Artificial Intelligence Algorithms to Predict College Students' Dropout: A Systematic Mapping Study

Henrique Soares Rodrigues, Eduardo da Silveira Santiago,
Gabriel Monteiro de Castro Xará Wanderley, Laura O. Moraes, Carlos Eduardo Mello,
Reinaldo Viana Alvares and Rodrigo Pereira dos Santos

Graduate Program in Computer Science (PPGI) at the Universidade Federal do Estado do Rio de Janeiro (UNIRIO),

Keywords: Artificial Intelligence, Machine Learning, Algorithm, Students, Dropout, College, University, Systematic Mapping Study.

Abstract: Higher Education Institutions (HEIs), including universities, colleges, and faculties, must develop strategies to mitigate students' dropout rates in undergraduate courses. This is crucial for fulfilling their social role, delivering high-quality professionals to society, contributing to economic development, and preventing the resource wastage. In this context, artificial intelligence (AI) algorithms have emerged as powerful tools capable of predicting dropout rates and identifying undergraduates at risk. This study aims to investigate and discuss the state-of-the-art in applying AI algorithms to address students' dropout. To achieve this objective, a systematic mapping study (SMS) was conducted, encompassing 223 studies at first. Finally, 23 studies were selected for in-depth analysis to explore the effectiveness of AI algorithms in predicting students' dropout. Furthermore, we identified key methodological design issues associated with the application of these AI algorithms, including common features and challenges in implementing these methodologies. This study contributes by providing practitioners and researchers with an overview of the main challenges faced by AI algorithms in predicting students' dropout, highlighting issues related to modeling, experimental methodology, and problem framing.

1 INTRODUCTION

Higher Education Institutions (HEIs) aspire for their students to undergo both academic and professional success, as it contributes to economic growth and social justice. However, one of the most problematic issues that HEIs face is the dropout of students (Realinho et al., 2022). The definition of dropout in this study is from Kehm *et al.* (Kehm et al., 2019): students leaving their university studies before having completed their study program and obtained a degree. Temporary dropout due to illness or pregnancy, for example, is not considered dropout in this context.

According to Bardagi *et al.* (Bardagi and Hutz, 2005), reducing the dropout rates at HEIs is not only an educational issue, but also an economic and political issue. The dropout reduction may have a positive impact on students' professional and financial trajectory, and it may reduce the waste of HEIs' resources. To address the student dropout issue in HEIs, artificial intelligence (AI) algorithms have been recognized as potential tools. They can identify students at risk of leaving educational institutions, enabling these institutions to develop policies that support students in continuing their studies until graduation. Therefore, this study focuses on the use of AI algorithms to predict dropout rates and identify undergraduate students at risk of dropping out.

The objective of this study is to identify the most common algorithms used to predict student dropout, the features used by these algorithms, and the typical challenges in their implementation. To do so, we conducted a systematic mapping study (SMS) to identify and analyze the existing literature on experiments using AI algorithms to predict dropout in HEIs, contributing to an overview of this issue.

The remainder of this paper is structured as follows: Section 2 details previous literature reviews on this topic; Section 3 presents the planning and conduction of this SMS; Section 4 details the results of this SMS; Section 5 discusses the findings of this SMS; Section 6 explores the threats to validity of the

Rodrigues, H., Santiago, E., Wanderley, G., Moraes, L., Mello, C., Alvares, R. and Santos, R.
Artificial Intelligence Algorithms to Predict College Students' Dropout: A Systematic Mapping Study.
DOI: 10.5220/0012348000003636

In *Proceedings of the 16th International Conference on Agents and Artificial Intelligence (ICAART 2024) - Volume 3*, pages 344-351
ISBN: 978-989-758-680-4; ISSN: 2184-433X

SMS; and Section 7 presents final remarks and future work.

2 RELATED WORK

Tete *et al.* (Tete et al., 2022) conducted a systematic literature review to analyze studies related to prediction models for student dropout from HEIs. The authors found that the most common algorithm is the Decision Tree. The most important features were grouped into five categories: socioeconomic (gender, age, professional position, income, ethnic group), academic (grades, Grade Point Average - GPA, frequency), psychological (learning difficulties, academic life satisfaction, sociability), health (well-being, diseases, health issues), and accessibility. This study did not identify any academic projects or actions to decrease student dropout.

Silva and Roman (Silva and Roman, 2021) also conducted a systematic literature review. They found that the most analyzed features in the studies relate to socio-demographic and academic factors, as well as psychological and motivational variables. They also concluded that the most frequently used algorithms are Naive Bayes, KNN, and Random Forest (a combination of several decision trees).

This study, as in the previous reviews, also investigates the most common algorithms and features used to predict student dropout. Our contribution is to also investigate the accuracies reached by the algorithms and the most common limitations and difficulties faced in the implementation of such algorithms, besides validating the previous results found in the literature.

3 RESEARCH METHOD

We performed an SMS based on Kitchenham and Charters (Kitchenham, 2012) and Petersen *et al.* (Petersen et al., 2015) guidelines, which prescribe the following phases: establish research scope, execute search, select studies, extract data, and perform analysis. The study was documented via Parsifal[1], an online tool to support SMS and it is detailed in the following subsections.

3.1 Search Strategy and Data Source

The research question that expresses the goal of this study was formulated following the criteria specified at the PIO (Population, Intervention, and Outcome), as shown in Table 1. Therefore, the formulated research question (RQ) is *"How are the artificial intelligence algorithms used to predict dropout rates among higher education students?"*.

Table 1: PIO structure to formulate the research question.

PIO	
Population	Higher Education Institutions Dropout
Intervention	Artificial Intelligence Algorithms
Outcome	Algorithms, Difficulties, Accuracies, and Features

The desirable outcome of this research is to understand which AI algorithms are most commonly used, which variables are used by these algorithms, how well these algorithms can predict student dropout in terms of accuracy, and the most common difficulties and limitations on the implementation of such algorithms. Moreover, to expand the comprehension of the research question, the following sub-questions (Sub-Q) were formulated:

(Sub-Q1): What are the biggest difficulties in using AI to predict university dropout rates?

(Sub-Q2): How do the AI algorithms use features to predict university dropout rates?

The sources to search by the existing studies were: ACM Digital Library, IEEE Xplore, and Scopus.

3.2 Search String

A generic search string was created from the keywords and their synonyms. Keywords were connected using the AND logical operator, whereas variations and synonyms were connected using the OR operator. The terms of the search string were selected to conduct a broader search including a wide range of studies. We tested different configurations of the search string in Scopus, which is considered the largest scientific publication database that indexes the most relevant publication venues. After calibrating the search string, the final version was:

("higher education" OR "college" OR "graduation" OR "university") AND ("predict*") AND ("artificial intelligence" OR "AI" OR "data science" OR "deep learning" OR "machine learning") AND ("drop off" OR "drop out" OR "dropout")

[1]https://parsif.al/

3.3 Selection Criteria

To properly address the research question and its sub-questions, we established selection criteria to include studies relevant to the topic and exclude those that are not. In this study, publication year was not deemed a relevant criterion. The adopted selection criteria are shown in Table 2. No criteria were set regarding the publication date, and studies from any country were considered acceptable.

Table 2: Selection Criteria.

Inclusion Criteria	
IC1	Study describes an AI technique for predicting dropouts in higher education.
Exclusion Criteria	
EC1	Study describes an AI technique for predicting dropouts in elementary, high school, or massive open online courses (MOOCs).
EC2	Duplicate study.
EC3	Study is not available for reading and data collection (files paid for or not made available by search engines).
EC4	Study is not peer-reviewed.
EC5	Secondary study.
EC6	Study is not written in English.
EC7	Study is not within the topic of AI techniques to predict higher education dropout.

3.4 Study Selection Process

After retrieving studies from the sources, the following filters were used to select the studies: I) title, abstract, and keywords screening; II) introduction and conclusion screening; and III) full text screening.

3.5 Data Extraction

We extracted the following data for each of the accepted studies: Study ID, reference, algorithm(s) used, features used, algorithm accuracy, and limitations of the study. The extracted data were saved in a spreadsheet form and later used to support the discussion of the SMS results.

4 RESULTS

In this section, we present the survey's main findings.

4.1 Sources of Studies

The number of studies retrieved from each source is described in Table 3. From the search in the chosen sources, 223 studies were retrieved: 31 were retrieved by IEEE Xplore, 3 by ACM Digital Library, and 189 by Scopus.

After applying the inclusion and exclusion criteria and filtering, 23 studies were selected, as shown in Tables 4 and 5. Not all features used in the studies are displayed in the tables. When multiple algorithms were used in a study, the algorithm with the highest accuracy was selected.

Table 3: Number of studies by source.

Quantity of studies by source	
IEEE Xplore	31
ACM Digital Library	3
Scopus	189

4.2 Filtering

The filtering process is described in Figure 1.

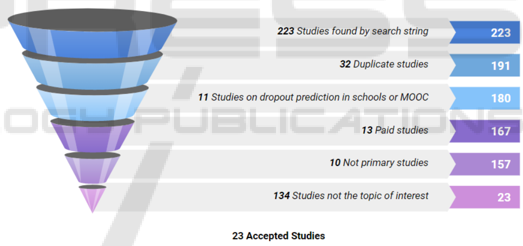

Figure 1: Studies' filtering process.

From the set of studies retrieved, 11 were excluded because the dropout prediction focused on basic education or MOOCs. Thirteen studies were excluded due to paid access. Moreover, some studies were about AI algorithms to predict academic performance, not focusing on dropout risk.

4.3 Country of Origin of the Studies

The selected studies analyzed dropout behavior in HEIs of different countries. Figure 2 describes the number of studies conducted in each country.

We identified one study from Portugal, Hungary, Vietnam, the United Kingdom of Great Britain and Northern Ireland, Costa Rica, the United States of America, Brazil, Malaysia, Saudi Arabia, Italy, and India. We identified 2 studies each from Colombia, Chile, and Spain. Finally, we identified 4 studies

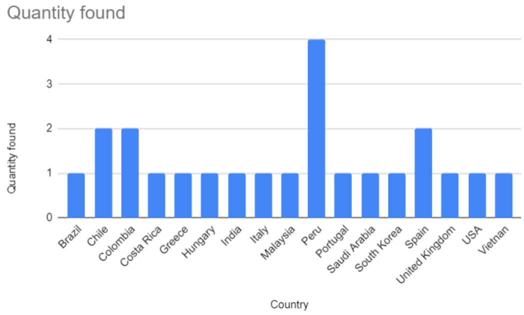

Figure 2: Studies' filtering process.

from Peru, the country with the most studies identified in this SMS. From the set of 23 selected studies, 10 were from Latin America, 7 were from Europe, 5 were from Asia, and one study was from North America.

4.4 Algorithms in the Studies

The percentage of each algorithm found in each study is described in Figure 3.

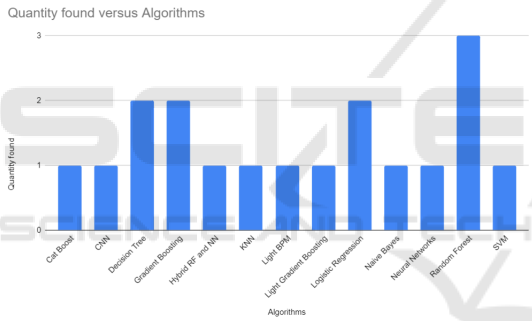

Figure 3: Algorithms explored in the selected studies.

In the case of studies that compared a set of algorithms, we only considered in Figure 3 the algorithm with the highest accuracy; therefore, it does not represent the total percentage for each algorithm used in the studies. In other words, it represents only the algorithm with the highest accuracy of each selected study. After performing data extraction from the selected studies, it was possible to answer the subquestions, presented in the next section.

4.5 (Sub-Q1): What Are the Biggest Difficulties in Using AI to Predict University Dropout Rates?

The most common difficulties and limitations are related to data availability and its small volume, they are usually data from the authors' HEIs affiliation. This limitation causes biases in the analyses. Another limitation is that data may differ in time, courses, and different HEIs, such as the behavior of dropout rates and student satisfaction with academic life.

By acknowledging and actively working to overcome these challenges, higher education institutions can harness the potential of AI algorithms to make significant strides in supporting student success and retention. Collaboration among institutions and researchers can facilitate the sharing of knowledge and resources, thus creating more robust, unbiased, and adaptable predictive models.

Two studies proposed two new algorithms to predict student dropout (S12 and S22). Both studies claimed very high accuracy for their algorithms, which should be replicated in other datasets to confirm such results.

4.6 (Sub-Q2): How do AI Algorithms use Features to Predict Higher Institutions' Dropout?

We found that the most commonly used variables to predict HEI student dropout can be grouped into socioeconomic (gender, age, professional position, income, ethnic group), academic (grades, GPA, frequency, scores at entrance exams, quantity of failed disciplines), and psychological (satisfaction with the academic life, sociability). The majority(11) of the analyzed studies used academic and socioeconomic variables, only a few used(2) psychological variables, and none used physical health and accessibility-related variables. Thus, we could not verify Tete *et al.* (Tete et al., 2022) results.

The analyzed studies on this SMS did not explore major differences between gender, ethnicity, and age group on the behavior of dropout prediction. However, it does not refute the existence of differences between these social groups.

The most important factors related to college students' dropout are academic performance, such as grades, GPA, attendance in class, and credits taken. The most important external factors are the psychological state of the student, such as satisfaction with academic life and addiction to drugs. Finally, the most used variables in AI algorithms to predict student dropout are related to academic performance.

5 DISCUSSION

Several researchers around the globe are investigating AI algorithms to predict student dropout, testing algorithms, such as Random Forest, Cat Boost, Logistic

Table 4: Results extracted from the studies.

ID	Reference	Algorithm	Main Variables	Best Accuracy	Limitations
S1	(Realinho et al., 2022)	Random Forest	Marital status, parent's formation	N/A	Bias may occur
S2	(Nagy and Molontay, 2023)	Cat Boost	Hungarian entrance exam scores, course, gender, age	0.84	Limited to Budapest
S3	(Osorio and Santacoloma, 2023)	Logistic Regression	Depression, drug addictions	0.80	Not mentioned
S4	(Anh et al., 2023)	Light Gradient Boosting	Grades in subjects, attendance in classes	0.95	Bias may occur
S5	(López-Angulo et al., 2023)	Structural Equation Modeling	Satisfaction with HEI	N/A	Satisfaction with academic life may change in time
S6	(Jimenez-Macias et al., 2022)	Random Forest	Grades, Employment, credits	0.99	Few data
S7	(Gutierrez-Pachas et al., 2023)	CNN	Grades, GPA, HDI	0.98	Unequal behaviours
S8	(Zihan et al., 2023)	Light BPM	Grades, GPA	0.93	Not mentioned
S9	(Kotsiantis et al., 2003)	Naive Bayes	Occupation, grades, attendance on tutoring	0.83	Not mentioned
S10	(Moseley and Mead, 2008)	Decision Tree	Grades, age, gender	0.94	Few data
S11	(Solis et al., 2018)	Random Forest	Average of Grades, academic records	0.91	Few data
S12	(Zhang and Rangwala, 2018)	Iterative Logistic Regression	Scores of SAT and ACT	0.98	New proposed algorithm
S13	(Pachas et al., 2021)	Random Forest	Quantity of fails	0.78	Lack of data diversity
S14	(Caselli Gismondi and Urrelo Huiman, 2021)	Neural Networks	Grades, use of mobiles	0.87	Not mentioned
S15	(Fernández-García et al., 2021)	Gradient Boosting	Not mentioned	0.72	Privacy issues
S16	(Santos et al., 2020)	Decision Tree	GPA, Entrance exam scores	0.95	Unbalanced classes

Regression, Neural Networks, Decision Tree, Naive Bayes, KNN, Gradient Boosting, CNN, Light Gradient Boosting, Light BPM, and SVM. Some selected studies in this SMS tested more than one algorithm. In such cases, this study reported the algorithm with higher accuracy. The Random Forest algorithm is the most frequent algorithm with better performance. Additionally, the difficulties reported are mostly related to the unavailability of large data sources because most of the analyzed studies used data provided

Table 5: Results extracted from the studies.

ID	Reference	Algorithm	Main Variables	Best Accuracy	Limitations
S17	(S Sani et al., 2020)	Gradient Boosting	Academic year, high-school GPA, channels of admission	0.93	Not mentioned
S18	(Uliyan et al., 2021)	Neural Networks	Grades, GPA	0.90	Not mentioned
S19	(Agrusti et al., 2020)	CNN	Not mentioned	0.94	Data accuracy required.
S20	(Opazo et al., 2021)	Gradient Boosting	Grades, GPA	0.69	Different HEIs may need different methods
S21	(Ramirez et al., 2022)	Random Forests	Grades, age, gender, academic credits	0.99	Not mentioned
S22	(Daza et al., 2022)	Hybrid Random Forest and Neural Networks	Gender, Age, Academic Credits	0.99	New proposed algorithm
S23	(Revathy et al., 2022)	K-nearest neighbors	Not mentioned	0.97	Not mentioned

by the authors' affiliated HEIs.

To develop a more reliable AI algorithm to predict student dropout, it is necessary to retrieve anonymized data from several HEIs in a large data source. However, it is a hard task to execute since different HEIs have different data formats, such as grades that can be expressed on a scale of 0 to 10, on a scale from F to A, or another format and variables, by different legislations, such as the General Data Protection Regulation (GDPR) from the European Union (European Commission, 2016) or the General Law on Data Protection (LGPD) from Brazil (Brasil, 2018).

Collaborative efforts among HEIs, researchers, and regulatory bodies are essential to overcome these challenges. Establishing data-sharing agreements that adhere to legal requirements while facilitating the exchange of anonymized data for research purposes can help unlock the potential for more reliable AI algorithms. Furthermore, initiatives to create standardized data formats and encourage transparency in data collection practices can contribute to the development of a more cohesive and effective research ecosystem focused on predicting student dropout.

The majority of the analyzed algorithms used data related to academic performance, such as grades and GPA, to predict student dropout, or concluded that such categories of features are the most significant for making such predictions. However, it was not explored how grades are influenced by another variable. In future work, it will be possible to investigate how AI algorithms predict academic performance, such as based on grades.

Another aspect to be explored is the influence of non-academic features on academic performance. These could include socioeconomic factors, such as family background, financial stability, and access to support services. Additionally, personal factors such as motivation, study habits, and mental health can significantly impact a student's grades. Investigating how these variables interact with academic performance can help create a more comprehensive understanding of the factors contributing to student dropout risk.

Moreover, a subject of interest could be the temporal aspect of academic performance prediction. Analyzing how students' grades evolve and how early warning signs in academic performance can be identified can be crucial for proactive interventions to prevent dropout. Furthermore, the application of advanced AI techniques, such as machine learning interpretability methods, could help shed light on how certain features or variables contribute to academic performance predictions. This can provide valuable insights into the underlying mechanisms that drive the results of AI models.

6 THREATS TO VALIDITY

The main threats to this SMS are related to the strategies adopted to create the search string, retrieve primary studies, and extract data from these primary studies. The completeness of this SMS may have been affected by the missing relevant primary studies because some of them may not be retrieved by the search string, or because some of them were excluded by EC3 due to paid access. The authors are aware that considering only peer-reviewed studies on the topic of using AI algorithms for predicting HEI student dropout does not allow for the generalization of the results, as there may be relevant content on this topic in grey literature, such as technical reports.

In addition, the quality of this SMS may also be influenced by potential biases introduced during the selection and inclusion of primary studies. The criteria used to determine which studies to include and exclude could inadvertently introduce bias, affecting the overall comprehensiveness and representativeness of the findings.

7 CONCLUSION

We performed an SMS in which 23 studies were selected for analysis. The results reveal that several HEIs around the globe are testing algorithms to predict student dropout, trying to find the most significant features, sharing their limitations, and trying to maximize the algorithms' accuracy.

From the results, we conclude that there is no specific recommended algorithm to predict higher education students' dropouts. Many studies test different algorithms to perform this task, looking for the one with the highest accuracy. In our search, the Random Forest algorithm was the one that had a better performance in most of the studies. The most recommended features are related to academic performance, such as grades, GPA, credits taken, and attendance in class. Psychological health features, such as satisfaction with academic life, drug addiction, and mental diseases are also present but are less used. The most common difficulties in implementing these AI algorithms are related to the unavailability of a large quantity of data to be used and the diversity of realities in which different HEIs and undergraduate courses are inserted.

Based on this study, we hope to contribute to the field by providing the current overview of the AI algorithms used in predicting HEI students' dropout.

ACKNOWLEDGEMENTS

This study was partly financed by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Universidade Federal do Estado do Rio de Janeiro (UNIRIO).

REFERENCES

Agrusti, F., Mezzini, M., and Bonavolontà, G. (2020). Deep learning approach for predicting university dropout: A case study at roma tre university. *Journal of e-learning and knowledge society*, 16(1):44–54.

Anh, B. N., Giang, N. H., Hai, N. Q., Minh, T. N., Son, N. T., and Chien, B. D. (2023). An university student dropout detector based on academic data. In *2023 IEEE Symposium on Industrial Electronics & Applications (ISIEA)*, pages 1–8.

Bardagi, M. and Hutz, C. S. (2005). Evasão universitária e serviços de apoio ao estudante: uma breve revisão da literatura brasileira. *Psicologia Revista*, 14(2):279–301.

Brasil (2018). Lei nº 13.709, de 14 de agosto de 2018. *Diário Oficial da República Federativa do Brasil*.

Caselli Gismondi, H. E. and Urrelo Huiman, L. V. (2021). Multilayer neural networks for predicting academic dropout at the national university of santa - peru. In *2021 International Symposium on Accreditation of Engineering and Computing Education (ICACIT)*, pages 1–4.

Daza, A., Guerra, C., Cervera, N., and Burgos, E. (2022). A stacking based hybrid technique to predict student dropout at universities. *J Theor Appl Inf Technol*, 100(13):1–12.

European Commission (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance).

Fernández-García, A. J., Preciado, J. C., Melchor, F., Rodriguez-Echeverria, R., Conejero, J. M., and Sánchez-Figueroa, F. (2021). A real-life machine learning experience for predicting university dropout at different stages using academic data. *IEEE Access*, 9:133076–133090.

Gutierrez-Pachas, D. A., Garcia-Zanabria, G., Cuadros-Vargas, E., Camara-Chavez, G., and Gomez-Nieto, E. (2023). Supporting Decision-Making Process on Higher Education Dropout by Analyzing Academic, Socioeconomic, and Equity Factors through Machine Learning and Survival Analysis Methods in the Latin American Context. *Education Sciences*, 13(2).

Jimenez-Macias, A., Moreno-Marcos, M., Merino, P., Ortiz, M., and Delgado-Kloos, C. (2022). Analyzing feature importance for a predictive undergraduate student dropout model. *Computer Science and Information Systems*, 20:50–50.

Kehm, B. M., Larsen, M. R., and Sommersel, H. B. (2019). Student dropout from universities in Europe: A review of empirical literature. *Hungarian Educational*

Research Journal, 9(2):147 – 164. Place: Budapest, Hungary Publisher: Akadémiai Kiadó.

Kitchenham, B. A. (2012). Systematic review in software engineering: Where we are and where we should be going. In *Proceedings of the 2nd International Workshop on Evidential Assessment of Software Technologies*, EAST '12, page 1–2, New York, NY, USA. Association for Computing Machinery.

Kotsiantis, S. B., Pierrakeas, C. J., and Pintelas, P. E. (2003). Preventing student dropout in distance learning using machine learning techniques. In Palade, V., Howlett, R. J., and Jain, L., editors, *Knowledge-Based Intelligent Information and Engineering Systems*, pages 267–274, Berlin, Heidelberg. Springer Berlin Heidelberg.

López-Angulo, Y., Sáez-Delgado, F., Mella-Norambuena, J., Bernardo, A. B., and Díaz-Mujica, A. (2023). Predictive model of the dropout intention of chilean university students. *Frontiers in Psychology*, 13:893894.

Moseley, L. and Mead, D. (2008). Predicting who will drop out of nursing courses: A machine learning exercise. *Nurse education today*, 28:469–75.

Nagy, M. and Molontay, R. (2023). Interpretable dropout prediction: Towards xai-based personalized intervention. *International Journal of Artificial Intelligence in Education*.

Opazo, D., Moreno, S., Álvarez Miranda, E., and Pereira, J. (2021). Analysis of first-year university student dropout through machine learning models: A comparison between universities. *Mathematics*, 9:2599.

Osorio, J. and Santacoloma, G. (2023). Predictive model to identify college students with high dropout rates. *Revista Electrónica de Investigación Educativa*, 25:1–10.

Pachas, D. A. G., Garcia-Zanabria, G., Cuadros-Vargas, A. J., Camara-Chavez, G., Poco, J., and Gomez-Nieto, E. (2021). A comparative study of who and when prediction approaches for early identification of university students at dropout risk. In *2021 XLVII Latin American Computing Conference (CLEI)*, pages 1–10.

Petersen, K., Vakkalanka, S., and Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and software technology*, 64:1–18.

Ramirez, J. A. R., Garcia-Bedoya, O., and Galpin, I. (2022). Maximizing student retention using supervised models informed by student counseling data. 3282:225–239.

Realinho, V., Machado, J., Baptista, L., and Martins, M. V. (2022). Predicting student dropout and academic success. *Data*, 7(11).

Revathy, M., Kamalakkannan, S., and Kavitha, P. (2022). Machine learning based prediction of dropout students from the education university using smote. In *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pages 1750–1758.

S Sani, N., Fikri, A., Ali, Z., Zakree, M., and Nadiyah, K. (2020). Drop-out prediction in higher education among b40 students. *International Journal of Advanced Computer Science and Applications*, 11:550–559.

Santos, G., Belloze, K., Tarrataca, L., Haddad, D., Bordignon, A., and Brandao, D. (2020). Evolvedtree: Analyzing student dropout in universities. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 173–178.

Silva, J. and Roman, N. (2021). Predicting dropout in higher education: a systematic review. In *Anais do XXXII Simpósio Brasileiro de Informática na Educação*, pages 1107–1117, Porto Alegre, RS, Brasil. SBC.

Solis, M., Moreira-Mora, T., Gonzalez, R., Fernandez, T., and Hernandez, M. (2018). Perspectives to predict dropout in university students with machine learning. In *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*, pages 1–6.

Tete, M. F., Sousa, M. d. M., de Santana, T. S., and Silva, S. F. (2022). Predictive models for higher education dropout: A systematic literature review. *Education Policy Analysis Archives*, 30:(149).

Uliyan, D., Aljaloud, A. S., Alkhalil, A., Al Amer, H. S., Mohamed, M. A. E. A., and Alogali, A. F. M. (2021). Deep learning model to predict students retention using blstm and crf. *IEEE Access*, 9:135550–135558.

Zhang, L. and Rangwala, H. (2018). Early identification of at-risk students using iterative logistic regression. In *2018 International Conference on Artificial Intelligence in Education*.

Zihan, S., Sung, S.-H., Park, D.-M., and Park, B.-K. (2023). All-year dropout prediction modeling and analysis for university students. *Applied Sciences*, 13:1143.