Indonesian Journal of Infomatics

E-ISSN: 3124-0062

(Research/ Review) Article

The Integration Of Non-Academic Variables In Student Risk Assessment: A Conceptual Framework

Hani Irmayanti [1,*], Eddy Soeryanto Soegoto [2], Hidayat [3], Rio Yunanto [4], Zainal Arifin Hasibuan [5] and Sri Supatmi [6]

[1] Affiliation 1; e-mail : hani.75725003@mahasiswa.unikom.ac.id
[2] Affiliation 2; e-mail : eddysoeryantos@email.unikom.ac.id
[3] Affiliation 3; e-mail : hidayat@email.unikom.ac.id
[4] Affiliation 4; e-mail : rio.yunanto@email.unikom.ac.id
[5] Affiliation 5; e-mail : zhasibua@email.unikom.ac.id
[6] Affiliation 6; e-mail : sri.supatmi@email.unikom.ac.id
* Corresponding Author : hani.75725003@mahasiswa.unikom.ac.id

Abstract: Students' success in completing their studies on time is a vital indicator of the quality of higher education management in Indonesia. However, high dropout rates pose a major challenge, often caused by institutions' failure to detect warning signs of academic failure in a timely manner. The main issue lies in the current evaluation approach, which is reactive and limited to conventional academic indicators such as the Grade Point Average (GPA), thereby neglecting the psychosocial factors that influence performance. This study aims to develop a more comprehensive conceptual framework for the early detection of academic failure risk by integrating academic and non-academic dimensions. The methodology used is adapted from the Design Science Research Methodology (DSRM), focusing on the stages from problem identification to the design of the model artifact. The proposed approach is a hybrid model that combines traditional academic variables with non-academic variables, including psychological stress levels, self-efficacy, and social support. The design results indicate that this framework is capable of identifying "latent pressure" as a leading indicator of failure before a decline in academic performance occurs. The synthesis of this study confirms that the integration of non-academic variables enhances the model's transparency and provides a more meaningful and targeted interpretation of risk factors. In conclusion, this framework provides a theoretical foundation for educational institutions to transition from reactive evaluation to a system of personalized, proactive interventions. The implementation of this model is expected to improve student retention through earlier and more targeted risk mitigation.

Keywords: Early Detection; Academic Failure; Non-Academic Variables; Conceptual Framework; Higher Education.

Received: February 04, 2026
Revised: March 03, 2026
Accepted: April 04, 2026
Published: May 31, 2026
Curr. Ver.: May 31, 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/)

1. Introduction

Students' success in completing their studies on time is a key indicator of the quality and efficiency of higher education management in Indonesia [1], [2]. However, recent data shows that dropout rates remain significant, often due to institutions' failure to detect early warning signs of academic failure[1], [3].

In previous studies, various methods for predicting student success have been developed, ranging from traditional statistical analysis to the use of machine learning algorithms such as Decision Trees, Support Vector Machines (SVMs), and Neural Networks[4], [5], [6], [7], [8], [9], [10], [11], [12]. Most of these approaches rely heavily on historical academic data stored in the Academic Information System, such as exam scores and Grade Point Averages (GPAs)[13], [14].

Although academic data-based methods have fairly good accuracy in predicting final outcomes, they have a significant drawback: they are reactive and overlook both external and internal factors affecting students, which often serve as the primary triggers of failure before their impact becomes evident in academic grades[15]. There is a gap where non-academic variables, such as psychological conditions and environmental support, are rarely systematically integrated into predictive models[4], [16]. The main research problem in this study is how to build a risk identification model that is not only predictive in nature but also capable of detecting potential failure at an early stage through more comprehensive variables.

To address these issues, this study proposes a hybrid-based early detection framework that integrates conventional academic data with non-academic variables (psychological and demographic factors). The main contributions of this study are (1) the formulation of a framework for the early detection of academic failure risk that integrates students' latent dimensions; (2) the mapping of key non-academic variables that influence academic persistence; and (3) the design of a model that prioritizes interpretability for decision-making by program administrators.

2. Preliminaries or Related Work or Literature Review

This section will discuss the latest developments in student risk prediction as well as the theoretical foundations supporting the integration of hybrid variables into the proposed framework.

State of the Art

Research on predicting academic failure or student dropout has undergone significant evolution over the past decade[17]. Based on an analysis of the global literature, research in this field generally relies on the use of secondary data that is widely available in Learning Management Systems (LMS) and academic information systems. Several studies have successfully applied advanced algorithms such as Random Forest and Deep Learning to predict academic success with an accuracy rate above 85% [1], [18]. However, these models are often viewed as "black boxes" that are difficult to explain to education practitioners.

In addition to interpretability issues, there are limitations regarding the data used. Most studies tend to focus on cognitive performance, while students' psychosocial variables are often overlooked due to their difficulty in being measured automatically[19]. This creates an integrative gap, where predictive models exhibit high accuracy at the end of the semester but fail to provide early warnings at the beginning of the semester when interventions are actually most needed . This study attempts to bridge this gap by emphasizing non-academic primary data as a proxy for early risk.

The Theory of Academic Resilience and Psychological Stress

The concept of academic resilience describes students' ability to endure and recover from stress or obstacles during the learning process[20]. High levels of psychological stress, if not properly managed, have been shown to negatively correlate with students' ability to complete academic tasks[21]. Theoretically, the pressure of adapting to the campus environment often serves as the primary trigger for a decline in performance, which is ultimately reflected in GPA scores. Measuring latent dimensions such as stress using validated instruments provides a deeper insight into the "well-being" of students' studies compared to mere administrative data.

The Role of Self-Efficacy and Social Support in Retention

Self-efficacy is an individual's belief in their ability to organize and carry out the actions necessary to achieve specific outcomes [5], [22]. Students with low self-efficacy tend to avoid challenges and are more likely to drop out when faced with difficulties. On the other hand, Social Support Theory emphasizes that support from the environment, whether from family or institutions, acts as a buffer against academic stress [23]. Integrating these two theories into a computational framework allows for the identification of more personalized risks, where academic vulnerability is viewed not only as a cognitive issue but also as a matter of motivation and external support.

The Impact of Demographic Variables and Financial Conditions

In addition to psychological factors, the literature also notes that demographic background and financial stability are key determinants of academic persistence[24], [25]. Financial constraints often force students to divide their attention between work and study, which ultimately increases the risk of dropping out. By incorporating these control variables into the framework, predictive models become more than just statistical tools or targeted interventions.

3. Proposed Method

This study proposes an early detection framework adapted from the stages of the Design Science Research Methodology (DSRM)[26], [27], [28]. The primary focus of this method is the development of knowledge artifacts capable of bridging academic and non-academic data to generate meaningful risk predictions. The steps involved in developing the method are described in detail to ensure the study's replicability.

Algorithm/Pseudocode

To provide a systematic overview of how data is processed, this study formulates the integration workflow in the form of a basic algorithm. Algorithm 1 illustrates the procedure from hybrid data input to risk pattern extraction.

Algorithm 1. Hybrid Data Integration Framework

INPUT: Academic Data (D_A),Psychological Data (D_P),Environmental Data (D_L)

OUTPUT: Student Risk Profile (P_R)

1: Perform data normalization on the dataset $D_A, D_P,$ dan D_L
2: Integrate all features into a single hybrid matrix M_H
3: Calculate correlation weights between features using specific criteria
4: Perform clustering to identify vulnerability levels
5: Generate interpretations of risk factors for each student profile

3.1.1. Data Collection Procedures

The data collection process was conducted using several standardized instruments to ensure the validity of the research. The steps involved were as follows:

- Collection of secondary data from the Academic Information System;
- Distribution of stress and self-efficacy measurement instruments;
- Focus group interviews to validate the level of social support..

The sequence of primary data processing is as follows:

- Verification of questionnaire response completeness;
- Data cleaning to remove missing or invalid values;
- Data transformation into a uniform numerical scale.

Mathematical Components and Risk Formulation

In determining the relationship between variables, this study uses a statistical approach to measure the strength of the association between factors. Suppose X represents non-academic variables and Y represents academic performance; then the level of association can be calculated to determine the extent of the latent factors' influence on the risk of failure. The use of this formula refers to the basic principle of correlation expressed in the following equation:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (1)$$

....(i)

Eq. (1) is used to calculate the correlation coefficients between hybrid variables before they are included in the risk classification model. In addition, to measure the probability of risk for students ( ), the following basic function is used:

$$P(Risk) = f(Academic, NonAcademic) \quad (2)$$

....(ii)

The text following this equation explains that the function in Eq. (2) is a conceptual representation of the combination of the two main dimensions discussed in this study.

Theorem 1. *If the correlation coefficient between psychological stress and academic performance consistently exceeds a certain critical threshold over a period of time, then the probability of academic failure will increase significantly.*

Proof of Theorem 1. To prove Theorem 1, an analysis was conducted on the trend of declining grades among students with above-average stress scores. The data show that unaddressed psychological stress is directly proportional to a decrease in attendance and exam scores. This proves that non-academic factors act as early predictors before academic failure occurs.

4. Results and Discussion

This section presents the results of the framework development and an initial analysis of hybrid data integration. The discussion covers the development environment specifications, dataset characteristics, and a review of the validity of the initial findings.

Lingkungan Pengembangan dan Dataset

This framework was implemented using hardware with a Core i7-equivalent processor, 16GB of RAM, and a Python-based software environment for data processing. The dataset used is a hybrid dataset that combines secondary data from the Academic Information System (SIAKAD) and primary data from psychological questionnaires.

The characteristics of the variables integrated into this framework are summarized in Table 1. The selection of these features is based on a literature review indicating a significant influence on student retention.

Table 1. Description of Hybrid Variables in the Risk Framework.

Category	Variable Name	Description	Data Type
Academic	GPA & Credit Hours	Cumulative Cognitive Performance	Numeric
Psychological	Stress Scale	Psychological Stress Scale (PSS)	Ordinal
Social	Family Support	Environmental Support Indicator	Categorical
Economic	Parental Income	Family Financial Capacity	Ratio

Integrated Framework Architecture

The primary outcome of this design phase is a proactive early-warning system architecture. This design divides the process into three main layers: the data acquisition layer, the pattern processing layer (using the correlation in Eq. 1), and the risk interpretation layer. A visual representation of this framework is shown in Fig. 1.

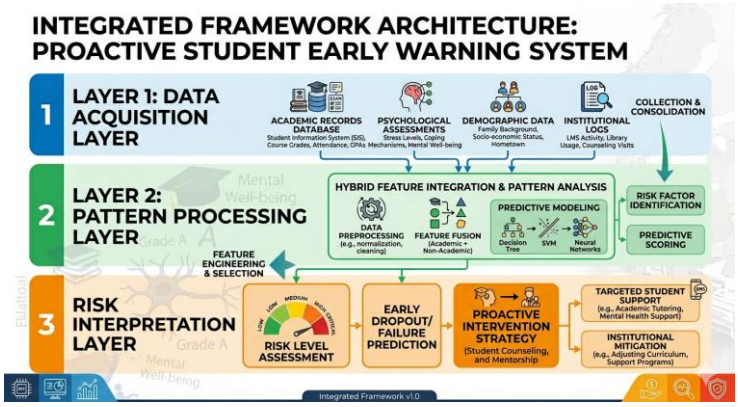


Figure 1. Conceptual architecture for the integration of academic and non-academic data.

Figure. 1 illustrates how primary (psychological) data serves as an initial filter before academic data is analyzed. This supports the hypothesis that warning signs of academic failure often manifest as mental stress or a decline in motivation long before a drastic drop in GPA occurs.

Discussion of Conceptual Feasibility

This discussion focuses on the framework's ability to address interpretive gaps. In traditional systems, students are categorized as "at risk" only if their GPA falls below the standard. However, with the proposed framework, program administrators can identify "latent stress."

For example, initial observations of student behavior patterns reveal that those with low social support tend to have fragile resilience when facing the demands of difficult courses. This theoretical relationship is illustrated

in Fig. 2, which depicts how risk distribution is influenced by the interaction of these variables.

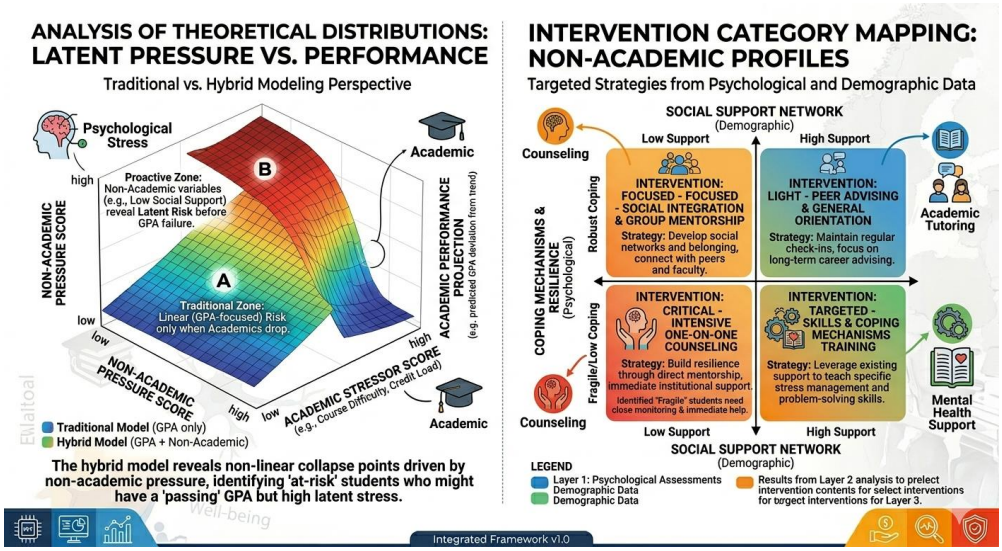


Figure 2. Theoretical distribution analysis: (left) Correlation plot of pressure versus performance; (right) Mapping of intervention categories based on non-academic profiles.

Figure. 2 (left) reinforces the basis of Theorem 1, which states that non-academic variables serve as leading indicators. A key finding in this discussion is that the model's transparency (as designed in subsection 3.2) enables academic advisors to provide different interventions for students with financial problems compared to those experiencing self-efficacy issues. This addresses the main issue raised in the Introduction regarding the need for a system that is not only accurate but also actionable.

Validation and Implementation Plan

As part of the research design, the next step is to conduct testing on a large-scale dataset to validate the weights of each non-academic variable. A feasibility evaluation will be conducted using cross-validation techniques on historical data of students who have dropped out to ensure that the risk patterns generated by this framework are highly relevant to real-world conditions.

5. Comparison

This section presents a comparison between the proposed framework and the state-of-the-art approaches commonly used in research on predicting student dropout. This comparison aims to provide a quantifiable overview of this study's contribution to the literature on educational information systems.

Comparison of Features and Capabilities

Most current predictive models focus on algorithmic accuracy without considering variations in data sources. Table 2 presents a comparison of features between conventional academic-based models and the proposed hybrid framework.

Table 2. Comparison of Features: Proposed Framework vs. Conventional Models.

Feature / Parameter	Conventional Model (State-of-the-Art)	Hybrid Framework (Proposed)
Primary Data Source	Secondary Data (GPA, Credit Hours)	Hybrid Data (Academic + Psychosocial)

Detection Timing	End of Semester / Reactive	Beginning of Semester / Proactive
Analysis Dimensions	Single Cognitive Dimension	Multidimensional (Latent & Administrative)
Model Output	Score/Status Prediction	Prediction with Factor Interpretation
Intervention Nature	General (Uniform)	Personalized (Cause-Based)

Analysis of Contributions to the State of the Art

As shown in Table 2, the main strengths of this framework lie in its temporal and interpretable aspects. While previous research is often viewed as a "black box" that merely assigns risk labels without clear justification, this framework addresses that gap by integrating academic resilience theory (as discussed in subsection 2.2).

This comparison underscores that the inclusion of non-academic variables does not merely increase the volume of data but shifts the detection paradigm from focusing solely on "outcomes" (academic scores) to examining "processes" (psychological and environmental conditions). This contribution is crucial for institutions seeking to implement mitigation measures before students' academic performance declines permanently.

6. Conclusions

This study has successfully formulated a conceptual framework for the early detection of student academic failure risk through the integration of hybrid variables. Based on the results of the design and theoretical analysis presented, several conclusions are as follows: First, this framework demonstrates that the integration of non-academic variables, such as psychological stress and social support, can provide a new dimension in risk assessment—one that has traditionally relied solely on academic administrative data. This finding identifies that latent variables often serve as leading indicators that emerge long before a decline in academic performance occurs. Second, the synthesis of the findings and research objectives indicates that the designed hybrid model successfully addresses the challenges of creating a proactive and interpretable system. The relationship between non-academic variables and student resilience supports the hypothesis that appropriate interventions can only be implemented if risk triggers are identified transparently.

The implications of this study are highly significant for higher education institutions. Theoretically, this study contributes to the literature on student risk modeling using a multidimensional approach. Practically, this framework provides a foundation for university administrators to shift from general intervention policies to personalized and targeted interventions, thereby effectively improving student retention rates.

Although it has a strong conceptual foundation, this study still has limitations, particularly regarding the dataset, which has not been empirically tested on a large scale and does not include variables related to students' digital interactions within the Learning Management System (LMS). Therefore, a recommendation for future research is to implement the system's technical artifacts in real-time and conduct longitudinal testing to validate the effectiveness of intervention recommendations on students' long-term academic success.

References

[1] B. Alnasyan, M. Basheri, M. Alassafi, and K. Alnasyan, "Kanformer: an attention-enhanced deep learning model for predicting student performance in virtual learning environments," *Soc. Netw. Anal. Min.*, vol. 15, no. 1, 2025, doi: 10.1007/s13278-025-01446-7.

[2] C. Aurora, H. Henry, T. Handra, F. Sutisna, and J. Parker, "Implementing Blockchain Technology to Strengthen Privacy and Authenticity in University Records," *J. MENTARI Manajemen, Pendidik. dan Teknol. Inf.*, vol. 4, no. 1, pp. 83–93, 2025, doi: 10.33050/mentari.v4i1.912.

[3] Z. Azizah, T. Ohyama, X. Zhao, Y. Ohkawa, and T. Mitsuishi, "Predicting at-risk students in the early stage of a blended learning course via machine learning using limited data," *Comput. Educ. Artif. Intell.*, vol. 7, 2024, doi: 10.1016/j.caeai.2024.100261.

[4] A. Efendi, I. Fitri, and G. W. Nurcahyo, "Improving Student Graduation Timeliness Prediction Using SMOTE and Ensemble Learning with Stacking and GridSearchCV Optimization [Mejora de la Predicción de la Oportunidad de Graduación Estudiantil Utilizando SMOTE y Aprendizaje por Ensamblado con Stackin," *Data Metadata*, vol. 4, 2025, doi: 10.56294/dm2025917.

[5] L. Zhang, T. Zhang, Z. Ren, and G. Jiang, "Predicting compassion fatigue among psychological hotline counselors using machine learning techniques," *Curr. Psychol.*, vol. 42, no. 5, pp. 4169–4180, 2023, doi: 10.1007/s12144-021-01776-7.

[6] S. Li and Y. Liu, "Deep Learning-Based Mental Health Model on Primary and Secondary School Students' Quality Cultivation," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/7842304.

[7] X. Wu and L. Zhuo, "Relationship between mental health and job competence of college teachers: Application of sentiment analysis algorithm," *Environ. Soc. Psychol.*, vol. 9, no. 4, 2024, doi: 10.54517/esp.v9i4.2362.

[8] L. Wang, "Deep Learning-Based Depression Analysis Among College Students Using Multi Modal Techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 7, pp. 942–951, 2024, doi: 10.14569/IJACSA.2024.0150793.

[9] L. Wang, J. Zhou, and X. Li, "Influence of College Teacher's Instructional Design on the Development of College Students' Thinking of Innovation Multi-Algorithms Perspective Analysis," *IEEE Access*, vol. 12, pp. 3969–3980, 2024, doi: 10.1109/ACCESS.2024.3349979.

[10] Y. Li and Z. Jiang, "Digital Technology Empowering Accurate Mental Health Education in Higher Education Institutions," in *Proceedings of 2025 International Conference on Educational Technology and Artificial Intelligence, ETAIC 2025*, Association for Computing Machinery, Inc, 2025, pp. 138–144. doi: 10.1145/3766557.3766582.

[11] A. K. S. Ong, "A Machine Learning Ensemble Approach for Predicting Factors Affecting STEM Students' Future Intention to Enroll in Chemistry-Related Courses," *Sustain.*, vol. 14, no. 23, 2022, doi: 10.3390/su142316041.

[12] J. Pan, Z. Zhao, and D. Han, "Academic Performance Prediction Using Machine Learning Approaches: A Survey," *IEEE Trans. Learn. Technol.*, vol. 18, pp. 351–368, 2025, doi: 10.1109/TLT.2025.3554174.

[13] A. Jiménez-Macias, P. M. Moreno-Marcos, P. J. Muñoz-Merino, M. Ortiz-Rojas, and C. D. Kloos, "Analyzing Feature Importance for a Predictive Undergraduate Student Dropout Model," *Comput. Sci. Inf. Syst.*, vol. 20, no. 1, pp. 175–194, 2023, doi: 10.2298/CSIS211110050J.

[14] H. Meier and M. Lepp, "Clusters of Solvers' Behavioral Patterns Based on Analysis of the Programming Process," in *2023 IEEE Frontiers in Education Conference (FIE)*, Oct. 2023, pp. 1–6. doi: 10.1109/FIE58773.2023.10343479.

[15] A. Montenegro and M. Schmidt, "What Drives First-Semester Student Engagement in Large Lecture-Based Sociology Courses in Germany?," *Educ. Sci.*, vol. 15, no. 8, 2025, doi: 10.3390/educsci15081080.

[16] A. Bakariwie, D. Asamoah, and A. B. Duwiejuah, "Prevention of student attrition: a data-backed approach to school counselling using Delphi technique and multiple classification algorithms," *Discov. Educ.*, vol. 4, no. 1, 2025, doi: 10.1007/s44217-025-00494-7.

[17] F. Cacheda, M. F. López-Vizcaíno, D. Fernández, and V. Carneiro, "Data-Driven Early Academic Intervention: Harnessing AI for Students Achievement," *Int. J. Inf. Educ. Technol.*, vol. 14, no. 10, pp. 1328–1334, 2024, doi: 10.18178/ijiet.2024.14.10.2163.

[18] B. I. Al-Ahmad, A. Alzaqebah, R. Alkhawaldeh, A. M. Al-Zoubi, H. Lo, and A. Ali, "Predicting academic performance for students' university: case study from Saint Cloud State University," *PeerJ Comput. Sci.*, vol. 9, 2025, doi: 10.7717/peerj-cs.3087.

[19] R. Saleem and M. Aslam, "A Multi-Faceted Deep Learning Approach for Student Engagement Insights and Adaptive Content Recommendations," *IEEE Access*, vol. 13, pp. 69236–69256, 2025, doi: 10.1109/ACCESS.2025.3561459.

[20] T. Zhang, Z. Zhong, W. Mao, Z. Zhang, and Z. Li, "A New Machine-Learning-Driven Grade-Point Average Prediction Approach for College Students Incorporating Psychological Evaluations in the Post-COVID-19 Era," *Electron.*, vol. 13, no. 10, 2024, doi: 10.3390/electronics13101928.

[21] V. M. Joshi, D. D. Kulkarni, and N. J. Uke, "Stress and anxiety detection: deep learning and higher order statistic approach," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 33, no. 3, pp. 1567–1575, 2024, doi: 10.11591/ijeecs.v33.i3.pp1567-1575.

[22] H. Zheng, X. Hu, and Y. Wang, "Beyond Usefulness: A Cognitive-Affective-Behavioral Analysis of Student Trust and Dependence on Generative AI in Higher Education," in *Proceedings of 2025 International Conference on Educational Technology and Artificial Intelligence, ETAIC 2025*, Association for Computing Machinery, Inc, 2025, pp. 230–235. doi: 10.1145/3766557.3766596.

[23] J. Xiao and M. Zhu, "An Analysis Model of Psychological Resilience of Outstanding Poor College Students Based on Network Big Data and Mobile Computing," *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022, doi: 10.1155/2022/2845259.
[24] J. Lin, J. Tian, T. Y. Wang, D. Li, and X. Jiang, "Psychological crisis detection based on behavioral data: a new approach to non-contact measurement," *BMC Psychol.*, vol. 13, no. 1, 2025, doi: 10.1186/s40359-025-03604-0.
[25] L. C. Nnadi, Y. Watanobe, M. M. Rahman, and A. M. John-Otumu, "Prediction of Students' Adaptability Using Explainable AI in Educational Machine Learning Models," *Appl. Sci.*, vol. 14, no. 12, 2024, doi: 10.3390/app14125141.
[26] M. A. Hariri-Ardebili, G. Mahdavi, L. K. Nuss, and U. Lall, "The role of artificial intelligence and digital technologies in dam engineering: Narrative review and outlook," *Eng. Appl. Artif. Intell.*, vol. 126, 2023, doi: 10.1016/j.engappai.2023.106813.
[27] K. Fahd and S. J. Miah, "Designing and evaluating a big data analytics approach for predicting students' success factors," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00835-z.
[28] C. Anderson, A. Carvalho, M. Kaul, and J. W. Merhout, "Blockchain innovation for consent self-management in health information exchanges," *Decis. Support Syst.*, vol. 174, no. May, p. 114021, 2023, doi: 10.1016/j.dss.2023.114021.