]. Nowadays, there are several proposals for performing appropriate cognitive stimulation; among the most outstanding are those related to the use of technology [4]. Within digital cognitive stimulation and rehabilitation, there is evidence that the use of Serious Games (SG) [1] allow the development of the brain's executive functioning and, in turn, help to improve cognitive skills with the potential to reduce problems caused by cognitive impairment [6]. Serious games are considered a promising solution that provides non-drug-based rehabilitation and treatment, aiming to satisfy, maintain, and even restore the patient's cognitive status through cognitive stimulation [7]. However, despite the importance of using serious games in the field of cognitive stimulation, there is a lack of a methodology that establishes in detail the steps to follow for the development of a SG oriented towards

[1]A serious game allows for the implementation of activities based on real-life scenarios; are intended to teach, as well as to transmit skills and information [5].

E-mail address: luisa.morales@inaoep.mx
https://doi.org/10.5281/zenodo.16916163

cognitive stimulation [6].

According to Garcia-Martinez et al. [3], the development of serious games that meet the requirements of having a medical perspective, usability, and user-based experience remains an unsolved problem. Attempting to solve this problem, they developed the COMFeeDY framework (Concepts, Objects, Mechanics, Feedback, and Dynamics). Having in mind to gain a better understanding of the relationship between older adults and the various technologies that can be used for cognitive stimulation, Palumbo [4] presented a survey to identify which are the most commonly used interactive devices for serious games, which are the cognitive functions that have been addressed by serious games supported by interactive technologies as well as the improvements on them after using these kinds of games. Lau & Agius [6] propose MCI-GaTE (MCI-Game Therapy Experience), which can be used for the implementation of SG that function as a tool capable of implementing physical and cognitive rehabilitation. This framework consists of four sectors which serve as a guide for the design and development of games: a) A player profile, b) Main elements that support recreational activities, c) Therapeutic components that support cognitive and physical rehabilitation through the implementation of tasks and scenarios that are tailored to the player's abilities, and d) Motivational components. MCI-GaTE was assessed by therapists who indicated that the framework has great potential for the design and implementation of therapeutic experiences. However, some aspects were highlighted as missing, such as the lack of times between game sessions, or that it is focused primarily on memory rehabilitation in people with cognitive problems, leaving aside all other conditions involved with MCI.

Attempting to contribute to solve this problem, we propose a methodology for designing serious games for cognitive stimulation in elderly people with cognitive problems. Our proposal includes aspects that have not been taken into account in the existing literature that are important to have a clear and comprehensive methodology. In addition, we also present a serious game denoted *Adventures in the corral, heroes of the flock*, which was designed and developed following the proposed methodology. For evaluation purposes, we asked a group of elderly people to interact with the serious game and report the results of a usability evaluation based on the player's experiences.

This paper is organized as follows. Section 2 introduces the methodology for the development of serious games for cognitive stimulation of the elderly with mild cognitive impairment. Section 3 describes the development of the case study based on the proposed methodology. Section 4 presents the results obtained from the usability study. Finally, Section 5 summarizes the conclusions and future work.

## 2. Proposed methodology

This section explains in detail each step of the proposed methodology to design serious games that can help to cognitively stimulate elderly people with mild cognitive impairment.

### 2.1 Identification of conditions related to mild cognitive impairment

The type of cognitive impairment to be addressed must be defined: *Mild Cognitive Impairment*, *Medium Cognitive Impairment or Mild Dementia*, *Medium Cognitive Impairment (or Prolonged Dementia)*, or *Severe Cognitive Impairment (or Severe Dementia)* [8]. Afterwards, the type of condition that is sought to be stimulated must be identified. Some of the conditions that are caused by MCI are *concentration*, *perception*, *communication*, *orientation*, *motor coordination*, *conceptualization*, *language*, *judgment*, *calculation*, and *memory* [9, 10, 11, 12].

### 2.2 Identification of cognitive stimulation activities for people with mild cognitive impairment

The activities to be developed in the SG must be identified. These activities will be part of the participants' cognitive stimulation. With this in mind, Table 1 shows some examples of activities that can be implemented within the SG and that can help stimulate specific conditions.

### 2.3 Identification of the type of participants

Participants must be chosen taking into account the classification of cognitive impairment and the condition(s) to be addressed. However, if it is unknown whether or not a participant has cognitive impairment, as a first point, some tests must be applied to determine his or her condition. For this, there are several tests to identify cognitive impairment such as: Mini Mental State Examination, Montreal Cognitive Assessment, Clock Drawing Test, Mini Cognitive Examination, and Five Words Test, among others.

### 2.4 Identification of inclusive design approaches

It is important to define the interaction techniques in terms of navigation and selection. Navigation and selection allow the participant to interact with the User Interface(UI) elements of the virtual environment of the SG through the use of hardware; these must be easy to use and learn in order to reduce the amount of effort, taking into account the limitations of older adults with mild cognitive impairment. Some examples of hardware used for interaction are: mouse, keyboard, touch screens, joystick, Kinect, and Wii [21]. Equally important at this stage, is the definition of the UI elements of the virtual environment of the SG, including: i) **Text content** should be clear, precise, and engaging. Besides, it should be written in a large font, preferably black, and avoid abbreviations. Technical and scientific language should also be avoided. ii) **Basic elements**

**Table 1.** Activities that can be implemented for stimulating specific conditions.

| Condition | Activities |
|---|---|
| Concentration | Finding identical figures, maze, word search, numbers or symbols, reading with simple activities, among others [13]. |
| Perception | Identifying identical objects, word and number searches, relate objects, describe images, etc. [14]. |
| Orientation | Locating oneself in time and space, identifying the current season, relating objects, writing personal information, etc. [15]. |
| Motor coordination | Mazes, completing drawings, puzzles, associating objects with their functions, etc. [13]. |
| Conceptualization | Identifying objects, giving meaning to an image, identifying geometric shapes, completing images, etc. [16]. |
| Language and speech | Forming and completing words, completing sentences, ordering phrases, etc. [17]. |
| Judgment | Associating objects, numerical and object series, joining related words, identifying patterns of different images, etc. [18]. |
| Calculation | Addition, chain operations, mental calculation, etc. [19]. |
| Memory | Memorizing details or images, making life books, etc. [20]. |

such as buttons, prompts, objects, etc., should be visible, large, and prominent. iii) **Images** they must be relevant to the context of the game being implemented and must also make sense. iv) **Color palette**, it is recommended to use a high-contrast color palette, as this can help improve visibility. Furthermore, it is also recommended not to include distractions in the virtual environment.

### 2.5 Design of competencies for cognitive stimulation

It is necessary to define how the activities selected (Section 2.2) will be implemented. Competencies must be defined so that the implementation of cognitive stimulation activity yields favorable results in the treatment of cognitive impairment.

### 2.6 Conceptualization of the type of game

The design characteristics of the game must be defined, as well as some other general aspects such as classification (pure progressive games, pure emergent games, etc.), genre (adventure, horror, fear, platforms, etc.), type of technology, mechanics, expected results, values, player effort, attachment to the result, negotiable consequences, and area of application. Some questions that can help identify such features are: *What is this game about?*, *How do I play?*, *What is the main idea or concept of the game?*, *Who are the characters that are going to appear in the story?*, *What are going to be the mechanics of the game*, and similar [22, 23, 24]. In addition, the type of approach to the game needs to be determined. Among the approaches to be considered, there are some focused on a specific theme, on the mechanics, dynamics, and emotional responses, or those that do not have limitations [25].

### 2.7 Narrative structuring

The narrative must have a structured process. To achieve this structure, it is necessary to define each of the chapters (levels) with their respective scenes. Once the story is defined, the dialogues or actions that will make it possible must be developed. After this, and as a final step, the relationship between the chapters, scenes, stories, dialogues, or actions with the cognitive stimulation activities proposed in phase two of the methodology must be defined. A narrative line can follow one of the types of stories: *linear stories*, *branching stories*, *open-ended stories*, *thematic settings*, among others [25].

Another aspect to keep in mind when creating a story is including and developing plots. There are three different types of plots: *i) Classic* based on the story of the main character versus the villain; *ii) Relationships* where there is a kind of relation (friendship, romantic, hatred, etc.) between the characters; and *Transformation* which narrates the internal development of characters through the need to overcome their goals. Afterwards, the script describing the structure of how the events will be developed must be defined. For doing this, some types of structure can be considered:

- **Classical**, composed by four elements: a) *Introduction* (describing the context of the world in which the story takes place, the game dynamics, and the rules must be provided), b) *Plot point* when something occurs and changes the course of how the main story is presented), c) *Knot* (it is the main part of the story, where the goal is to escalate the problem to return to the initial or future calm point), d) *Climax* (which seeks to confront each of the problems presented).

- **Modern**, seeks to explore the character's internal conflicts, whether their hidden desires, fears, insecurities, dilemmas, or some type of psychological or emotional conflict. Its main objective is to develop the story by focusing on the protagonist's internal journey, self-discovery, or the resolution of internal conflicts.

Once the script is finished, the dialogues must be defined, which are all those arguments that will be said by the characters within the game or are all those texts that will be shown during the game [26]. Besides, the characters in the game need to be described; there are four kinds of characters: *protagonist*, *antagonist*, *secondary characters*, and *non-playable character*[2], which are those within the game that are controlled by artificial intelligence, rather than by a player.

At this stage, it is necessary to design the chapters that comprise the process by which environments, challenges, and experiences are structured and developed, which players must experience as they progress through the game [27]. It is also required to define the position of the game elements and the mobility of objects.

#### *2.7.1 Labeling and evaluation of cognitive activities.*

As a final step in developing this stage, the story situations in which the cognitive activities chosen in point II of the methodology will be developed must be defined, and the method of implementation must also be specified. In this final part, the moments, characters, objects, dialogues, etc., that will be involved in order to include the cognitive activities within the story, chapters, and scenes must be specified.

### 2.8 Adaptive structuring and personalization design

The user profile must be defined, including the patient's personal data (social and morbid data [12, 28]), and a record of all the activities performed, as well as their times and scores. In addition, it is important that the user profile allows the inclusion of personalized characteristics based on their abilities, needs, preferences, and condition, ensuring an accessible and effective interaction [29]. Finally, the user profile can also record patient preferences so that the game elements are tailored to these needs. In this sense, elements, decorations, objects, animals, etc., can be added or removed, text can be enlarged, audio elements can be increased in volume, audio can be slowed down, etc.

#### *2.8.1 Labeling and identifying inclusive design approaches*

The inclusive aspects of gameplay, narrative, and game mechanics must be defined, but this time taking into account the final game structure. This means that if it is decided to use a gripper for game interaction, it must be specified which movements or actions can be performed with it. Finally, it is necessary to define how adaptability will be implemented within the game. Adaptability within the context of rehabilitation or stimulation ensures that the best decisions are made to guarantee intelligent rehabilitation, and this is achieved through the control of auditory, haptic, visual, and other aspects.

#### *2.8.2 Inclusion of artificial intelligence*

Artificial intelligence provides certain tools that help within the mechanics, narratives, sounds, history, scenarios, etc. [30]. Some of these tools are the following [31]: movement techniques, complex movements, decision-making, and adaptation to the user.

### 2.9 Design of orientation and motivation strategies

The strategies that will be used to guide the players through the first steps of the SG's gameplay need to be defined. Also, the points and rewards system that will be implemented within the SG's gameplay must be explained, taking into account the narrative and mechanics of the SG. It is highly recommended to design a **Tutorial** to help the player become familiar with the game mechanics. Tutorials are the levels prior to the main game and the basis of it, where the player is expected to become familiar with the gameplay of the SG through the interaction with objects using the main commands. These should be presented naturally and simply, in addition to not being immersive, in order not to affect the gameplay of the main game and not to ruin the player's motivation [25]. Concerning the motivation strategies, it is important to specify how the rewards will be granted following the narrative and mechanics of the game. As players develop a greater attachment to the outcome or reward, the more interested and enthusiastic they become about the game. Similarly, the greater the commitment and effort that a player makes to complete a challenge, the greater the emotional burden related to the result and the rewards [23].

## 3. Evaluation of the proposed methodology

To validate the proposed methodology, a case study was conducted. In particular, it was used for the design and implementation of a serious game named as *Adventures in the corral, heroes of the flock*. Next, we describe the different stages of the methodology applied to this case.

1. **Identification of conditions related to mild cognitive impairment:** The primary focus will be on single-domain non-amnestic mild cognitive impairment. Particularly, the goal is to stimulate only the area of **concentration**.

[2]https://dictionary.cambridge.org/es/diccionario/ingles/npc

2. **Identification of cognitive stimulation activities for people with mild cognitive impairment:** Identification of identical objects.

3. **Identification of the type of participants:** Inclusion criteria:

   - No tests will be performed to determine whether or not the patient suffers from any type of Mild Cognitive Impairment.
   - Age range: People over 60 years old.
   - Sex : Indistinct.
   - Level of education: Not an important factor.

   Exclusion criteria:

   - Participants with Parkinson's disease will be excluded, as this condition can interfere with game results due to the relationship between control and the participant's spontaneous movements. Participants with depression will also be excluded, due to the fact that according to [32] this kind of rehabilitation has less positive impact on them.

4. **Identification of inclusive design approaches:**

   - ***Navigation and selection:*** Two different hardware components were used for navigation and selection: a video game controller and a Bluetooth VR remote controller. In both cases, the corresponding joystick can move the elements in four directions: up, down, left, and right. Further, when using the video game remote controller for selection, it is required to use the button located at the bottom of the set of buttons positioned on the right side of the device; and when using the VR Bluetooth remote controller, the"C" key is the one used for selecting an element. Figure 1 shows both devices and their respective parts used for navigation (in green) and for selection (in red).
   - ***Textual content***. The SG includes instructions that are short, clear, and concise, using informal language; in addition, text will be written in large fonts.
   - ***Basic elements within the game***. All the visual elements of the game are placed at a considerable size, so that they are visible at all times.
   - ***Visual elements and images.*** Visual elements or images used within the game (animals, objects, characters) are represented as close to reality as possible.
   - ***Use of colors***. The colors used are as close to real-life elements as possible. For elements such as the menu, dark background colors and light text colors are used.

5. **Design of competencies for cognitive stimulation**. The main objective of this SG is to provide stimulation in the area of concentration for people with MCI who have concentration problems; however, it can also help stimulate *perception* and *memorization*. Regarding the design of the skills, the goal is for the person to be able to stimulate the area of concentration by carrying out an activity of identifying identical objects. The main theme of the game is for the player to be able, in the first instance, to identify the animal corresponding to the instruction provided and, subsequently, to identify the correct corral thanks to this same activity. Likewise, as mentioned above, this game can also stimulate the area of memorization and perception, which is carried out through the presence of short and clear instructions that will be said only once by voice and that will indicate to the player the order in which the animals must be taken to the corral; in this way, the player must read an instruction, identify the animal that the player wants to herd, identify the corral to which it belongs, and take it there.

6. **Development of the conceptualization of the game type**. This SG is going to be a pure progressive game; it will have an adventure genre, and it will have an "A" rating (this means that its content will be suitable for all audiences). It is about a farmer who seeks to return all the animals that have escaped to their respective barnyard; to do this, the player must herd each one of them until none are left free. To correctly implement the game play within *Adventures in the corral, heroes of the flock*, the following rules and results must be followed:

   (a) *Rule:* The animals must be herded in the order indicated in the instructions that will be displayed on the screen.
   *Result*: If the player herds the animals following the instructions, they can move on to the next step, which is to attempt to lasso the animal. Otherwise, they will not be able to lasso the animal because it will not be active.

   (b) *Rule*: To herd an animal, the player must approach it and throw a lasso by pressing the corresponding key.
   *Result*: For the player to activate the lasso function, the player must be close enough to the animal. Otherwise, even if the animal matches the order of the instructions, the lasso will not be able to be thrown.


**Figure 1.** Devices used for in-game interaction. The parts used for navigation (green) and selection (red) are highlighted.

(c) *Rule*: Once the animal is connected with the lasso, the player must walk to the animal's correct pen.
*Result*: If the player successfully moves the animal close to its pen, the player will be able to move on to the next step, which is to put it in the pen. Otherwise, the player will not be able to put the animal in the pen, as the gameplay of the SG does not allow animals to be placed in incorrect pens.

(d) *Rule*: Once the player is close enough to the corresponding corral, the player must press a key to release the animal and have it go into its corral.
*Result*: If the player is positioned close enough to the pen of the linked animal, the player must press the correct key to allow the animal to enter the pen. Otherwise, the player will not be able to leave the animal inside its pen.

(e) *Rule*: The game will have a total of 5 levels, each one with a different time limit in seconds ranging from 300 up to 540. All animals must be properly housed in their pens within the established time limit.
*Result*: If all the animals are placed in their pens within the allotted time, the game is won; otherwise, the game is lost.

There are two ways in which a game can finish:

(a) *Option 1*: The game ends when the player has completed all five levels of the game.
*Result:* The player wins the game.

(b) *Option 2*: When the player does not finish putting the animals in their respective pens in time and once the game-end screen appears, click on "Exit" the game ends.
*Result*: The player loses the game and can choose to play again or exit the game.

Figure 2 presents a graphical representation of the elements implemented within the rules and the corresponding results.

**Player efforts**.

- The game has five levels, which the player will progress through automatically; that is, each time they complete a level, they will move on to the next. The difficulty of the gameplay increases with each level.
- Each time the player links an animal, the player will receive 50 points.
- Each time the player places the animal in its corral, the player will receive 100 points.
- In each level, the player can receive a maximum of points: 750 for levels one and two, 1050 for the third one, and 1200 and 1500 for levels fourth and fifth, respectively.

**Negotiation consequences**.

- Points will be used as in-game rewards, which will increase as the player follows the instructions.
- Celebratory sounds will be used as a reward each time the player follows the instructions within the game.

7. **Narrative structuring**.

- **Story**. The story is linear, and the SG starts with the following description: "*One day, a farmer was putting all the animals on her farm in their respective pens; however, due to her carelessness, where the player didn't lock the pens, most of the animals managed to escape. Now the farmer must put all the escaped animals in their respective pens, and the player must do it quickly, because if night falls and the animals are not kept safe, they can be devoured by predators. This is because every night, wolves come*

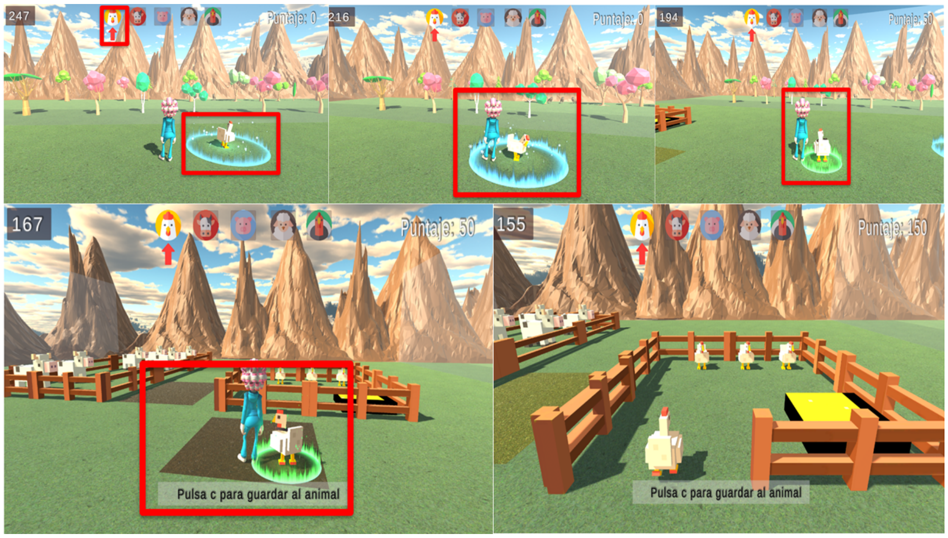



**Figure 2.** From top left, clockwise. First box: The instructions indicate grabbing the chickens, so it is necessary to get close to them. Second box: To grab the chicken, the player must get close enough to it; in this case, the player must be within the circle surrounding it. Third box: Once the player is close enough and presses the correct key, the circle surrounding the chicken turns green. Fourth box: Once the chicken is grabbed, it must be taken to its pen and a key pressed to enter it. Fifth box: Since the chicken was taken to the correct pen and the correct key was pressed, the chicken was successfully saved, and a reward was received.

*down from the hills in search of food and destroy everything in their path, whether animals or people.*" [3].

- **Character design**. There are three types of characters in the SG. The **Farmer** is the main character in the story and interacts with the animals to herd them into their corrals. The **animals** (cows, pigs, sheep, chickens, turkeys) are about to escape from their respective holding place. And finally, the **wolves**, which are predatory animals that will appear to eat the farm animals if the farmer doesn't manage to get them back into the corral in time.
- **Chapters design**. The game has five different chapters. In each of the chapters, the gameplay will be the same; the only difference within the chapters is the kind and number of animals that will appear: *Level 1*: Chickens and wolves; *Level 2*: Chickens, cows, and wolves; *Level 3*: Chickens, cows, pigs, and wolves; *Level 4*: Chickens, cows, pigs, sheep, and wolves; and *Level 5*: Chickens, cows, pigs, sheep, turkeys, and wolves.
- **Labeling and evaluation of cognitive activities**. As mentioned previously, the activity selected for cognitive stimulation of concentration is the identification of identical objects. This activity is carried out within the game when the farmer identifies which corral the animal the player must herd belongs to and puts it there.

8. **Adaptive structuring and personalization design**.

- *User profile development.* No user profile was developed for this game.
- *Labeling and identifying inclusive design approaches*
  - (a) Both the video game controller stick and the VR Bluetooth remote controller stick will be used for the farmer's mobility (i.e., navigation) as well as to link the animal and place it inside the pen (i.e., selection).
  - (b) As for UI elements, each of them is implemented with a considerable size.
  - (c) Large fonts will be used to provide instructions within the SG.
  - (d) The colors used in the game will be as similar as possible to reality, to not confuse players within the SG.

[3] A video with an example of the serious game is available at https://drive.google.com/drive/folders/11_XUmGXqgs2Poey2gtzR4VQayDylaU4E?usp=sharing

- *Inclusion of artificial intelligence.* For this game, artificial intelligence was not included.

9. **Design of orientation and motivation strategies**.

- *Tutorial design.* For the development of this SG, a video tutorial was created in which the functionalities of the remote control are explained. This video explains in detail and through examples how to interact with each of the objects within the game.
- *Points and rewards system.* The reward system within this SG will be simple, with points awarded only for specific actions.

## 4. Assessing *Adventures in the Corral, Heroes of the Flock*

In order to assess the SG developed following the aforementioned methodology and also to determine whether it meets the objective of providing cognitive stimulation for elderly people with mild cognitive impairment, we carried out a usability study [4] as played by a group of 21 people in a community center located in the suburbs of Puebla, Mexico. We asked permission from the respective authorities, explaining the objectives of our research. All participants signed an informed consent before participating in this evaluation. The evaluation was conducted in two phases. In the first phase, a total of 12 people participated, all female, with ages ranging from 61 to 83 years old, where only 16.66% of the population had ever used video games in their lives. In the second phase, 9 people participated, all female, with ages ranging from 67 to 79 years old, with no previous experience in the use of video games.

In both phases, participants were asked to sit at a table in front of a computer. In the first phase, the players used a video game remote control for playing. In the second phase, people had the opportunity to choose a device for playing, either a video game remote control or a VR Bluetooth remote control. Only two people decided to use the former, and the remaining decided to use the latter. Once the interaction began, participants were informed that they could end the test at any time if they felt uncomfortable. Afterwards, they were asked to complete a short tutorial and start the game evaluation. In case people had any doubts about the interaction within the game, the researchers who conducted the test were allowed to interact. Likewise, every time a good action was performed, people were praised as an attempt to boost their confidence. Finally, every time the players ran out of time or finished a level, they were given the option to continue playing or end the game, to make them feel comfortable and not to feel pressure to continue playing, avoiding having a bad experience while using the SG.

We designed a 17-item Likert-scale (rated from 1 to 5) questionnaire, including aspects regarding usability, usefulness, and intention of use, to get some insights into the users' experience after using the SG. In particular, the questionnaire designed to assess usability was divided into the following sections: aesthetics, learning, operability, challenge, satisfaction, design, attention, and relevance. Overall, the results of the participants' perception suggest a *general acceptance* of the SG, with an average greater than or equal to 4.44 for most items, as shown in Table 2. The lowest value with regards to the *relevance* aspect, which suggests that participants may prefer other forms of exercise since, for most of them, this was the first time interacting with a video game.

From the results presented above, we may conclude that in general, the structure of the game was perceived as correct, which indirectly serves to evaluate the proposed methodology. No negative comments were expressed by the participants; instead, scores greater than 4 were obtained for most aspects. Furthermore, a possible explanation for the *relevance* aspect lower than the 4 score is the lack of previous experience of the participants in video games.

Despite the positive results obtained, there is room for improvement for the SG (which also applies to the methodology) according to the feedback provided by the participants in the evaluation. This is summarized in the following recommendations:

- **Paying attention to the duration of the game**. When designing activities, consider that participants with mild cognitive impairment may perform less well than healthy participants. An ideal scenario is that real patients evaluate the game, but when this is not possible, it is suggested to measure the time it takes a healthy subject to perform the activity and then multiply it by 3 times, trying to ensure that they have time enough to perform the activity.

- **Settings of the device for interacting with the game**. The movement speed for the characters must be lower than in a regular video game, since this can cause the elements to move very quickly, confusing the users.

- **Using levels**. When the SG includes levels, instead of continuously implementing them, adjust them within a main menu or within the user profile to make browsing easier.

- **Size of the landscape**. When the SG map is large, it is important to ensure that the player sees all the elements s/he will be interacting with from the start, in order to facilitate interaction with the game.

[4] No clinical trials were conducted in this study, as it is an initial phase of evaluation. Rigorous clinical evaluations are planned for future phases.

**Table 2.** Results of the usability study for phase 1 (E1) and phase 2 (E2).

| Aspect | E1 | E2 | Related questions |
|---|---|---|---|
| Aesthetics | $n \geq 4.58$<br>$SD \leq 0.62$ | $n \geq 4.44$<br>$SD = 0.5$ | *"The elements that appear are pleasant and similar to reality."*<br>*"The letters, their size and the colors seemed adequate to me."* |
| Learning | $n = 4.58$<br>$SD = 0.49$ | $n = 4.44$<br>$SD = 0.63$ | *"It was easy for me to learn to play."* |
| Operability | $n \geq 4.58$<br>$SD \leq 0.5$ | $n \geq 4.44$<br>$SD \leq 0.68$ | *"The rules of the game were clear and concise."*<br>*"The tutorial kept me from making mistakes."* |
| Challenge | $n \geq 4.25$<br>$SD \leq 0.60$ | $n = 4.44$<br>$SD = 0.5$ | *"The game's difficulty was appropriate."*<br>*"The pace of the missions varied appropriately."* |
| Satisfaction | $n \geq 4.58$<br>$SD \leq 0.49$ | $n \geq 4.44$<br>$SD \leq 0.5$ | *"I managed to overcome the obstacles and advance in the game."*<br>*"I feel satisfied with the exercise I completed."* |
| Design | $n = 4.25$<br>$SD = 0.60$ | $n = 4.44$<br>$SD = 0.5$ | *"Some elements of the game made me laugh."* |
| Attention | $n \geq 4.17$<br>$SD \leq 1.28$ | $n \geq 4.00$<br>$SD \leq 1.29$ | *"I was so focused on the game that I lost track of time. "*<br>*"I forgot about my surroundings while playing."* |
| Relevance | $n = 3.50$<br>$SD = 1.32$ | $n = 3.67$<br>$SD = 1.49$ | *"I prefer to exercise with this game than with other forms"* |

## 5. Conclusions

This paper presents a methodology for the design of serious games for cognitive stimulation of elderly people with mild cognitive impairment. Despite the growing interest in the use of computers for rehabilitation purposes, there is a lack of standard methodologies guiding the process of developing serious games for this purpose. We propose a set of steps that can help people involved in such a process to consider different aspects, ranging from stylistic design to the activities to be performed by patients. In addition, we developed a serious game following the proposed methodology. This game was evaluated by a group of elderly people, who interacted with it and then answered a usability questionnaire regarding their satisfaction with the game. The results obtained in usability evaluation show good levels of satisfaction, which gives evidence that the methodological approach presented helps in the development of effective and motivating serious games for this specific target population. As future work, we are interested in evaluating the usefulness of the proposed methodology from the perspective of software developers.

## Ethics statement

According to the guidelines of the Instituto Nacional de Astrofísica, Óptica y Electrónica Ethics Committee, this study was exempt from full ethical review. Nevertheless, all participants were informed about the study and provided voluntary consent prior to their involvement.

## CRediT authorship contribution statement

**Luisa Andrea Morales-García:** Conceptualization, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization and Writing – original draft. **Luis Enrique Sucar:** Conceptualization, Formal analysis, Investigation, Project administration, Resources, Supervision, Visualization and Writing – review & editing. **Delia Irazú Hernández-Farías:** Conceptualization, Formal analysis, Investigation, Project administration, Resources, Supervision, Visualization and Writing – review & editing. **Alberto L. Morán:** Conceptualization, Formal analysis, Investigation, Project administration, Resources, Supervision, Visualization and Writing – review & editing.

## Declaration of Generative AI and AI-assisted technologies in the writing process

The authors utilized Grammarly and ChatGPT to refine sentence structure and enhance readability. No content was generated by AI; all scientific insights and original ideas are the authors' own.

## Declaration of competing interest

The authors declare no competing interests.

## References


[1] R. Petersen, O. Lopez, M. Armstrong, T. Getchius, M. Ganguli, D. Gloss, ..., and A. Rae-Grant, "Practice guideline update summary: Mild cognitive impairment: Report of the guideline development, dissemination, and implementation subcommittee of the american academy of neurology," *Neurology*, vol. 3, no. 90, pp. 126–135, 2018.

[2] A. Larner, "Diagnosis of dementia and cognitive impairment," *Diagnostics*, vol. 4, no. 9, p. 180, 2019.

[3] S. García-Martínez, F. Orihuela-Espina, L. Sucar, A. Moran, and J. Hernández-Franco, "A design framework for arcade-type games for the upper-limb rehabilitation," *2015 International Conference on Virtual Rehabilitation (ICVR)*, pp. 235–242, 2015.

[4] R. Petersen, O. Lopez, M. Armstrong, T. Getchius, M. Ganguli, D. Gloss, ..., and A. Rae-Grant, "Serious games to cognitively stimulate older adults: a systematic literature review," *Proceedings of the 13th ACM International conference on pervasive technologies related to assistive environment*, pp. 1–10, 2020.

[5] A. Lobo, P. Saz, G. Marcos, J. Día, C. de la Cámara, T. Ventura, ..., and S. Aznar, "Juegos serios: alternativa innovadora," *Conocimiento Libre y Educación (CLED)*, vol. 2, no. 2, 2014.

[6] S. Lau and H. Agius, "A framework and immersive serious game for mild cognitive impairment," *Multimedia Tools and Applications*, vol. 20, no. 80, pp. 1183–31 237, 2021.

[7] G. Borrego, A. Morán, V. Meza, F. Orihuela-Espina, and L. Sucar, "Key factors that influence the ux of a dual-player game for the cognitive stimulation and motor rehabilitation of older adults," *Universal Access in the Information Society*, vol. 20, pp. 767–783, 2021.

[8] E. de Postgrado de Psicología y Psiquiatría. (2022) Tipos de deterioros cognitivos, causas y tratamiento. escuela de postgrado de psicologÍa y psiquiatrÍa. [Online]. Available: https://eepsicologia.com/deterioros-cognitivos-tipos-causas/

[9] R. Barzola, "Diagnóstico y tratamiento del deterioro cognitivo en al adulto mayor en el primer nivel de atención," *Catálogo Maestro De Guías De Prácticas Clínicas*, vol. 1, pp. 8–18, 2012.

[10] N. I. on Aging. (2021) What is mild cognitive impairment? [Online]. Available: https://www.nia.nih.gov/health/memory-loss-and-forgetfulness/what-mild-cognitive-impairment

[11] M. Clinic. (2024) Mild Cognitive Impairment. [Online]. Available: https://www.mayoclinic.org/es/diseases-conditions/mild-cognitive-impairment/symptoms-causes/syc-20354578

[12] L.-E. M. C.-A. B. C.-F. J. M. M. T.-P. S. y. V.-M. M. Rojas-Zepeda, C., "Factores de riesgo sociodemográficos y mórbidos asociados a deterioro cognitivo leve en adultos mayores," *Cuadernos de Neuropsicología/Panamerican Journal of Neuropsychology*, vol. 2, no. 15, 2021.

[13] N. Jiménez and M. Torrecillas. (2014) Ejercicios de estimulación cognitiva. [Online]. Available: https://www.kernpharma.com/sites/default/files/documentos/Cuadernillo%20Ejer%20Estimulaci%C3%B3n%20Leve2.pdf

[14] E. Ferrada, R.and Sandoval, B. Carrillo, and A. Aedo. (2020) Cuadernillo de estimulaciÓn cognitiva y fÍsica. [Online]. Available: https://www.med.ufro.cl/departamentoenfermeria/images/descargas/1.Cuadernillo-estimulacion-cog-fis.pdf

[15] N. A. Cognitiva. (2024) Ejercicios de estimulación cognitiva, memoria. [Online]. Available: https://academianeurona.com/ejercicios/

[16] K. Pharma. (2024) Ejercicios de estimulación cognitiva. nivel básico 2. [Online]. Available: https://www.fundacioace.com/uploads/files/categorias/ejercicios-online/kern-pharma/kern-pharma-estimulacion-cognitiva-nivel-basico-2.pdf

[17] G. Mora, P. Ramírez, L. Zeledón, E. Carrillo, M. Batista, and K. Robles. (2019) Estimulación cognitiva para personas adultas mayores: abordaje multidisciplinario. [Online]. Available: https://centrodiurnodeeltejar.org/images/ejercicios/libro-estimulacion-cognitiva-centro-diurno.pdf

[18] P. Espíndola, "Recurso para educadores sociales: Actividades de estimulación cognitiva," Trabajo de Fin de Grado, Universidad de Barcelona, 2015. [Online]. Available: https://www.grupdem.com/uploads/documentos/recurso-para-educadores-sociales.pdf

[19] A. P. Elche. (2024) Cálculo matemático. [Online]. Available: https://parkinsonelche.es/category/servicios/estimulacion-cognitiva/calculo-matematico/

[20] F. Rodríguez, C. Salazar, and K. Carrera. (2023) Manual de estimulación cognitiva para personas adultas mayores. [Online]. Available: https://www.gob.mx/cms/uploads/attachment/file/891719/Manaul_de_Estimulaci_n_Cognitiva_en_PAMS.pdf

[21] E. Verhulst, F. Banville, P. Richard, S. Tabet, C. Lussier, Massicotte, and P. Allain, "Navigation patterns in elderly during multitasking in virtual environment," *Human Interface and the Management of Information: Supporting Learning, Decision-Making and Collaboration: 19th International Conference*, pp. 176–188, 2017.

[22] B. Brathwaite and I. Schreiber, *Challenger for game designers*. Course Technology, 2009.

[23] H. Fuster, "El diseño de videojuegos: Diversión y potencial adictivo como dos caras de la misma moneda," *Rev. esp. drogodepend*, pp. 49–64, 2020.

[24] P. Alsina, *Jugabilidad: arte, videojuegos y cultura.* Artnodes, 2007, vol. 1, no. 7.

[25] B. Martín, "Game Design Aspectos que consiguen que un juego sea divertido y motive a seguir jugando," Trabajo de fin de grado, Universidad de Alicante, 2022.

[26] M. Moore, *Basics of Game Design.* CRC Press, 2011.

[27] E. B. . M. School. (2024) Diseño de niveles de videojuegos: qué es y principios básicos. [Online]. Available: https://www.esic.edu/rethink/tecnologia/diseno-niveles-videojuegos-que-es-principios-basicos-c

[28] D. Castilla, C. Suso-Ribera, I. Zaragoza, A. Garcia-Palacios, and C. Botella, "Designing icts for users with mild cognitive impairment: a usability study," *International Journal of Environmental Research and Public Health*, vol. 14, no. 17, p. 5153, 2020.

[29] J. Leyva Regalón and I. Mayol Céspedes, "Los juegos serios en el entrenamiento y la rehabilitación cognitiva," *Revista Cubana de Informática Médica*, vol. 2, no. 11, pp. 140–157, 2019.

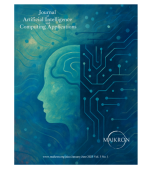
[30] J. Mendoza Solórzano, "Revisión de la literatura sobre el uso de la inteligencia artificial con enfoque a su aplicación en los videojuegos," Tesis de licenciatura, Universidad Politecnica Salesiana, 2024. [Online]. Available: https://dspace.ups.edu.ec/bitstream/123456789/28119/4/UPS-GT005498.pdf

[31] J. Duch, H. Tejedor Navarro, S. Kanaan-Izquierdo, and G. Escudero, *Inteligencia artificial para videojuegos.* Universitat Oberta de Catalunya, 2019.

[32] M. Reinlieb, L. Ercoli, P. Siddarth, N. Cyr, and H. Lavretsky, "The patterns of cognitive and functional impairment in amnestic and non-amnestic mild cognitive impairment in geriatric depression," *The American Journal of Geriatric Psychiatry*, vol. 12, no. 22, pp. 1487–1495, 2014.

Journal of Artificial Intelligence and Computing Applications (2025) 3(1): 49–60

Journal of Artificial Intelligence and Computing Applications



*Research article*

# Relationship between academic engagement and burnout syndrome in Mexican students: a PLS-SEM analysis

**Irving A. Ramírez-Muñoz**[1], **Vanesa Avalos-Gaytán**[1,*], **Igor Barahona**[2], **Valeria Soto-Mendoza**[1], **and Gabriela Linares Acuña**[3]

[1]Universidad Autónoma de Coahuila, Centro de Investigación en Matemáticas Aplicadas, Unidad Saltillo, Mexico.
[2]King Fahd University of Petroleum and Minerals, Department of Information Systems and Operations Management, Dhahran, Saudi Arabia.
[3]Universidad Autónoma de Coahuila, Facultad de Psicología, Unidad Saltillo, Mexico.

**ABSTRACT**

Academic burnout is a growing concern in higher education, characterized by emotional exhaustion, cynicism, and a reduced sense of accomplishment. In contrast, academic engagement – defined as a positive, energetic, and committed state toward learning – has been identified as a protective factor and even an antidote to burnout. While most studies in this area have focused either on theoretical model development or on validating measurement instruments, few address both simultaneously. Moreover, research using Structural Equation Modeling (SEM) has predominantly been conducted in Europe and the United States, leaving Latin American contexts underexplored. A literature review revealed only nine studies on academic burnout in Mexico, underscoring the need for further investigation in the region. This study aims to bridge that gap by validating adapted versions of the Maslach Burnout Inventory-Student Survey (MBI-SS) and the School Engagement Measure, and by developing a SEM to examine how academic engagement influences burnout levels among students at a public university in northern Mexico. The findings are expected to contribute to the understanding of student well-being in Latin America and to offer validated tools for measuring and addressing academic burnout.

**Keywords:** burnout syndrome, academic engagement, structural equation model

## 1. Introduction

Burnout syndrome has its origins in the medical field [1]. However, today it has impacted several other contexts, so much so that the World Health Organization (WHO) included it in its International Classification of Diseases (ICD-11) in 2019 [2]. Originally defined within the context of healthcare professions, burnout was characterized as a combination of chronic emotional exhaustion, fatigue, and depersonalization toward patients [3]. Academic burnout has become an increasing concern in the educational field, particularly among university students [4]. In this setting, the condition has evolved to describe a stress response marked by negative attitudes and feelings towards school and the role of the student [3].

Although students are not legally considered employees, from a psychological standpoint, their academic activities can reasonably be regarded as a form of work. The primary distinction between formal employment and academic activity lies in the absence of a direct monetary exchange. However, this difference may be at-

E-mail address: vanesaavalos@uadec.edu.mx
https://doi.org/10.5281/zenodo.16916210

tenuated by the fact that students often maintain an active relationship with educational institutions and may receive financial support such as scholarships. Additionally, students are frequently exposed to performance pressure and demanding academic environments, both of which have been linked to declines in physical and mental health [5].

In contrast, engagement has emerged as a key construct for understanding the student experience, particularly in relation to academic performance, well-being, and retention. It is defined as a positive emotional state characterized by energy, dedication, and active involvement in academic activities [6]. Unlike burnout – marked by exhaustion, cynicism, and depersonalization – engagement is considered its positive counterpart and even a potential antidote [7]. Engagement comprises emotional, cognitive, and behavioral dimensions, reflecting the student's connection to their learning environment, academic tasks, peers, and instructors [8]. This commitment is shaped by individual factors such as self-efficacy, self-esteem, and positive emotions, as well as contextual elements including the quality of the university environment, academic workload, and time devoted to extracurricular activities [7, 6].

Numerous studies have shown that academic burnout and engagement are closely related, yet often opposing, constructs within the educational context [4, 9]. While burnout is characterized by emotional exhaustion, cynicism, and a diminished sense of personal accomplishment, engagement is associated with vigor, dedication, and deep absorption in academic activities [6, 4]. However, the relationship between burnout and engagement is not strictly dichotomous; rather, it may reflect a continuum in which students can experience elements of both to varying degrees.

Building on the preceding discussion, the main objective of this study is to examine the relationship between academic engagement and academic burnout among university students in northern Mexico. To this end, Partial Least Squares Structural Equation Modeling (PLS-SEM) is used to analyze the influence of emotional, cognitive, and behavioral engagement on burnout dimensions, namely emotional exhaustion, cynicism, and reduced academic efficacy. The remainder of the paper is structured as follows: the next section presents the literature review; Section 3 describes the materials and methods; Sections 4 and 5 discuss the results and provide the conclusions, respectively.

## 2. Literature review

Academic burnout has become a growing concern in educational settings (see Table 1), particularly in countries like China, where its prevalence among high school students has prompted research using the Job Demands-Resources (JD-R) model. This model helps explain how academic workload, teacher relationships, and personal resources (e.g., self-efficacy, self-esteem, optimism) contribute to either the development or prevention of burnout through protective factors [10].

The most widely used instrument to assess academic burnout is the Maslach Burnout Inventory-Student Survey (MBI-SS), a student adapted version of the MBI-General Survey. It includes 15 items measuring emotional exhaustion, cynicism, and personal efficacy [11, 12].

Recent studies have explored the relationship between burnout, academic engagement, and affective traits such as attachment. High attachment-related anxiety has been shown to reduce academic engagement and increase burnout symptoms, which may influence students' decisions to persist or drop out of their studies [13]. While engagement is positively associated with academic performance and satisfaction, burnout is more strongly linked to dropout intentions.

Despite its importance, research on engagement faces methodological challenges, including overreliance on self-reported data and limited focus on outcomes. Scholars have called for more diverse methodologies, including observational approaches and culturally varied samples [8]. For example, a study with German university students found that 6.7% exhibited both high engagement and high burnout, suggesting a more complex relationship than previously assumed [7].

Cross-cultural studies in Finland [4], Spain [6], Canada [13], India, and Romania [9] consistently support a negative correlation between burnout and engagement. These findings highlight the potential of fostering academic engagement as a strategy to reduce burnout and improve student retention.

## 3. Materials and methods

This research utilizes the PLS-SEM technique to estimate and validate a causal model composed of latent variables. PLS-SEM focuses on maximizing the explained variance in endogenous constructs and is particularly suitable for complex models, non-normal data, and theory development contexts [20, 21].

### 3.1 Model specification

Let $\boldsymbol{\xi} = [\xi_1, \xi_2, \ldots, \xi_p]^\top$ denote the vector of exogenous latent variables and $\boldsymbol{\eta} = [\eta_1, \eta_2, \ldots, \eta_q]^\top$ denote the vector of endogenous latent variables. The structural model is represented as:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{1}$$

where $\mathbf{B}$ is a $q \times q$ matrix of path coefficients among endogenous constructs (excluding diagonal elements), $\boldsymbol{\Gamma}$ is a $q \times p$ matrix representing the influence of exogenous constructs on endogenous constructs, and $\boldsymbol{\zeta}$ is a vector of structural disturbances.

For reflective measurement models:

$$\mathbf{x} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}, \quad \mathbf{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon} \tag{2}$$

where $\mathbf{x}$ and $\mathbf{y}$ are vectors of observed indicators, $\boldsymbol{\Lambda}_x$ and $\boldsymbol{\Lambda}_y$ are loading matrices, and $\boldsymbol{\delta}$ and $\boldsymbol{\epsilon}$ are measurement errors.

**Table 1.** Related work comparison.

| Authors | Instrument | Variables | | Approach | | Method | Population | Region | Sample size | Sampling technique |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Dependent | Independent | Validation | Model | | | | | |
| Yavus & Dogan (2014)[14] | MBI-SS | Academic burnout | Student's characteristics | X | | Confirmatory factor analysis | High school students | Turkiye | 1020 | Not specified |
| Veiga et. al (2016)[15] | SES with Four-Dimensional Scale | Engagement | N/A | X | | Confirmatory factor analysis | High school Students | Portugal | 685 | Convenience |
| Hederich-Martínez & Caballero-Domínguez (2016) [11] | MBI-SS | Academic burnout | N/A | X | X | Factor analysis | Students | Colombia | 820 | Not specified |
| Grilo et. al. (2019) [5] | Precursors of burnout | Academic burnout | Behavioral stress | | X | PLS-SEM | Student volunteers | Arizona, USA | 374 | Convenience |
| Virga et al. (2020) [9] | PysCap, UWES-S MBI-SS, UBOS | Engagement | Psychological capital, academic performance, and boredom | | X | PLS-SEM | University students | India and Romania | 420 | Convenience |
| De la Fuente et. al.(2020)[6] | AEQ, EEC-Short MBI-SS,UWE | Engagement & burnout | Achievement emotions, coping mechanisms | | X | SEM and logistic regression | Students | Spain | 642 | Convenience |
| Brubaker et. al. (2020) [3] | PSS, PSQI and MBI-SS | Perceived stress, burnout levels, and sleep quality | Intervention | | X | Quasi-experimental | Medical students | Ohio, USA | 57 | Convenience |
| Bumbacco and Schafre (2020)[13] | EDA and SBI | Engagement and academic burnout | Attachment | | X | Qualitative correlational study | First year college students | Canada | 290 | Intencional |
| Kiema-Junes et. al. (2020)[4] | UWES-S and SBI-9 | Burnout and engagement | Self-perceived social skills | | X | Linear regression | Students | Finland | 351 | Not specified |
| Smith & Emerson (2021) [16] | Connor Resilience GHQ-12 and MBI-SS | Academic burnout and psychological distress | Resilience | | X | SEM | Undergraduate accounting students | USA | 443 | Convenience |
| Teuber et. al. (2021)[10] | QARCA-C and MBI-SS | academic burnout | workload, academic demands, teacher-student relationships and optimism | | X | SEM | High school students | China | 1083 | Not specified |
| Reyna-Castillo et al. (2022)[17] | MBI-SS and EMEDO | Academic burnout | Sociodemographic antecedents | X | X | PLS-SEM | Students | Mexico and Colombia | 235 | Convenience |
| Fiorilli et al. (2022) [18] | Burnout BAT-C-Short | Burnout | Gender and employment status | | X | MANOVA | Students | Italy | 494 | Snowball |
| Olson et. al. (2023) [7] | UWES-S 9 and MBI-SS | Engagement and student burnout | Work overload and academic satisfaction | | X | SEM, Regression analysis | University students | Germany | 3451 | Convenience |
| Gutu et al. (2024) [19] | Instrument made by the authors | Level of engagement | Higher Education digitalization and academic leadership | | X | PLS-SEM | Students | Romania | 2272 | Not specified |
| This work (2025) | MBI-SS and School Engagement Measure-MacArthur | Academic burnout | Academic engagement | X | X | PLS-SEM and Confirmatory factor analysis | Higher education students, anonymous and voluntary | Mexico | 552 | Convenience |

For formative constructs:

$$\boldsymbol{\xi} = \mathbf{W}_x^\top \mathbf{x}, \quad \boldsymbol{\eta} = \mathbf{W}_y^\top \mathbf{y} \tag{3}$$

where $\mathbf{W}_x$ and $\mathbf{W}_y$ are weight matrices estimated to maximize the $R^2$ of the endogenous latent constructs.

### 3.2 Estimation procedure

PLS-SEM was performed using SmartPLS (v4.1), employing a three-stage iterative algorithm:

1. Initial approximation of latent variable scores based on proxies,
2. Estimation of inner model relationships using weighted least squares,
3. Update of weights and loadings until convergence is achieved.

Bootstrapping with 5,000 subsamples was used to assess the significance of path coefficients and measurement parameters.

### 3.3 Measurement model evaluation

For reflective indicators, the following criteria were used:

- Indicator reliability: Outer loadings $\lambda_i \geq 0.70$
- Internal consistency: Composite reliability (CR) $\geq 0.70$
- Convergent validity: Average Variance Extracted (AVE) $\geq 0.50$

$$\text{AVE} = \frac{1}{k}\sum_{i=1}^{k} \lambda_i^2 \tag{4}$$

For formative constructs, multicollinearity was assessed via Variance Inflation Factor (VIF), ensuring VIF $< 5$, and significance of outer weights was tested using bootstrapped $t$-statistics.

### 3.4 Structural model evaluation

The structural model was evaluated through:

- Coefficient of determination ($R^2$) for endogenous constructs,
- Effect size ($f^2$) for each path:

$$f^2 = \frac{R^2_{\text{included}} - R^2_{\text{excluded}}}{1 - R^2_{\text{included}}} \tag{5}$$

- Predictive relevance ($Q^2$) using the blindfolding procedure,
- Model fit using Standardized Root Mean Square Residual (SRMR), with threshold $< 0.08$.

### 3.5 Mediation and multigroup analysis

Mediation effects were examined by calculating the product of coefficients ($\beta_{ab} = \beta_a \cdot \beta_b$), and significance was tested via bootstrapped confidence intervals. Multigroup analysis (PLS-MGA) was conducted to explore differences in path coefficients across predefined subgroups using non-parametric techniques.

## 4. Participants and procedure

The study focused on students from a public university in Northern Mexico. A sample of size $n = 552$ students was collected, with 343 identifying as women, 202 as men, and only 7 as non-binary, representing 62.1%, 36.5%, and 1.3%, respectively. The average age was 19 years with a standard deviation of $\pm\,5.4$ years. The sample was obtained using a convenience sampling method. Participation was anonymous and voluntary; however, participants were given the option to register an email address if they wished to receive their results. To collect the information, assistance from specific teachers was requested, as well as support from the university tutoring department and the official announcements department.

### 4.1 Instruments and their validation

MBI - Student Survey [11] and School Engagement Measure - MacArthur [22] were adapted and used for this work. They were also validated by calculating the Cronbach's alpha coefficient [23]. Then we conducted an exploratory factor analysis (EFA) as suggested [24].

***MBI - student survey***

To determine levels of burnout, a translated and adapted version of the MBI-SS [11] was used. This questionnaire consists of six items measuring academic personal efficacy (e.g., "I can effectively solve problems related to my field of study."), five items measuring emotional exhaustion (e.g., "The academic activities in my field have left me emotionally drained."), and four items measuring cynicism (e.g., "I have become distanced from my field of study because I think it will not be useful for my professional development.").

The results in Table 2a show that all Cronbach's alpha values exceed 0.70, indicating that the items exhibit good reliability to measuring each factor [25]. On the other hand, Table 2b reports a KMO value of 0.89 and a $p-value$ of 0.001 in Bartlett's test [26], which suggests that applying an EFA is appropriate. Table 3 presents the factor loadings of the questionnaire items where *ago* refers to *emotional exhaustion* items, *cin* refers to *cynicism* items and, *efi* refers to *academic efficacy* items, with its associated number of item. It can be seen that all the items have satisfactory loadings, exceeding 0.40. Item *cin4* shows loadings in two factors, and only item *efi5* did not load significantly on any factor. These results confirm that the items appropriately measure the theoretical factors proposed in the MBI-SS and that the

**Table 2.** Results to Cronbach's alpha and the exploratory factor analysis for the MBI-SS.

**(a)** Cronbach's alpha coefficients for each factor in the MBI-SS.

| Factor | Cronbach's Alpha |
|---|---|
| Exhaustion (ago) | 0.8602 |
| Efficacy (efi) | 0.7792 |
| Cynicism (cin) | 0.8404 |

**(b)** Results from the exploratory factor analysis (EFA) for the MBI-SS.

| Statistic | Value |
|---|---|
| KMO Overall | 0.89 |
| Bartlett's Test $\chi^2$ | 3436.32 ($p < 0.001$) |
| Total variance explained | 51% |
| RMSR | 0.02 |
| RMSEA (90% CI) | 0.043 (0.033, 0.054) |
| Tucker-Lewis Index (TLI) | 0.967 |

**Table 3.** Factor loadings for the MBI-SS.

| **Item →** | *ago1* | *ago2* | *ago3* | *ago4* | *ago5* | *cin1* | *cin2* | *cin3* | *cin4* | *efi1* | *efi2* | *efi4* | *efi6* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **-Emotional exhaustion** | 0.67 | 0.68 | 0.72 | 0.81 | 0.70 | | | | 0.47 | | | | |
| **-Personal efficacy** | | | | | | | | | | 0.56 | 0.72 | 0.72 | 0.73 |
| **-Cynicism** | | | | | | 0.64 | 0.65 | 0.77 | 0.61 | | | | |

instrument can be reliably used for this purpose.

#### *School engagement measure - MacArthur*

Regarding academic engagement, a translation and adaptation of the MacArthur School Engagement Measure (SEM-MacArthur) was also carried out [22]. This questionnaire consists of five items measuring behavioral engagement (e.g. "I follow the rules at my faculty"), six items measuring emotional engagement (e.g., "I feel happy about the work I have to do"), and eight items assessing cognitive engagement (e.g., "I review my activities and school attitude for mistakes").

Table 4a shows that the emotional and cognitive factor scores exceed the required threshold of 0.70. Only the behavioral factor falls below this threshold (0.67). However, when item *com2* corresponding to this factor is removed, the coefficient increases to 0.71, suggesting that this item presents internal consistency issues and must be eliminated. Furthermore, the results in Table 4b show a KMO value of 0.90 and a significant result in Bartlett's sphericity test ($p < 0.001$), indicating that the conduct of an EFA is appropriate. Finally, Table 5 shows that the vast majority of questionnaire items load exclusively onto a single factor, only the items "*com2*" and "*efi6*" did not exhibit adequate loadings during the exploratory factor analysis (EFA), leading to their removal from the instrument. As a result, the final version of the scale includes 17 of the 19 items originally proposed.

Overall, these findings provide strong empirical support for the validity and reliability of the School Engagement Measure in the studied sample. The instrument demonstrates adequate internal consistency, a well-defined factor structure, and appropriate conditions for the application of exploratory factor analysis, thus confirming its suitability for assessing academic engagement in this context.

## 5. Results and discussion

Two key constructs emerge from the Structural Equation Modeling (SEM) proposed: academic engagement and academic burnout, which are discussed in the following subsections.

### 5.1 Academic burnout

Regarding academic burnout, it is composed of three factors: academic personal efficacy, emotional exhaustion, and cynicism. Given the number of items and the fact that all MBI-SS questions use a Likert scale from 0 to 6, theoretical maximum scores for efficacy, exhaustion, and cynicism would be 36, 30, and 24 points, respectively. Table 6 presents the descriptive statistics obtained from the university student sample.

For the personal academic efficacy factor, the students reported a mean score of 26.45 points and a median of 28 points, with a standard deviation of approximately 7 points. This suggests that most university students perceive themselves as highly effective in their academic activities. In contrast, the emotional exhaustion factor showed a mean of 17.78 and a median of 19, with a standard deviation of 7.86, indicating that students are moderately emotionally exhausted, although with considerable variability. Finally, the cynicism fac-

**Table 4.** Results to Cronbach's alpha and the exploratory factor analysis for the MBI-SS.

**(a)** Cronbach's alpha coefficients for each factor in SEM-MacArthur.

| Factor | Cronbach's Alpha |
|---|---|
| Behavioral (com) | 0.6703 |
| Emotional (emo) | 0.8970 |
| Cognitive (cog) | 0.8261 |

**(b)** Results from the exploratory factor analysis for the School Engagement Measure.

| Statistic | Result |
|---|---|
| KMO Overall | 0.90 |
| Bartlett's Test $\chi^2$ | 4007.16 ($p < 0.001$) |
| Total variance explained | 45% |
| RMSR | 0.03 |
| RMSEA (90% CI) | 0.044 (0.036, 0.052) |
| Tucker-Lewis Index (TLI) | 0.952 |

**Table 5.** Factor loadings for the School Engagement Measure.

| **Item** $\rightarrow$ | *com1* | *com2* | *com3* | *com4* | *com5* | *emo1* | *emo2* | *emo3* | *emo4* | *emo5* | *emo6* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **-Behavioral** | 0.47 | | 0.65 | 0.65 | 0.49 | | | | | | |
| **-Emotional** | | | | | | 0.82 | 0.73 | 0.72 | 0.69 | 0.83 | |
| **Item** $\rightarrow$ | *cog1* | *cog2* | *cog3* | *cog4* | *cog5* | *cog6* | *cog7* | *cog8* | | | |
| **-Cognitive** | 0.52 | 0.59 | 0.61 | 0.72 | 0.78 | 0.59 | 0.46 | 0.47 | | | |

**Table 6.** Descriptive statistics of the factors in MBI-SS.

| Factor | Efficacy | Exhaustion | Cynicism |
|---|---|---|---|
| Items | 6 | 5 | 4 |
| Mean | 26.45 | 17.78 | 6.48 |
| Median | 28 | 19 | 4.5 |
| Standard deviation | 7.02 | 7.86 | 6.51 |
| Skew | -0.88 | -0.33 | 0.97 |
| Kurtosis | 0.23 | -0.89 | 0 |

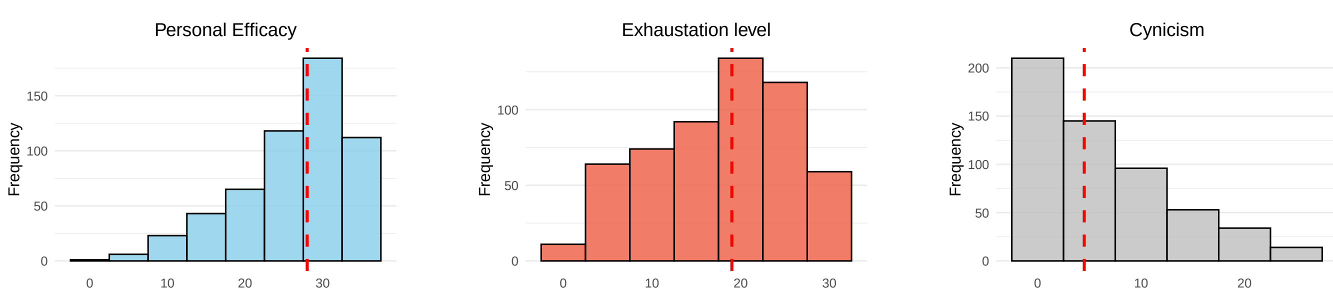



**Figure 1.** Histograms of the factors Efficacy, Exhaustion and Cynicism in students.

tor revealed a mean of 6.48 and a median of 4.5, accompanied by a standard deviation of 6.51. Although these results imply that most students exhibit low levels of cynicism, the relatively high standard deviation suggests that a smaller subgroup reports significantly higher levels of cynicism, as we can see in Figure 1.

Figure 1 displays the histograms of the score distributions for each factor of MBI-SS. It is visually evident and supported by the skewness and kurtosis values in Table 6 that none of the factors follows a normal distribution. Therefore, all group comparisons will be conducted using nonparametric tests.

Regarding academic burnout, the literature (Section 2) indicates that, for a student to be considered to experience a severe and significant level of burnout, three criteria must be met: low academic efficacy, high emotional exhaustion, and high cynicism. However, the syndrome may also be present in moderate or mild forms when one or two of these conditions are observed. In contrast, a student is considered free of academic burnout when they exhibit high levels of academic efficacy along with low levels of cynicism and emotional exhaustion. In this regard, Figure 2 illustrates the levels of academic burnout in the student sample from the Universidad Autonoma de Coahuila. Based on the criteria mentioned above, almost half of the students (47.6%) exhibit some level of academic burnout. In addition, at least 1 in 4 students (28.1%) show moderate to high levels of burnout syndrome, which puts them at risk of experiencing its negative consequences.

### 5.2 Academic engagement

Academic engagement is composed of three factors: behavioral engagement, cognitive engagement, and emotional engagement. It is important to note that the response scale for this questionnaire was a Likert-type scale ranging from 1 to 5. Theoretical maximum scores for each dimension are 25, 40, and 30, respectively. Table 7 presents the descriptive statistics for the academic engagement factors observed in the university student sample.

Table 7 reports that the behavioral engagement factor has a mean of 16.4, a median of 17, and a standard deviation of 2.38. These results suggest a high level of behavioral engagement with low variability, indicating consistency across responses. The negative skewness suggests a higher concentration of students scoring above the mean in this dimension. For the cognitive engagement factor, the mean score of 25.69 and median of 25.5 indicate a moderately high level of engagement. However, the standard deviation of 6 points reflects greater variability in responses compared to the other two dimensions, suggesting a broader range of cognitive involvement among students. Regarding emotional engagement, the mean of 20.38 and median of 20 indicate comparatively higher levels relative to the other factors. The standard deviation of 4 points denotes moderate dispersion, while a skewness value of -0.14 suggests a slight leftward asymmetry, implying that higher scores are more frequent.

In summary, all three academic engagement factors exhibit left skewed or near symmetric distributions, with overall scores clustering in the moderate-to-high range. This indicates that the majority of students in the sample demonstrate medium to high levels of academic engagement. This trend is also observable in Figure 3.

### 5.3 The SEM of academic burnout and academic engagement

Figure 4 presents the proposed model, which follows a hierarchical structure. In this model, the factors Behavioral, Cognitive, and Emotional operate sequentially to explain the higher-order construct of Academic Engagement, which in turn negatively predicts Academic Burnout, a second-order construct composed of personal efficacy, exhaustion, and cynicism.

First, the model indicates that students who pay attention in class, complete assignments on time, and remain focused during school hours are more likely to exhibit stronger cognitive engagement ($\beta = 0.524,\ p < 0.001$). This suggests that adherence to academic norms and personal responsibility enhances students' willingness to engage in deeper, reflective academic processes.

Furthermore, the cognitive engagement factor, which includes behaviors such as studying outside of class, reading supplementary materials, and discussing academic topics with others, exerts a significant positive effect on emotional engagement ($\beta = 0.417,\ p < 0.001$). This implies that students who invest in constructing knowledge beyond assigned tasks are also more likely to experience positive emotions in their academic life, such as satisfaction, interest, and enjoyment.

Finally, emotional engagement significantly reduces academic burnout levels ($\beta = -0.871,\ p < 0.001$). Specifically, students who report feeling happy, motivated, and satisfied with their academic activities tend to show lower levels of cynicism and exhaustion, and higher levels of personal efficacy.

These findings support the hypothesis that affective well-being within the university environment serves as a key protective factor against the psychological strain associated with academic demands.

#### *5.3.1 The SEM comparison between gender and semester*

Regarding the potential differences between the groups analyzed in this study, a multigroup analysis (MGA) was conducted using the *SEMinR* package in R [27]. This approach allows for the comparison of structural equation models (SEMs) across different groups to identify statistically significant differences. In this case, Table 8 presents the results of the MGA comparing gender (209 males vs. 343 females) and semester (360 students before their fifth semester vs. 192 students in their fifth semester or later).

For the gender comparison, the overall estimate mean (0.0231) is relatively small, and while the beta

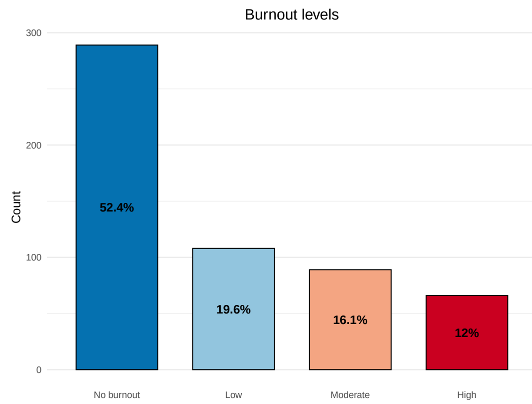



**Figure 2.** Levels of academic burnout in the university sample.

**Table 7.** Descriptive statistics of the factors in the School Engagement Measure.

| Factor | Behavioral | Cognitive | Emotional |
|---|---|---|---|
| Items | 4 | 8 | 6 |
| Mean | 16.43 | 25.69 | 20.38 |
| Median | 17 | 25.5 | 20 |
| Standard deviation | 2.38 | 6.03 | 4 |
| Skew | -0.58 | -0.02 | -0.14 |
| Kurtosis | -0.01 | -0.11 | -0.5 |

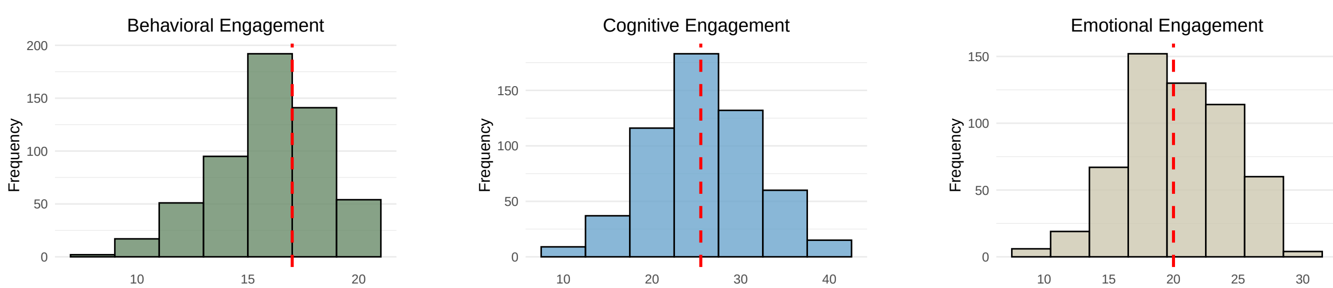



**Figure 3.** Histograms of the factors Behavioral, Cognitive and Emotional Engagement in the Student Engagement Measure.

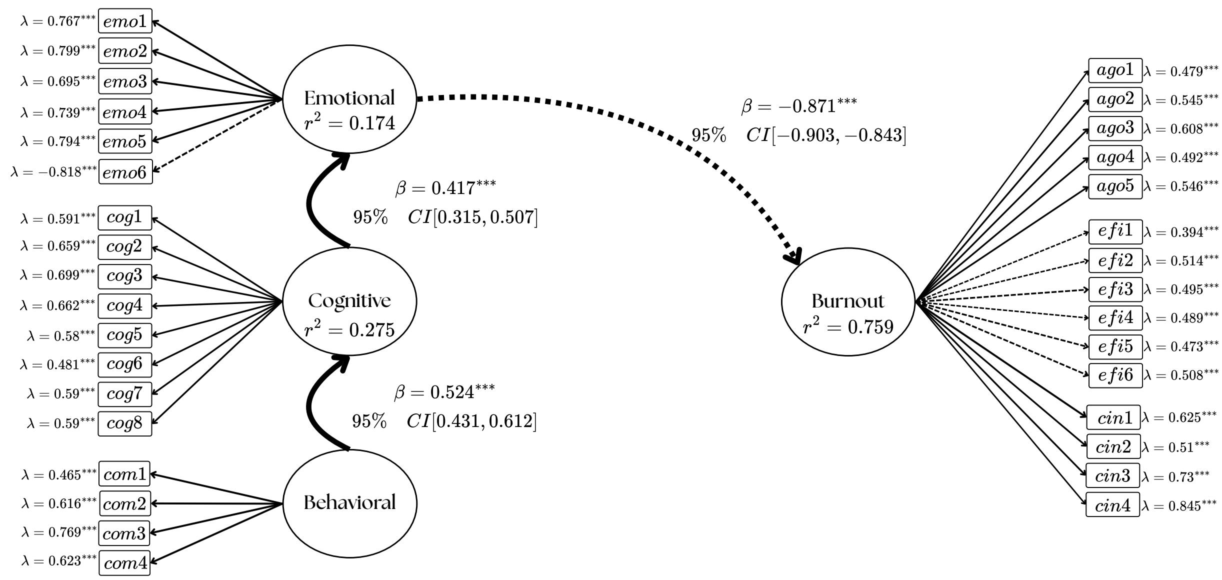



**Figure 4.** Levels of academic burnout in the university sample.

coefficient for male students (0.0290) is slightly higher than that for female students (0.0186), this difference is not substantial within the context of the model. The average p-value of 0.4801 indicates that, overall, the differences between the SEMs for male and female students are not statistically significant in this context.

For the comparison between the early and late semesters, Table 8 shows that the estimated mean (0.0231) is similar to that observed in the gender comparison, indicating that the average difference between these two semester groups is also small. Students in the first semesters (0.0230) exhibit a slightly weaker structural relationship compared to those in the later semesters (0.0465), suggesting a change in academic dynamics as students progress through their studies. However, the $p - values$ for this comparison confirms that there is no significant statistical evidence to support meaningful differences in the SEM relationships across these two groups. In other words, although students in higher semesters tend to exhibit slightly stronger effects, these differences are not statistically significant.

## 6. SEM & AI

This research, which focuses on analyzing academic burnout and engagement through SEM, can be enriched by incorporating artificial intelligence (AI)-based tools from both theoretical and practical perspectives. Two complementary approaches emerge from recent studies on intelligent technology adoption: the use of automated care systems in digital health [28] and the application of smart technologies in tourism experiences enhanced by augmented reality [29].

First, AI can be conceived as a key component in the development of adaptive dashboards for personalized academic support, inspired by digital medical platforms such as "eDoctor apps". These systems could be integrated with data collected via the MBI-SS and SEM-MacArthur instruments, generating a dynamic psychoeducational profile for each student. This profile would enable real-time monitoring of latent variables such as academic efficacy, emotional exhaustion, and cynicism—core constructs in our academic burnout model. The descriptive and structural analyses in this study revealed that students with higher engagement levels report significantly lower exhaustion and cynicism. Consequently, an AI-enhanced platform that optimizes these factors through personalized interventions could serve as a preventive tool, especially beneficial for institutions managing large student cohorts with limited human resources. Second, following Namahoot's proposal [29], the adaptation of AI-based "smart assistant" models—similar to virtual tourist guides—to academic contexts is proposed. In this setting, these assistants would serve as "cognitive and emotional tutors", supporting students throughout their academic journeys. Such emotional support would be particularly valuable for students reporting low efficacy or high cynicism, clearly identifiable segments in our structural model.

Both proposals position artificial intelligence as a mediating technology that, beyond its informational role, plays an active part in promoting academic well-being. Future research should evaluate the technical feasibility and user acceptance of such resources among students and faculty, incorporating variables such as attitude toward AI, intention to use, and perceived usefulness into extended SEM models. This would enable the development of evidence-based intervention pathways that foster ethical and personalized use of AI in higher education environments.

**Table 8.** MGA results comparing gender and semester.

| Group | Estimate mean | $\beta$-group 1 | $\beta$-group 2 | p-value | p-value min |
|---|---|---|---|---|---|
| Male vs. Female | 0.0231 | 0.0290 | 0.0186 | 0.4801 | 0.3002 |
| First vs. Last semesters | 0.0231 | 0.0230 | 0.0465 | 0.5222 | 0.0837 |

## 7. Conclusions and future work

This study examined the structural relationships between academic engagement and academic burnout among university students in Northern Mexico, using a second-order PLS-SEM. By conceptualizing academic engagement through its behavioral, cognitive and emotional dimensions, and academic burnout through emotional exhaustion, cynicism, and reduced academic efficacy, the model explained a substantial 75.9% of the variance in burnout. The results provide compelling evidence for the protective effect of emotional engagement against burnout, as demonstrated by the strong negative path coefficient ($\beta = -0.871$; 95% CI [$-0.903$, $-0.843$]) linking emotional engagement with burnout levels.

The findings corroborate theoretical models such as the JD-R and Conservation of Resources (COR) frameworks, suggesting that emotionally invested students exhibit lower susceptibility to burnout [10], [9]. Notably, cognitive engagement emerged as a positive antecedent of emotional engagement, indicating that students who go beyond basic academic requirements—by reflecting critically and engaging in peer discussions—also tend to report more positive academic emotions. These insights align with prior research emphasizing the mediating role of engagement in the relationship between personal resources (e.g., psychological capital) and burnout [13], [7], [9].

Multigroup analysis revealed no statistically significant differences across gender or semester level, suggesting the robustness of the model across demographic subgroups. However, the absence of significant moderation by these variables opens the door for future exploration of alternative moderators, such as socioeconomic status, attachment style, psychological capital, and institutional support systems.

In practical terms, the structural model validated in this study provides a foundation for future work. Pertinent research efforts may focus on operationalizing these findings through the development of predictive models—such as binary or ordinal logistic regression—to classify students at risk of burnout based on engagement profiles. Furthermore, longitudinal studies are recommended to capture the dynamic evolution of burnout and engagement over time, as cross-sectional designs may obscure causal inferences.

Additionally, the implementation of targeted interventions aimed at fostering emotional engagement—such as mentorship programs, positive psychology training, or resilience workshops—should be prioritized and empirically evaluated within Mexican universities. These initiatives are particularly relevant in the context of Mexico's ongoing educational transformation, where digitalization and shifting pedagogical models are altering traditional student-instructor dynamics. Institutional support for these programs is essential, as engagement can be significantly disrupted by socioeconomic disparities, limited student services, and inconsistent technological infrastructure—challenges that are especially salient in public universities across the country.

For higher education practitioners and decision-makers in the Mexican education system, the results of this study offer actionable insights for designing policies and allocating resources to strengthen protective factors against academic burnout. The validated SEM framework not only deepens theoretical understanding but also enables predictive diagnostics that can be embedded into institutional early-alert systems. By identifying students at risk through engagement profiles, universities can deliver timely, data-driven interventions, thereby enhancing student retention, academic success, and psychological well-being. In a national context where dropout rates and mental health issues are pressing concerns, these findings provide a critical roadmap for implementing evidence-based strategies that promote both educational quality and equity. Thus, this research directly aligns with Mexico's educational policy goals of improving inclusion, academic performance, and long-term student development.

## Ethics statement

This study used anonymized data collected via an online survey. No personally identifiable information was collected. Ethical approval was deemed unnecessary according to the guidelines of UADEC (https://www.uadec.mx/transparencia/acceso-a-la-informacion/53-aviso-de-privacidad-y-derechos-arco/).

## CRediT authorship contribution statement

**Irving A. Ramírez Muñoz:** Conceptualization, Methodology, Software, Validation, Formal analysis, data curation, Visualization, Writing-original draft, Writing-review & editing. **Vanesa Avalos-Gaytán:** Conceptualization, Methodology, Software, Validation, Formal analysis, data curation, Visualization, Writing-original draft, Writing-review & editing. **Igor Barahona:** Conceptualization, Methodology, Software, Validation, Formal analysis, data curation, Visualization, Writing-review & editing. **Valeria Soto-Mendoza:** Formal analysis, Visualization, Writing-review & editing. **Gabriela Linares-Acuña:** Formal analysis, Visualization, Writing-review & editing.

## Declaration of generative AI and AI-assisted technologies in the writing process

The authors utilized Grammarly and ChatGPT to refine sentence structure and enhance readability. No content was generated by AI; all scientific insights and original ideas are the author's own.

## Declaration of competing interest

All authors declare that there are no potential conflicts of interest—financial or personal—that could have influenced the results or interpretations of this work.

## Acknowledgements

The authors would like to thank the International Conference on Artificial Intelligence for Mental Health (ICAIMH) 2025 and the anonymous reviewers for their valuable comments and suggestions, which will improve the quality of this article.

## References

[1] H. J. Freudenberger, "Staff burn-out," *Journal of social issues*, vol. 30, no. 1, pp. 159–165, 1974.

[2] World Health Organization. (2019) Burn-out an "occupational phenomenon": International classification of diseases. [Online]. Available: https://www.who.int/news/item/28-05-2019-burn-out-an-occupational-phenomenon-international-classification-of-diseases

[3] J. R. Brubaker, A. Swan, and E. A. Beverly, "A brief intervention to reduce burnout and improve sleep quality in medical students," *BMC medical education*, vol. 20, pp. 1–9, 2020.

[4] H. Kiema-Junes, M. Hintsanen, H. Soini, and K. Pyhältö, "The role of social skills in burnout and engagement among university students," *Electronic journal of research in educational psychology*, vol. 18, no. 1 (50), pp. 77–100, 2020.

[5] L. M. Grilo, A. Mubayi, K. Dinkel, B. Amdouni, J. Ren, and M. Bhakta, "Evaluation of academic burnout in college students. application of sem with pls approach," in *AIP Conference Proceedings*, vol. 2186, no. 1. AIP Publishing, 2019.

[6] J. de la Fuente, F. Lahortiga-Ramos, C. Laspra-Solís, C. Maestro-Martín, I. Alustiza, E. Aubá, and R. Martín-Lanas, "A structural equation model of achievement emotions, coping strategies and engagement-burnout in undergraduate students: A possible underlying mechanism in facets of perfectionism," *International journal of environmental research and public health*, vol. 17, no. 6, p. 2106, 2020.

[7] N. Olson, R. Oberhoffer-Fritz, B. Reiner, and T. Schulz, "Study related factors associated with study engagement and student burnout among german university students. front public health 11: 1168264," 2023.

[8] K. Salmela-Aro, X. Tang, J. Symonds, and K. Upadyaya, "Student engagement in adolescence: A scoping review of longitudinal studies 2010–2020," *Journal of Research on Adolescence*, vol. 31, no. 2, pp. 256–272, 2021.

[9] D. Vîrgă, M. Pattusamy, and D. P. Kumar, "How psychological capital is related to academic performance, burnout, and boredom? the mediating role of study engagement," *Current Psychology*, vol. 41, no. 10, pp. 6731–6743, 2020.

[10] Z. Teuber, F. W. Nussbeck, and E. Wild, "School burnout among chinese high school students: the role of teacher-student relationships and personal resources," *Educational Psychology*, vol. 41, no. 8, pp. 985–1002, 2021.

[11] C. Hederich-Martínez and C. C. Caballero-Domínguez, "Validación del cuestionario maslach burnout inventory-student survey (MBI-SS) en contexto académico colombiano," *CES Psicología*, vol. 9, no. 1, pp. 1–15, 2016. [Online]. Available: http://www.scielo.org.co/scielo.php?pid=S2011-30802016000100002&script=sci_arttext

[12] E. Demerouti, A. B. Bakker, F. Nachreiner, and W. B. Schaufeli, "The job demands-resources model of burnout." *Journal of Applied psychology*, vol. 86, no. 3, p. 499, 2001.

[13] C. Bumbacco and E. Scharfe, "Why attachment matters: first-year post-secondary students' experience of burnout, disengagement, and drop-out," *Journal of College Student Retention: Research, Theory & Practice*, vol. 24, no. 4, pp. 988–1001, 2023.

[14] G. Yavuz and N. Dogan, "Maslach burnout inventory-student survey (MBI-SS): A validity study," *Procedia - Social and Behavioral Sciences*, vol. 116, pp. 2453–2457, 02, 2014.

[15] F. Veiga, "Assessing student engagement in school: Development and validation of a four-dimensional scale," *Procedia - Social and Behavioral Sciences*, vol. 217, pp. 813–819, 02 2016.

[16] K. Smith and D. Emerson, "Resilience, psychological distress, and academic burnout among accounting students†," *Accounting Perspectives*, vol. 20, 03 2021.

[17] M. Reyna-Castillo, M. A. Pulgarín-Rodríguez, A. H. Ríos-Serna, and A. Santiago, "PLS-SEM validation for burnout measures in latino college students: A socially sustainable educational return," *Sustainability*, vol. 14, no. 21, 2022. [Online]. Available: https://www.mdpi.com/2071-1050/14/21/14635

[18] C. Fiorilli, D. Barni, C. Russo, V. Marchetti, G. Angelini, and L. Romano, "Students' burnout at university: The role of gender and worker status," *International Journal of Environmental Research and Public Health*, vol. 19, no. 18, 2022. [Online]. Available: https://www.mdpi.com/1660-4601/19/18/11341

[19] I. Gutu, C. Medeleanu, and R. Asiminei, "The limits of learning engagement and academic leadership within the higher education digitalization process - analysis by using PLS SEM," *PLOS ONE*, vol. 19, 11 2024.

[20] H. Wold, "Model construction and evaluation when theoretical knowledge is scarce: Theory and application of partial least squares," in *Evaluation of Econometric Models*, J. Kmenta and J. B. Ramsey, Eds. Academic Press, 1980, pp. 47–74. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780124165502500078

[21] J. F. Hair, C. M. Ringle, and M. S. and, "Pls-sem: Indeed a silver bullet," *Journal of Marketing Theory and Practice*, vol. 19, no. 2, pp. 139–152, 2011. [Online]. Available: https://doi.org/10.2753/MTP1069-6679190202

[22] J. A. Fredricks, A. L. Reschly, and S. L. Christenson, "Interventions for student engagement: Overview and state of the field," *Handbook of student engagement interventions*, pp. 1–11, 2019.

[23] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.

[24] R. Aguayo, C. Vargas, I. Emilia, and L. M. Lozano, "A meta-analytic reliability generalization study of the maslach burnout inventory," *International journal of clinical and health psychology*, vol. 11, no. 2, pp. 343–361, 2011. [Online]. Available: http://www.redalyc.org/articulo.oa?id=33716996009

[25] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," *International journal of medical education*, vol. 2, p. 53, 2011.

[26] N. Shrestha, "Factor analysis as a tool for survey analysis," *American journal of Applied Mathematics and statistics*, vol. 9, no. 1, pp. 4–11, 2021.

[27] S. Ray and N. Danks, "Seminr." [Online]. Available: https://cran.r-project.org/web/packages/seminr/vignettes/SEMinR.html

[28] Z. Yue, Y. Qin, Y. Li, J. Wang, S. Nicholas, E. Maitland, and C. Liu, "Empathy and burnout in medical staff: mediating role of job satisfaction and job commitment," *BMC public health*, vol. 22, no. 1, pp. 1–13, 2022.

[29] K. S. Namahoot, W. Punpairoj, C. Wattana, and V. Rattanawiboonsom, "Augmented reality technology adoption in tourism using structural equation model," *International Journal of Professional Business Review: Int. J. Prof. Bus. Rev.*, vol. 8, no. 8, p. 61, 2023.

# Journal of Artificial Intelligence and Computing Applications

MAIKRON

Volume 3 · Issue 1 · January–June 2025 · Open Access (CC BY 4.0) · ISSN Pending ·
DOI 10.5281/zenodo.16958297 · Published by Maikron · maikron.org/jaica