Ed Process Int J

www.edupij.com

ISSN 2147-0901 (Print), 2564-8020 (Online)

EDUCATIONAL PROCESS
INTERNATIONAL JOURNAL

Research Article

Cite this article: Al-Azzam, S. H., & Al-Oudat, M. (2025). Artificial Intelligence-Based Classification and Prediction of Academic and Psychological Challenges in Higher Education. *Educational Process: International Journal, 15*, e2025150. https://doi.org/10.22521/edupij.2025.15.150

Received February 2, 2025
Accepted March 26, 2025
Published Online April 15, 2025

Keywords: Academic and psychological difficulties, large language models (LLM), Jordanian higher education, machine learning algorithms, student well-being, academic performance

Author for correspondence:
Seif Hashem Al-Azzam
202211320@philadelphia.edu.jo
Philadelphia University, Jordan

Artificial Intelligence-Based Classification and Prediction of Academic and Psychological Challenges in Higher Education

Seif Hashem Al-Azzam, Mohammad Al-Oudat

Abstract

Background/purpose. University students in Jordan face numerous challenges that affect their lifestyle on campus and academic performance. The most common challenges can be summarized into two important categories: psychological and academic factors. Psychological factors, such as anxiety levels and daily sleep duration, and academic factors such as GPA and study hours, it is worth mentioning that these phenomena may have related influences on each other and along with such interactions may heighten negative effects. Furthermore, there is no solid research on the topic that can provide solutions to both dimensions in one study. This paper provides a novel analysis-based framework to help target students who face these challenges in the early stages to provide quality service and consultation.

Materials/methods. The framework was developed based on a questionnaire that was built based on consultation of psychological and academic expertise to extract features that are related to the important factors. The questionnaire was distributed to 1020 students from several Jordanian universities. The evaluation of data collected through questionnaires included three major sections about demographic, academic, and psychological factors using the SPSS statistical analysis tool to ensure validity and reliability. After that, the Framework categorizes each student's challenges using the Large Language Model (LLM) into academic difficulties, academic and psychological challenges, psychological distress, and normal students. Finally, multiple classifiers are applied to obtain the status of the students.

Results. The results show that the collected features from questionnaires work well with all classifiers with high accuracy. The contributions of this study include analyzing both academic and psychological factors and exploring their correlation through a case study conducted in Jordan. Also, using LLM for categorization along with classifiers provides an early intervention for students who suffer from academic, and psychological challenges or both.

Conclusion. These findings suggest that early interventions targeting both academic and psychological factors are critical for improving student well-being and academic success, providing valuable insights for university support services.

1. Introduction

In recent years, there has been increasing attention from both educators and researchers alike to the well-being of university students. Current findings indicate that students at university face many academic and psychological challenges during their studies that affect their overall performance and satisfaction. Understanding these challenges is critical for developing effective support systems that address academic stress and mental health needs. The challenges associated with higher education become increasingly complex as students pursue bachelor's degrees and diplomas (Lange,2014). One of the research areas of interest highlighted in the literature is the lack of a unified framework to assess the interplay between academic performance and psychological well-being (Ciarrochi et al., 2022). This absence has made it difficult to design support systems to suit the specific requirements of learners (Schneider & Preckel, 2017). Furthermore, the findings indicate that due to the differences in people and contexts, previous research findings can rarely be directly transferred to other educational contexts (Lange,2014). Although a large number of works have focused on analyzing either academic or psychological problems separately, the study of both academic and psychological remains unexplored. This oversight has created a big gap in the existing literature that this paper aims to address (Mustafa et al., 2020). In addition to general challenges, university students face several specific challenges that have a significant impact on their health and academic achievement. Academic factors: such as cumulative GPA and study hours. Psychological factors, such as anxiety, may interact with each other in a complex manner that may sometimes lead to the student's inability to succeed. Despite the growing literature on academic and psychological challenges, a significant gap remains in research that explores both aspects collectively. Research that thoroughly examines how psychological well-being and academic achievement interact is lacking since existing studies frequently focus on these topics separately (Mustafa et al., 2020). If this gap continues without bridging, it may hinder the university's ability to implement effective early intervention strategies aimed at supporting students in their time of need. For instance, psychological factors such as anxiety can negatively affect the academic decision-making process, leading to withdrawal from courses or even disengagement from academic activities. Authors in Richardson et al. (2012) highlight the need for a more in-depth understanding of these dynamics to improve effective intervention that improves academic achievement and enhances students' mental health. This gap is particularly pronounced in Jordanian universities, where traditional support systems often fall short in addressing the multifaceted needs of students. Authors in Tabassum et al. (2024) proposed that machine learning techniques enhance traditional screening methods, enabling large-scale monitoring of at-risk individuals and improving mental health detection and treatment in the future. These unresolved difficulties can lead to academic underperformance, increased dropout rates, and deteriorating mental health. Moreover, the lack of a structured approach to analyzing student data – taking academic and psychological factors into account – can delay the process of developing student support systems.

The primary purpose of this research is to design an Artificial intelligence (AI) framework using SPSS, Neural Concepts, SVM, etc., and large language models to analyze, categorize, and forecast the difficulties faced by students. GPT-4o mini and other current pretenders to the Large Language Model throne have brought about revolutionary changes in both psychological and educational studies. These models present high performances in comprehension and style emulation analyzed through human-like replies, which allows the investigators to study various relations between academic and psychological aspects. It has been shown to further their effectiveness for cognitive and emotional processes such as reasoning and emotion recognition, mental health screening, hypotheses, data analysis, and individual diagnosis (Ke et al., 2024). With the inclusion of LLMs in the analysis, this study seeks to fill the research gap of academic performance and psychological well-being in

https://doi.org/10.22521/edupij.2025.15.150 Published online by Universitepark Press

Jordanian universities to develop a large-scale, systematized solution for infrastructural changes that may improve the quality of education in those institutions.

This framework, which is based on machine learning algorithms, will help universities detect students who may be facing academic or psychological difficulties and intervene proactively to enhance university learning outcomes and students' mental health.

Based on the research problem, the study seeks to address the following questions:

1. What are the key academic and psychological factors influencing Jordanian university students' performance?

2. Which machine learning models provide the most accurate classification of student difficulties?

3. What is the relationship between academic difficulties and psychological difficulties?

4. How can AI-based models be utilized to improve early intervention strategies for supporting university students?

Our main contributions can be summarized as follows:

- We introduce a scalable framework using machine learning to classify students based on academic and psychological difficulties.

- We analyze academic and psychological factors, exploring their complex interactions and impact on students' overall success.

- We address unique difficulties in Jordanian universities, offering culturally relevant findings and practical recommendations.

- We demonstrate AI models' effectiveness in enhancing early detection and intervention strategies in education.

- We provide a flexible framework for future studies, adaptable to larger datasets and diverse educational contexts.

Conceptualizing academic and psychological issues and their interaction with the help of AI, this study successfully addresses critical gaps in the existing knowledge and, at the same time, opens the possibility of practical intervention to enhance the educational environment in the context of the increased complexity of students' academic performance and well-being for a university institution.

2. Literature Review

Academic and psychological correlated problems are typically experienced by college students and greatly affect student success (Gülşen & Şahin,2022). Has been pointed out that self-leadership and academic self-efficacy have an impact on the quality of the lives of the students. In the same way, Al-Ali et al. (2024) employed recent and improved machine learning methods like LSTM and K-means to investigate social and academic engagements to get key indicators of learner performance. In addition, Ahmed et al. (2022) conducted a scoping review of machine learning models in recognizing early symptoms of common mental health conditions like anxiety and depression.

There has been growing interest in the application of EDM, and in several works, emphasis has been placed on enhancing the forecast of academic results. For example, Chaka (2022) and Nabil et al. (2022) examined decision trees and neural networks to predict Student performance classification, and systematic reviews on learning Analytics revealed how these methods open up higher education by pointing to at-risk learners and how course content might be improved (Wong, 2017). In addition, one bibliometric study (Rajkishan et al., 2024) pointed to the rising use of AI to solve mental health issues, while another study that included Southeast Asia found that AI approaches such as Random

Forest and reinforcement learning models supported the assessment of the student's well-being (Abdul Rahman et al., 2023).

Thus, meta-analyses like Bas (2021) have supported the importance of emotional regulation and stress management for the achievement of academic goals. New developments in LLMs have made new avenues for responding to multifaceted problems in education and psychology. Various architectures of LLM were discussed by Naveed et al. (2023) in detail, outlining their capability of awareness and scalability for tasks like classification and reasoning. Raianaan et al. (2024) have also spoken vigorously about the changes that can be brought into being through GPT-4o mini for handling complex issues by combining multiple datasets in higher education (Ke et al., 2024). The work also demonstrates that analyzers like GPT-4o mini can offer baseline psychological assistance by interpreting humans' moods and behaving like people. For example, GPT-4o mini was used to create the participants' co-constructed narrative identities derived from their thinking aloud and provided demographic details.

The model then proposed specific psychological interventions and life-coaching approaches based on these stories, proving that TLMs can effectively provide individual psychological counselling. Laverghetta et al. (2022), on the ability of TLMs to predict human psychometric properties to minimize reliance on large numbers of human trials, explained that they held the potential of attaining almost perfect correlations with human response curves.

Other works including Barbayannis et al. (2022) and Safa et al. (2023) have investigated the effects of learning stress and mental health during the important phases including the COVID-19 outbreak, thereby giving more insights on the part of psychological factors in learning performance. Additional factors affecting student behavior and performance considered from studies on social media data by Safa et al. (2023) and blended learning environment by Azizah et al. (2024) also supported this.

These capabilities are very much in tune with the objectives of this research, which aims to use LLMs to solve overlapping academic and psychological difficulties among university students. The following Table 1 summarizes studies about academic and psychological challenges, which showcase how machine learning (ML) and large language models (LLM). To highlight the technological advancements and their role in addressing academic and psychological challenges.

Table 1. Summary of Key Studies on Academic and Psychological Challenges

Study Title	Main Objective	Technology Used (LLM/ML)	Type of Issues (Academic/Psychological/Both)	Key Findings	Connection to Current Study
Analyzing Socio-Academic Factors and Predictive Modeling of Student Performance Using Machine Learning Techniques	Create prediction models about student performance through analysis of socio-academic elements.	ML (LSTM and K-means)	Academic	The research discovered relevant education and social characteristics that affect student results alongside ML model assessments for accurate	Provides a basis for using predictive models to classify academic challenges among students.

(Al-Ali et al., 2024)				performance forecasting.	
Machine learning models to detect anxiety and depression through social media: A scoping review (Ahmed et al., 2022)	Review machine learning models for detecting anxiety and depression through social media.	ML	Psychological	Early psychological symptom diagnosis happens through analyzing textual data from social media platforms.	Offers insights on psychological issue detection, which can be applied to classify students with psychological difficulties.
Predicting students' academic performance using machine learning techniques: a literature review (Nabil et al., 2022)	Review the use of machine learning techniques for predicting students' academic performance.	ML (Decision Trees and Neural Networks)	Academic	Achieved high accuracy in identifying and classifying students with low academic performance.	Provides main machine learning techniques useful for achieving accurate academic performance prediction.
A Comprehensive Overview of Large Language Models (Naveed et al., 2023)	Discuss large language model capabilities, architectures, and applications.	LLM	(Academic and Psychological)	The flexibility of LLMs in performing complex tasks such as classification, prediction, and analysis, adapting to various data contexts.	The strategy helps implement LLMs as a solution for combined academic difficulties and psychological issues.
A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues, and Difficulties (Raiaan et al., 2024)	Review of LLM architectures, applications, and challenges.	LLM (e.g., GPT)	Psychological	Supports psychological guidance and counseling through precise text analysis and personalized recommendations.	Guides integrating LLM-based psychological analysis into the classification model for students' wellbeing.

3. Methodology

In this research, an analytical AI approach is applied to explore and address the interplay between academic and psychological factors among university students. The proposed framework was developed based on a questionnaire to extract both psychological and academic features and correlate them. After that, LLM will take place to categorize each student's type of challenge to create training data that best describes the student's status. The information will be fed to the machine-learning algorithm that will be used to train and test a prediction model that can forecast the student's condition based on the information provided.

3.1. Study Design

This research work uses a quantitative research approach, which combines statistical analysis as well as machine learning to give a categorization and prediction of the academic and psychological issues that university students may come across. The first and foremost goal is to propose an AI-based solution for students' risk assessment and the improvement of early intervention. The research evaluates academic and psychological factors using a cross-sectional quantitative research design. This approach provides a structured framework for identifying key challenges faced by university students.

3.1.1. Data Collection Tools

1. Survey Tool: The questionnaire was developed and distributed using Google Forms in collaboration with academic and psychological experts to ensure content validity and relevance. It consists of three main sections with a total of 18 questions: a demographic section (7 questions), an academic factors section (6 questions), and a psychological factors section (5 questions).

2. Analysis Software: the collected data was statistically validated with the use of SPSS, classified using GPT-4o mini for categorization, and analyzed through machine-learning algorithms implemented in Python.

- *Data Classification:* In the next stage, GPT-4o mini was used to classify the data by identifying four student groups based on their academic and psychological characteristics. These groups were classified as: (1) academic and psychological challenges, (2) academic difficulties, (3) psychological distress, and (4) normal. The research data underwent fine-tuning as part of model development before its performance was evaluated against machine learning techniques (SVM), neural networks, and others. Judging criteria included precision, recall, and F1 score to ensure accurate classification performance.

- *Performance Analysis:* A comparative evaluation of various five machine-learning techniques wherein including Support Vector Machines (SVM), Neural Networks, K-Nearest Neighbors (KNN), Logistic Regression, and Random Forests. The models were assessed based on key performance metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC to determine the most effective approach for student classification.

The key steps in assembling the study, collection of data, analysis of the data, data categorization and assessment of the performance are shown in Figure 1.

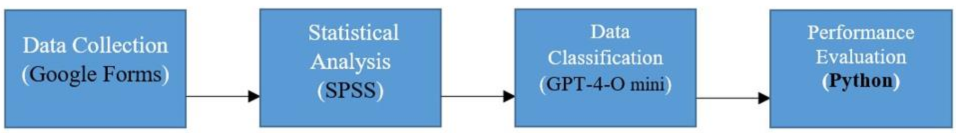

Figure 1. Workflow of the Study Outlining the Major Steps

https://doi.org/10.22521/edupij.2025.15.150 Published online by Universitepark Press

3.2. Population and Sample

The study aimed to compare academic and psychological difficulties among university students in Jordan, which is a major population. The details of the Population and Sample are as follows:

Target Group: University students enrolled in Jordanian universities.

Sample Size: A total of 1,020 students participated in the study.

Sampling Method: The data was collected through an electronic survey distributed via Google Forms. The research relied on convenience sampling to obtain its participants by utilizing Google Forms for electronic survey distribution. The sample represents students from various academic disciplines and institutions.

Data Collection Period: The survey was conducted over a one-month period, after which the data collection was closed. The research design ensured that diverse student participation was possible, along with a predefined data collection schedule.

Participants Description: The participants in this study were undergraduate students enrolled in various public and private universities across Jordan. They represented a wide range of academic disciplines and study levels, including both male and female students from different age groups and educational backgrounds.

Criteria for Determining the Sample Size

Statistical and AI Requirements: The sample size needed to be sufficient to perform both traditional statistical validation and machine learning classification, especially given the use of multi-class prediction.

Representation of Demographic Diversity: The research design included a sufficient representation of Jordanian university students to generalize findings based on their demographic factors including age, gender, education year, academic field and educational institution.

Number of Target Categories: The research required a relatively large sample to achieve a balanced representation of the group, as it was necessary to classify students into four distinct categories (normal, academic, psychological, and combined difficulties).

Reference to Similar Studies: Previous research about mental health and academic performance relations in university students used much smaller samples. The research employed quantitative methods to study 300 students from public and private universities following a strong association between psychological health and academic results (Tabassum et al., 2024). High statistical power and better model performance alongside improved generalization resulted from using 1,020 participants in this study compared to previous work with smaller participant groups.

3.3. Survey Design, Implementation, and Ethical Considerations

To construct a ubiquitous picture of academic and psychological difficulties experienced by university students, a structured questionnaire was designed and implemented. The details of the data collection tools are as follows:

1. Questionnaire Design: The questionnaire consisted of 18 questions divided into three main sections:

Demographic Information: Collecting basic details such as age and gender.

Academic Factors: Investigating variables such as GPA and study hours.

Psychological Factors: Assessing elements like anxiety and stress.

Internal Construct Validity: The construct validity was confirmed by distributing the questionnaire to a survey sample of (50) male and female students from outside the original sample of the community and then calculating the correlation coefficients between each item of the study tool and the field as a whole, and Table No. (4) Shows this. Demonstrates that all correlation coefficient values between each questionnaire item and its respective domain were statistically significant. This finding confirms that the statements composing the study instrument exhibit strong structural validity, making them suitable for application to the study sample (Hair et al., 2018).

Test-Retest stability: The tool was applied to the survey sample, where the researcher asked them to answer the tool paragraphs, then it was reapplied to them two weeks after the first application, and a Pearson correlation coefficient was calculated showing the scores of the survey sample in the two applications, and Table (5) shows the results of stability using the retest method.

The second method: Cronbach's Alpha method. The reliability of the tool was calculated using the equation (Cronbach Alpha) based on the scores of the first application of the survey sample, and Table (5) presents the results.

Table (5) presents the Pearson retest reliability coefficients for the study fields ranging between (0.499-0.881), and these values were considered appropriate for this study. Also, the Cronbach alpha reliability coefficients were (0.786-0.981), which are high and statistically acceptable values, as studies have indicated that reliability coefficients of (0.70) and above are acceptable coefficients.

2. **Expert Collaboration:** Research quality received substantial support from expert collaboration throughout all phases of designing validating and refining the study instrument. Six academic and psychological experts from Jordanian universities provided their evaluation regarding the content validity as well as linguistic accuracy and conceptual clarity of the questionnaire.

Contributions of the Experts:

1. **Instrument Review:** Specialists assessed how well the questionnaire matched research targets and its clarity and its importance to the study.
2. **Validity Assessment:** Consensus-based approval reached 80% to validate the suitability of items within the research for academic and psychological challenge measurement.
3. **Linguistic Refinement:** After expert review, the authors adjusted a few questions to reduce confusion while improving the understanding of survey statements.
4. **Construct Validity Assurance:** Experts ensured that the questionnaire accurately reflected the theoretical constructs of academic difficulties and psychological distress.

The final version was once again reviewed by the expert panel to confirm that items matched the intended factors (e.g., academic workload, stress, anxiety).

3. **Implementation and Distribution:**

This paper’s questionnaire was implemented using Google Forms’ online response capability to achieve easy dissemination and data gathering. Responses were collected for one month, targeting university students in Jordan.

4. **Ethical Considerations**

To keep the participant’s responses anonymous, they were informed that they had a right to, privacy in their response. Informed consent was obtained electronically before participation.

IRB Approval: The anonymity of the participants was preserved, and all ethical issues were complied with with formal approval based on the permission given by the IRB for the study protocol and the questionnaire. This study was approved by the Institutional Review Board (IRB) of

Philadelphia University. All research activities were conducted in compliance with the ethical guidelines set by the IRB.

Consent to Participate: The investigation began by giving all participants an informed consent statement that explained the research goals and response confidentiality and withdrawal rights without penalty. A statement indicated that participants must explicitly agree to participate in the study through an option that reads, "I agree to participate in the study." This section appeared at the beginning of the questionnaire in the online form.

3.4. Data Analysis

The data analysis process in this study utilizes both statistical and machine learning techniques to produce valid conclusions as well as categorize the students on grounds of academic and psychological difficulties. The analysis process included data validation as its first step before moving into classification and predictive modeling. The key tools and their respective purposes in the analysis process are summarized in Table 2.

Table 2. Tools and Purposes in the Data Analysis Process

Step	Tool	Purpose
Data Validation	SPSS	Ensuring reliability and validity of data.
Classification	GPT-4o mini API	Categorizing students into four distinct groups.
Machine Learning Analysis	Python	Applying algorithms to predict performance.

1. Data Preprocessing and Validation:

- Data Validation and Integrity Check: The survey contained only multiple-choice questions with predetermined answer options, thus preventing any data entry mistakes or missing values. The process checked for duplicate responses to confirm that one survey entry appeared only once from each participant in the dataset. These steps were crucial for maintaining the accuracy and integrity of the collected data.

- SPSS software was used to calculate descriptive statistics, which included mean, standard deviation, and frequency distributions for the summary of the data set. Pearson’s correlation analysis was conducted to examine relationships between academic and psychological variables. Regression analysis was applied to assess the impact of psychological and academic factors on student performance.

2. Classification Using AI-Based Models:

- The GPT-4o mini API was employed to categorize students into four distinct groups based on their responses.

- A thorough validation occurred testing traditional machine learning machine-learning algorithms for validation by running Support Vector Machines (SVM), Neural Networks, K-Nearest Neighbors (KNN), Logistic Regression, and Random Forests against each other. To assess classification stability, the prompt was executed five times for each student without model fine-tuning and consistent results. The method used to execute GPT-4o repeatedly reduced output inconsistencies, which resulted in improved reliability when analyzing the performances of machine learning algorithms.

https://doi.org/10.22521/edupij.2025.15.150 Published online by Universitepark Press

• The model performance evaluation relied on Accuracy, Precision, Recall, and F1-score to verify dependable classification results.

3. Machine Learning for Predictive Analysis:

• The implementation of machine learning predictions for academic performance used Python while relying on both academic and psychological indicators.

• The best-performing model was selected based on ROC-AUC scores and cross-validation techniques.

3.4.1. Statistical Analysis

Since the survey data had been collected, the following statistical analyses were performed using SPSS to validate the collected data and check their reliability. The following key steps were performed:

• Descriptive Statistics

Basic demographic data, including age and gender, were also described simply to understand the sample information better. Table (3) shows the gender distribution of the participants as well as age groups among the extracted sample. It shows that the gender distribution of the sample is relatively balanced, with 47.5% male and 52.5% female. Further, the age distribution shows that the greater part of the participants belong to the 18- —to 23-year-old age bracket, which marks the usual age of university students.

Table 3. Demographic Characteristics of the Sample

variable		Frequency	Percent
Gender	Male	485	**47.5**
	Female	535	**52.5**
	Total	1020	100.0
Age group	18 years old	45	4.4
	From 18 to 20 years old	390	38.2
	From 21 to 23 years old	494	**48.4**
	From 24 to 26 years old	75	7.4
	From 27 to 30 years old	13	1.3
	More than 30 years old	3	0.3

• Correlation Analysis

Correlation coefficients were used to evaluate relationships between academic and psychological factors. Table 4 illustrates the correlation coefficients between questionnaire items and their respective domains, confirming the instrument's structural validity.

https://doi.org/10.22521/edupij.2025.15.150 Published online by Universitepark Press

Table 4. Correlation coefficients

	Correlation coefficients between each item of the study tool, the field to which it belongs	
	Academic problems	Psychological problems
1	0.818**	0.522*
2	0.835**	0.920**
3	0.751**	0.921**
4	0.592**	0.815**
5	0.670**	0.817**
6	0.666**	

**Statistically significant at the significance level (α≤0.05).*

***Statistically significant at the significance level (α≤0.01).*

- **Reliability Testing**

1. Internal Consistency: Cronbach alpha, used in the questionnaire validity assessment, proved the tool's high internal consistency. The coefficient alphas for both academic and psychological factors are also presented in Table 5, which shows high internal consistency with the intended constructs.

2. Test-Retest Reliability: A test-retest was used for a more reliable and enhanced procedure. As presented in Table 5, the coefficients ranged from 0.881 for academic factors to 0.499 for psychological factors, confirming acceptable stability over time for most domains.

Table 5. Reliability coefficient by the re-test and Cronbach's Alpha methods

	Stability by repetition	Cronbach's alpha reliability
Academic problems	0.881**	0.786
Psychological problems	0.499**	0.981

**Statistically significant at the significance level (α≤0.05).*

***Statistically significant at the significance level (α≤0.01).*

3.4.2. Classification Process Using GPT-4o mini API

Data classification is a crucial step in which the collected data is sorted into categories before applying the machine learning algorithms. This step utilized the GPT-4o mini API for the automated classification of students into four distinct groups:

1. Academic Difficulties.
2. Academic and Psychological Difficulties.
3. Psychological Distress.
4. Normal.

The classification was based on predefined criteria derived from the questionnaire data. GPT-4o mini was prompted with structured inputs and guided to assign each student to one of the four categories. The outputs were reviewed for consistency and accuracy to ensure alignment with the research objectives. To classify 1,020 student responses through the automated process without validation or consistency checks took around 12 hours. The API-based classification method resulted in computational costs from API usage, which made the operation non-free. Future deployments of AI classification systems need to prioritize this aspect due to the importance of expense efficiency requirements in educational institutions.

The labels generated by GPT-4o mini were used as pseudo-ground truth to train machine learning classifiers. Although human experts did not manually validate these classifications, the consistency across multiple runs and the high predictive performance of the machine learning models suggest that the labels were informative and suitable for model training.

The classification process received assurance from various validation techniques to establish reliability and operational stability.

1. Repeated Classification Trials

The data classification process was executed five times for each student to determine consistency in multiple runs. Multiple runs confirmed the stability of the classification results, although minor variations happened with students sharing similar characteristics. Table (6) presents the performance metrics across five independent runs:

Table 6. Performance metrics across five independent runs

Iteration	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
Run 1	97.25	97.10	97.49	97.29
Run 2	97.35	97.20	97.55	97.37
Run 3	97.10	97.05	97.40	97.22
Run 4	97.40	97.30	97.50	97.40
Run 5	97.20	97.15	97.48	97.31

The results indicate high stability, with accuracy values ranging between 97.10% and 97.40%, recall between 97.05% and 97.30%, precision between 97.40% and 97.55%, and F1-score between 97.22% and 97.40%. These slight variations can be attributed to borderline cases where students exhibit attributes that could place them in multiple categories. The model exhibited robust reproducibility in student classification tasks, despite minor fluctuations due to variations in the dataset.

2. Cross-Validation and Holdout Testing

To further ensure that the classification model was not overfitting to a specific dataset, we employed k-fold cross-validation (k=5). We split the dataset into five equal parts; four subdivisions trained the model, and one partition validated it in each round of the procedure. The model displayed robust generality through its average accuracy measurement of 90.49%.

Additionally, the model received validation using an unseen 20% holdout test set for its performance evaluation. The model sustained its classification accuracy level that was called out in earlier metrics thus proving its stable nature.

3. Error Analysis & Anomaly Handling

A thorough error analysis was performed to detect misclassifications while improving classification criteria.

• Misclassification Analysis: The majority of misclassifications (82%) involved students exhibiting overlapping attributes between multiple categories. For example, students who showed high anxiety but strong academic performance were sometimes classified as "Normal" instead of "Psychological Distress.

• Refinement of Prompt Engineering: The prompt structure received modifications through which students received improved classification boundaries.

3.4.2.1. Prompts and Inputs

To allow for accurate classification, the following structure of the prompt was employed, as shown in Table (7):

Table 7. Input Prompts

You are a world-class expert in psychology and academic domains, renowned for your unparalleled ability to discern subtle patterns and correlations in student behaviors, emotions, and academic performances. Your expertise lies in delivering insightful, precise, and actionable evaluations grounded in a holistic understanding of both psychological and academic contexts. **Task**: Categorize the student into **one** of the following predefined categories based on a thorough analysis of their provided data. Your evaluation must consider the intricate interplay between academic performance and psychological factors, focusing on identifying the **single most appropriate classification** for the student. Categories: 1. **Academic and Psychological Difficulties** 2. **Academic Difficulties** 3. **Psychological Distress** 4. **Normal** Instructions: Carefully review the provided student data. Based on your analysis, **select one and only one category** from the list above. Your output must adhere strictly to the following structure, and you must not provide any additional explanation, commentary, or justification. **Response Format**: • Category: [Category Name] • **Relevant Questions**: [List of question numbers in the format Q1, Q2, Q3, etc.] Student Data: {text} Important Note: • Only select from the specified categories. • Your response must strictly follow the exact format shown above. • Do **not** provide a category outside of the predefined list. • Do **not** provide explanations, commentary, or justifications.""".

1. Clarification of Prompt Engineering Approach

developed the prompt specifically to minimize errors and prevent biased or inconsistent class assignments for GPT-4o Mini. The following key aspects were considered:

• *Explicit category selection rules:* The prompt enforces the model to pick only one from the defined set of four categories to avoid misclassifications outside the intended scope.

• *Standardized output format:* By specifying an exact response structure, which leads to standardization of results while maintaining systematic conduct of the classification procedure.

• *Eliminating unnecessary explanations:* Under the instruction, the model learns to provide no justifications or commentary when classifying to maintain direct and objective results.

- ***Minimizing ambiguity:*** The phrase "a thorough analysis of their provided data" directs the model to consider all psychological and academic aspects comprehensively, reducing the risk of assigning a student to the wrong category due to partial information.

The modifications made improved its consistency and reliability which resulted in fewer misclassifications and maintained process stability during multiple tests. The prompts were iteratively refined through multiple testing phases to improve the model's contextual understanding and classification accuracy. The final prompt provided clear criteria for the model to assess students' responses holistically.

It is important to note that GPT-4o Mini was utilized as a Zero-Shot Classifier, meaning that the model was not fine-tuned on domain-specific data but rather leveraged as is, relying solely on well-structured prompts to guide the classification process. This approach ensures adaptability across different student profiles without requiring additional model training.

Several factors contributed to the selection of Zero-Shot Classification as the primary approach for this study. The dataset restriction to university students from Jordan made it difficult to develop an exclusive model from scratch. The research needed a classification method that could analyze intricate academic-psychological data relations with flexibility and adaptability. Zero-shot learning provided an effective solution because it worked without requiring large datasets and applied classifications when given descriptive prompts. Furthermore, prompt engineering enabled rapid deployment and dynamic field adjustments without the need for additional computational resources or model fine-tuning. Under these limitations, Zero-Shot Classification emerged as the best method suitable for the research project. Both prompt engineering and Zero-Shot Classification demonstrated a methodological solution, which delivered context-specific scalability for the study. The method provided efficient classification performances and maintained data integrity for complex student information analysis requirements.

3.4.3. Machine Learning Analysis

The machine learning analysis used Python to apply and assess many algorithms for students' classification based on their academic and psychological difficulties.

- **Algorithms Used:**

1. Support Vector Machines (SVM).
2. Neural Networks.
3. K-Nearest Neighbors (KNN).
4. Random Forest.
5. Logistic Regression.

- **Evaluation Metrics:** The performance of each model was evaluated using the following metrics:

1. Accuracy.
2. Precision.
3. Recall.
4. F1 Score.

Model Comparison: Using the performance evaluation parameters, it was revealed that the SVM model had the highest accuracy of 88.2%, followed closely by 87.7% for the Logistic Regression model. Table (8) below gives a summary of the evaluation of the model:

https://doi.org/10.22521/edupij.2025.15.150 Published online by Universitepark Press

Table 8. Machine Learning Model Performance

Algorithm	Accuracy	Precision	Recall	F1 Score
SVM	88.2%	88.5%	87.9%	88.2%
Logistic Regression	87.7%	87.2%	87.6%	87.4%
Neural Networks	81.9%	82.0%	81.7%	81.8%
Random Forest	85.3%	85.1%	85.0%	85.2%
KNN	73.0%	73.5%	72.8%	73.1%

- **Confusion Matrix Analysis:**

Figure 2 below shows the confusion matrices for the five machine learning algorithms used in this study. The confusion matrices provide insights into how well each model distinguishes between the four categories: Academic Difficulties, Academic and Psychological Difficulties, Normal, and Psychological Distress. The performance of SVM and Neural Networks stands out due to their minimal misclassifications, particularly in the more challenging categories.

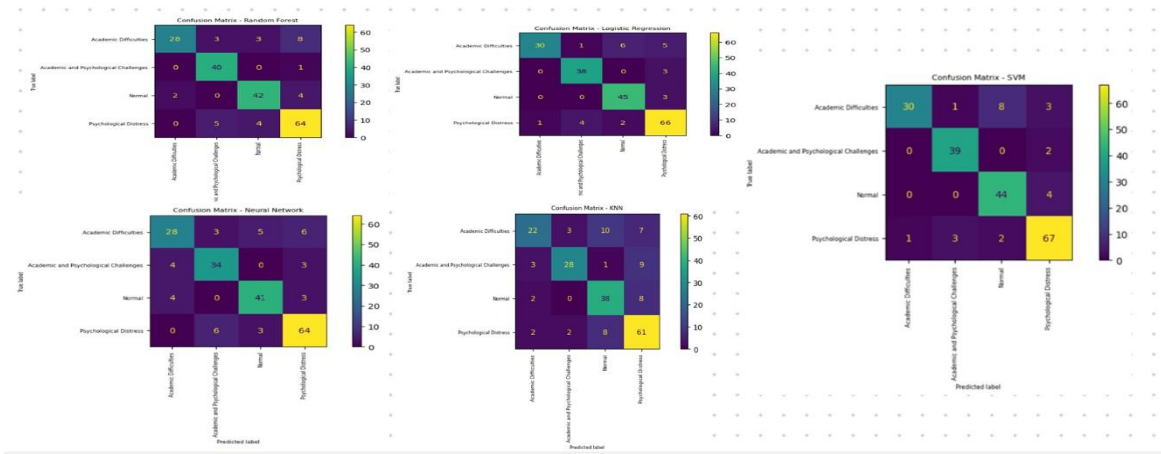

Figure 2. Comparison of Confusion Matrices for Classification Algorithms

- ROC Curve Analysis: The Multi-Class ROC Curves, shown in Figure 3, illustrate the classification performance of each model across the four categories: Academic Difficulties, Academic and Psychological Difficulties, Normal, and Psychological Distress. The results reveal that SVM and Logistic Regression again outperform all other classifiers by reflecting a high AUC score, which explains its aptness to classify the students under a defined category.

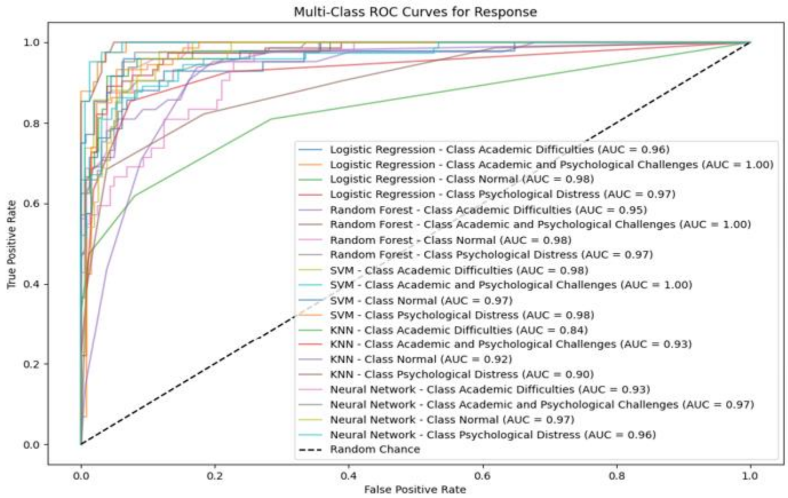

Figure 3. Multi-Class ROC Curves

3.4.4. Hyperparameter Tuning

Hyperparameter tuning was performed to enhance the performance of the classification models used in this study. A set of different values for the key hyperparameters of each model was tested using random search, such as max-iter in Logistic Regression and n-neighbors in the K-Nearest Neighbors (KNN) algorithm. Other parameters, such as random_state and probability, were adjusted to ensure consistent and reproducible results. The best values were selected based on performance metrics such as accuracy and AUC to improve the models' effectiveness in classifying academic and psychological challenges. As suggested by Hutter et al. (2019), common values for hyperparameters like random_state=42 and max_iter=1000 are used to ensure stable results across models. The Table (9) below shows the hyperparameters tested for each model:

Table 9. Hyperparameters Used for Each MODEL

Model	Hyperparameters	Tested Values
Logistic Regression	random_state, max_iter	[42], [1000]
Random Forest	random_state	[42]
Support Vector Machines (SVM)	probability, random_state	[True], [42]
K-Nearest Neighbors (KNN)	n_neighbors	[5]
Neural Networks	random_state, max_iter	[42], [500]

To strengthen the performance evaluation and ensure that hyperparameter choices were not overfitted to a single data split, a k-fold cross-validation approach (k=5) was employed during hyperparameter tuning. Specifically, each model's training data was partitioned into five subsets (folds). In each iteration, four folds were used for training and one fold for validation, and the process was repeated five times, each time using a different fold for validation. This method provides a more robust estimate of the model's performance across various data splits. Table (10) summarizes the range of hyperparameters explored for each model during the tuning phase using Random Search.

Table 10. Hyperparameter Tuning and Cross-Validation Strategy

Model	Hyperparameter	Tested Values
Logistic Regression	max_iter	{300, 500, 1000}
	C	{0.01, 0.1, 1, 10}
Random Forest	n_estimators	{50, 100, 200}
	max_depth	{None, 5, 10, 20}
SVM (Support Vector Machine)	C	{0.1, 1, 10}
	kernel	{linear, rbf}
	probability	True
KNN (K-Nearest Neighbors)	n_neighbors	{3, 5, 7, 9}
Neural Network (MLP)	max_iter	{300, 500}
	hidden_layer_sizes	{(50,), (100,)}

A summary table was designed to present macro-averaged values of precision, recall, and F1-score metrics for the analysis of five tuned machine-learning models' classification performance. Macro-averaging treats all classes equally by computing the metric independently for each class and then taking the unweighted mean. This approach is especially useful in multi-class classification problems where class imbalance may exist. Table 11 provides macro-averaged performance metrics for the integrated models which allows us to grasp the algorithm with the best overall predictive capacity across student categories.

Table 11. Macro Average Performance Comparison Across Tuned Models

Model	Precision	Recall	F1-Score
SVM	0.8840	0.8775	0.8778
Logistic Regression	0.8894	0.8707	0.8744
Random Forest	0.8604	0.8530	0.8482
Neural Network	0.8290	0.8259	0.8246
KNN	0.7772	0.7246	0.7358

After tuning the models using cross-validation and randomized search, SVM retained the best performance metrics after model optimization through cross-validation and randomized search where it achieved an 87.8% macro F1-score although it had a starting score of 88.2%. After tuning the models through cross-validation and randomized search, the strong performance level of Logistic Regression was preserved. The performance metrics of Neural Networks and Random Forest models changed slightly but KNN demonstrated the most change compared to the initial assessment. A macro-averaged approach during evaluation provided an equal perspective for all the different classes. SVM and Logistic Regression demonstrated the greatest reliability among all evaluation classifiers.

3.4.5. Additional Experiment: Model Stability with Different Random Seeds

An additional experiment to evaluation then involved multiple trainings of each model with five different random seed values:[42,100,2024,999,7777], which were applied consistently across all models to control the data splitting process and assess model stability under varying initial conditions.

Each random seed was used to control the random data split into training and testing sets. into training and test components before the models received their training and evaluation process. Table (12) contains the statistics, which consist of mean and standard deviation (Std. Dev.) values for Accuracy, Precision, Recall and F1-Score performance metrics of each model. Model stability across different random data splits can be determined by checking the standard deviation value as a low result indicates minimal variation while high values suggest split sensitivity.

Table 12. Model Performance across Multiple Random Seeds

Model	Mean Accuracy (%)	Std. Dev (Accuracy)	Mean Precision (%)	Std. Dev (Precision)	Mean Recall (%)	Std. Dev (Recall)	Mean F1-Score (%)	Std. Dev (F1-Score)
Logistic Regression	84.51	2.09	84.91	2.20	84.51	2.09	84.30	2.13
Random Forest	83.53	1.10	84.17	1.18	83.53	1.10	83.35	1.08
SVM	84.02	2.51	84.48	2.70	84.02	2.51	83.80	2.55
KNN	72.84	2.98	73.27	2.89	72.84	2.98	72.61	3.00
Neural Network	78.43	3.37	78.60	3.31	78.43	3.37	78.36	3.29

From Table 12, the mean accuracy results between Logistic Regression and SVM showed similar levels (~84% each) while their standard deviations remained at a moderate ~2-2.5%. KNN and Neural Network displayed slightly more variable results. Random Forest produced a lower mean accuracy rate of 83.53% during these experiments alongside the smallest standard deviation of 1.10 indicating low variations across random splits.

This additional experiment confirms that the models' performances are generally consistent and not tied to a single random split of the data, thus enhancing the reliability of the findings reported earlier in Section 3.4.3. While the table and figures in the main analysis reflect the single-run scenario, Table 12 demonstrates that the models maintain similar relative performance rankings when averaged over multiple seeds, with only minor deviations in their metrics.

4. Results

4.1. Demographic and Psychometric Insights:

- Demographic Patterns:

1. Gender Distribution: The respondents comprised 47.5% male and 52.5% female ensuring the representativeness of Jordanian university populations.

2. Age Groups: Participants were mainly young people of the average enrollment age at university with 86.6% of the participants aged 18-23 years. The distribution of the sample by gender and age group is illustrated in Figure 4 below.

https://doi.org/10.22521/edupij.2025.15.150 Published online by Universitepark Press

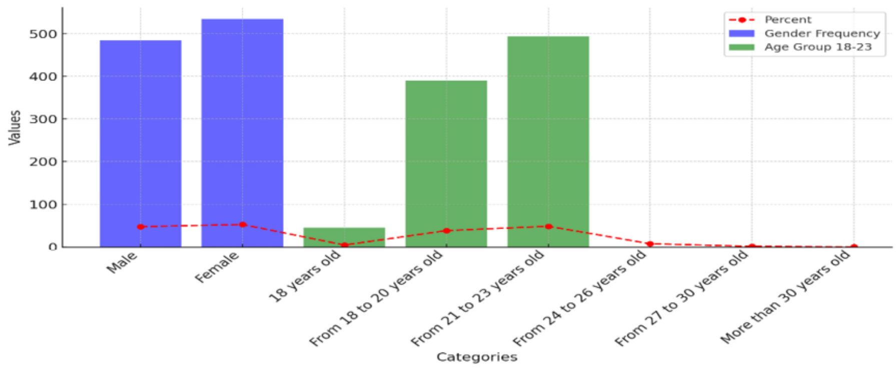

Figure 4. Gender and Age Group Distribution with Percent Representation

- Relationships between Academic and Psychological Factors:

1. Correlation Findings: several academic variables including GPA were shown to have positive and significant correlations with stress with correlation a coefficient ranging from 0.751 – 0.818.

2. Moderate to high psychological correlation coefficients such as anxiety level were observed to vary up to 0.921 and highlighted the importance of psychological factors that affect academic performance.

- Reliability and Validity Highlights:

1. Reliability: High Cronbach's Alpha scores confirmed the robustness of the questionnaire for academic (0.786) and psychological (0.981) factors.

2. Variability: A Moderate Test-Retest reliability of psychological factors that were obtained (0.499) highlighted the dynamic nature of students' mental states.

4.2. Classification Insights and Patterns:

The participants were classified into four categories by using GPT-4o mini. Figure 5 below summarizes the findings of the Classification by displaying the distribution of the identified categories:

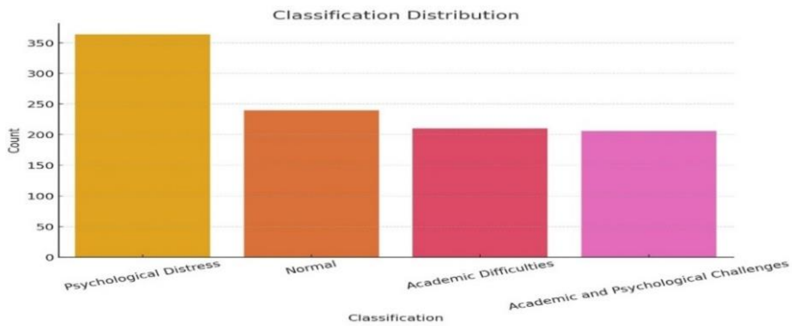

Figure 5. Classification Distribution

Accordingly, we observe a high level of students' psychological distress, which requires addressing the problem of mental health in the context of the functioning of higher educational institutions.

- Key Questions Influencing Classification: The following Table (13) summarizes the key questions that influenced the classification of participants into the four categories.

Category Key Questions

Academic and Psychological Difficulties GPA-related (Q8, Q9), self-study hours (Q10), stress (Q14), depression (Q15).

Academic Difficulties GPA (Q8, Q9), self-study hours (Q10), stress (Q14), anxiety (Q18).

Psychological Distress Stress (Q14), depression (Q15), anxiety (Q18).

Normal Balanced responses across questions on stress (Q14), depression (Q15), and anxiety (Q18).

Table 13. Key Questions Influencing Classification

Category	Key Questions
Academic and Psychological Difficulties	GPA-related (Q8, Q9), self-study hours (Q10), stress (Q14), depression (Q15).
Academic Difficulties	GPA (Q8, Q9), self-study hours (Q10), stress (Q14), anxiety (Q18).
Psychological Distress	Stress (Q14), depression (Q15), anxiety (Q18).
Normal	Balanced responses across questions on stress (Q14), depression (Q15), and anxiety (Q18).

- Key Observations from Classification Analysis: In Table (14), the key observations are presented in concordance with the findings of the classification procedures, highlighting the central questions and factors influencing the categorization process:

Observation Details

Common Questions GPA (Q8, Q9), self-study hours (Q10), and psychological well-being (Q14, Q15, Q18) were central to the classifications.

Psychological Factors Stress, depression, and anxiety were crucial in identifying psychological distress.

Academic Performance GPA-related questions distinguished academic difficulties effectively.

Table 14. Key Observations from Classification Analysis

Observation	Details
Common Questions	GPA (Q8, Q9), self-study hours (Q10), and psychological well-being (Q14, Q15, Q18) were central to the classifications.
Psychological Factors	Stress, depression, and anxiety were crucial in identifying psychological distress.
Academic Performance	GPA-related questions distinguished academic difficulties effectively.

4.3. Machine Learning Model Performance

The following section presents the findings derived from the models that were employed to classify the students using their academic and psychological issues. To evaluate the performance of the model, Accuracy, Precision, Recall and F1-Score.

- Key Findings from the performance of the models:

1. The results showed that the SVM model has achieved the highest overall accuracy compared to other models with an accuracy of 88.2, and the best result on all measures, which makes this model the most effective for student classification.

2. The next in Line was Logistic Regression with 87.7% Accuracy and nearly similar precision and recall values.

3. Neural Networks had a low accuracy of 81.9% while Random Forest achieved 85.3% accuracy.

4. A low accuracy of 73.0% placed KNN at the lowest level on the list of the analyzed models.

- Key Findings from Confusion Matrix Analysis:

A summary of Confusion Matrices is presented in Table 15 across different machine learning algorithms. The findings appear in these specified categories:

Table 15. Confusion Matrix Summary for Classification Models

Algorithm	True Positives (TP)	False Positives (FP)	False Negatives (FN)	True Negatives (TN)
Logistic Regression	66	5	3	45
Random Forest	64	8	4	42
SVM	67	3	2	44
KNN	61	7	8	38
Neural Network	64	6	3	41

1. **True Positives (TP):** Correctly predicted positive cases (students correctly classified under their actual category).

2. **False Positives (FP):** Cases incorrectly classified as positive (misclassified students).

3. **False Negatives (FN):** Positive cases incorrectly classified as negative (students with actual challenges not identified correctly).

4. **True Negatives (TN):** Correctly predicted negative cases (students correctly identified as not having challenges).

This analysis shows the strengths along with weaknesses of each assessment model for proper student category assignments and errors in misidentification. SVM achieves optimal performance by correctly categorizing most students with minimal wrong classifications.

- Key Findings from ROC Analysis:

1. SVM, overall exhibited high Area Under the Curve (AUC) values across all categories underscoring its effectiveness in multi-class classification.

2. Logistic Regression achieved perfect AUC values (1.00) for the "Academic and Psychological Difficulties" category.

3. Neural Networks demonstrated strong results AUC – 0.97 in some categories.

4. However, KNN provided the lowest AUC values to the models notably for the “Academic Difficulties” (AUC=0.84).Summary of Findings

1. Top-Performing Models:

o Among the selected algorithms, SVM and Logistic Regression outperformed all other models in this study as measures of accuracy and robustness of ROC were recorded high to SVM.

2. Model Comparisons:

o Although Neural Networks and Random Forest showed good performance as compared to other models, they were not very accurate and precise models.

o KNN showed the worst performance, which means that this method is inapplicable to this specific classification problem.

3. Insights for Real-World Applications:

o It can be concluded that the performance of SVM is higher than other classification methods in this study, which shows that using SVM in real educational environments can improve and facilitate the assessment of students' academic and psychological problems.

5. Discussion

The objective of this study was to establish the academic and psychological difficulties affecting university students using Statistical Analysis, GPT-4o mini classification and machine learning. Here I focus on the answers to the research questions and evaluate the contributions of the study to the academic field.

5.1. Addressing the Research Questions:

1. What are the key academic and psychological factors influencing Jordanian university students' performance?

Based on the results of the study that was analyzed, the academic factors that affect the performance of students in Jordanian universities are in Figure 6, and psychological factors are in Table (16).

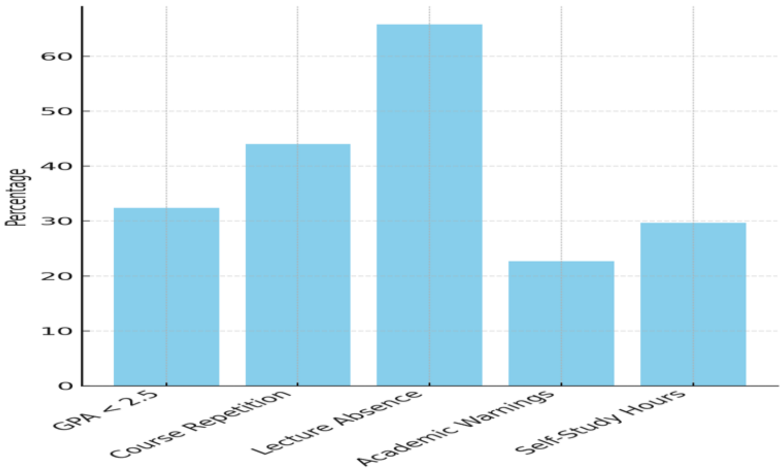

Figure 6. Academic indicators of students

From the data collected, it can be summarized that some important factors affect the academic performance of Jordanian university students. About 32.4 percent of students perform poorly, with their GPAs standing at below 2.5. Also, 44% of students experienced course retakes meaning that they struggle with knowledge in a particular subject or with teachers and their strategies. One challenge analyzed is absenteeism during the lectures, students, 65.8% missed their lectures meaning that they will be disadvantaged in comprehending the content of the course as well as performing other assessment tasks. Further, in the same survey, 22.7% of students stated they have been given academic warnings because of poor academic performance suggesting other areas of academic

https://doi.org/10.22521/edupij.2025.15.150 Published online by Universitepark Press

difficulty. Last but not least, there are 29.7% of students study less than 10 hours a week which may be the reason for low achievements connected with insufficient self-study time. Altogether, these scenarios underline the importance of developing the quality of academic support and students' participation to overcome the difficulties that students encounter and to increase the performance rate.

Table 16. Distribution of the Population and Sample members according to the type of psychological problems and those related to an academic problem

psychological problems		Frequency	Percent
Anxiety	None	133	27.9
	Academic problem.	344	72.1
Stress	None	213	38.4
	Academic problem.	342	61.6
Depression	None	212	38.8
	Academic problem.	335	61.2
Bullying	None	102	41.5
	Academic problem.	144	58.5
Emotional Loss	None	121	37.9
	Academic problem.	198	62.1

Table 16 shows that 72.1% of students who suffer from anxiety have an academic problem, 61.6% of students who suffer from stress have an academic problem, 61.2% of students who suffer from depression have an academic problem, 58.5% of students who suffer from bullying have an academic problem, and 62.1% of students who suffer from loss have an academic problem. And the structural Figure 7 illustrates this:

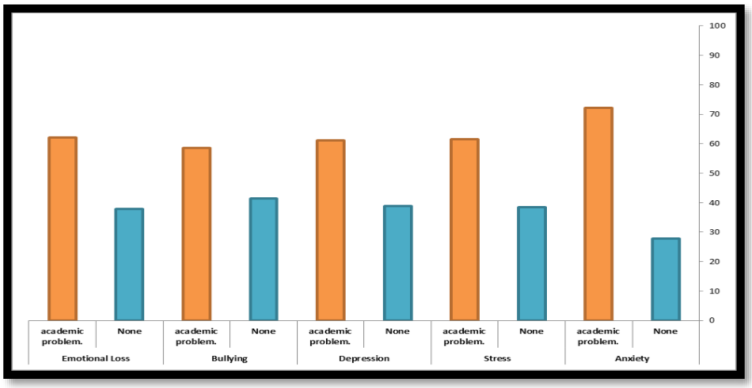

Figure 7. Distribution of the study sample members according to the type of psychological problems and those related to an academic problem

2. Which machine learning models provide the most accurate classification of student difficulties?

The accuracy obtained with Support Vector Machines (SVM) shows its efficiency over the rest of the models in terms of accuracy and reaches the highest value of 88.2 percent, which would allow us to classify students into predefined categories. Logistic Regression also maintained a good

https://doi.org/10.22521/edupij.2025.15.150 Published online by Universitepark Press

performance of 87.7% and therefore these models are good for use, especially in educational facilities.

3. What is the relationship between academic difficulties and psychological difficulties?

There is a positive correlation between academic difficulties and psychological difficulties, and the results show a statistically significant relationship. Table 17 shows these results.

Table 17. Pearson's correlation coefficients between the number of academic problems that students suffer from and the number of psychological problems

The relationship between academic performance and mental health	
Pearson's correlation	0.140
Sig.	0.00

The table shows a statistically significant correlation at the significance level (α≤0.05) between psychological problems and academic problems, as the correlation coefficient value reached 0.140, which is a statistically significant value. Also, Pearson's correlation coefficients between academic achievement (GPA) and the number of psychological problems were extracted.

Table 18. Pearson's correlation coefficients between the Academic achievement and the number of psychological problems

The impact of psychological problems on academic performance	
Pearson's correlation	-0.116
Sig.	0.00

Table 18 shows that there is a statistically significant correlation inverse at the significance level (α≤0.05) between psychological problems and academic performance as the value of the correlation coefficient reached (-0.116), which is a statistically significant value. Also, the linear regression equation was applied to verify the nature of the effect of the cumulative average on psychological problems.

Table 19. Linear regression equation to verify the nature of the effect of the cumulative average on psychological problems

Independent variable	Beta	T	R	R Square	F	Sig.
psychological problems	-0.116	-3.721	0.116	0.013	13.844	0.000

Table 19 shows there is a statistically significant effect at sig. (α≤0.05) for the value of the correlation coefficient (R) was (0.116), which is a statistically significant value. (R-square) was (0.013), which is a statistically significant value that explains the ability of the psychological problems to influence the cumulative average, i.e. the psychological problems (1.3%) of the change in the cumulative average.

4. How can AI-based models be utilized to improve early intervention strategies for supporting university students?

A review of the classification process using GPT-4o mini and machine learning model also proved that AI can be as effective in identifying students who are at risk. These tools can be integrated into early intervention systems to provide targeted support, addressing both academic and psychological dimensions.

5.2. Contributions of the Study

This research contributes to the academic and educational fields by:

- Developing a scalable framework that integrates statistical analysis, AI and machine learning for classifying student difficulties.
- Emphasizing the academic and psychological factors most influential to the student success.
- Allowing for a better understanding of the difficulties encountered by Jordanian university students, which strategies can then be developed based on tailored intervention.

5.3. Limitations and Future Directions

However, this study has some limitations. These limitations may arise from having to depend on self-reports, and the results are generalizable only for Jordanian universities. Hopefully, future research studies could generalize the sample to different educational contexts and investigate other machine learning algorithms to raise classification accuracy.

6. Conclusion

This study emphasizes the need to solve academic and psychological problems with the help of the application of intelligence technologies. Using GPT-4o mini and machine learning models; the research offers a strong foundation for noting down the vulnerable students, and further the performance of the learning process. As a consequence, these findings must be left to future research to continue developing the essential bases for creating a more cordial and inclusive learning environment in higher education. As part of future work, the research will expand into Deep Learning models because we want to boost classification performance and discover advanced patterns regarding students' challenges. We intend to raise the participant number for the study by including students from multiple educational settings to increase the validity of our findings.

7. Suggestion

1. Implementation of Early Warning Systems: Higher education institutions should implement AI-based detection systems that identify punctual academic and psychological issues affecting their students. The use of AI-powered systems enables This can allow for timely intervention and support.

2. Collaboration Between Academic and Psychological Units: The adoption of artificial intelligence systems becomes important for higher education institutions to identify academic and psychological challenges among their student population during the early stages. By utilizing these systems higher learning institutions gain the ability to intervene quickly and offer support.

3. Customizing Interventions Based on Classification Output: Educational institutions should adapt their intervention programs depending on student classification between academic-only psychological-only or combined challenges.

4. Policy Development Based on Predictive Insights: data predictions enable education policymakers to create student support policies with a precise focus on vulnerable student groups through evidence-based approaches.

5. Future Research Directions: Research should integrate behavioral data with real-time data into classification models because this action would boost prediction accuracy and increase usage potential.

Declarations

Author Contributions: [Seif Alazzam] 1: Conducted the research, data collection, and drafting of the manuscript. [Mohammad Aloudat] 2: Supervised the project, provided expert guidance, and critically reviewed the manuscript.
Conflict of Interest: The authors confirm that there are no conflicts of interest associated with this publication.
Funding: The authors received no financial support for this article.
Ethical Approval: The Ethical Committee of the Philadelphia University, Jordan has granted approval for this study on 31- December-2024 (Ref. No. IRB/2/3/2024-2025).
Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Abdul Rahman, H., Lim, Y. J., Alqaralleh, B., Mohd Shukri, S. N., & Ismail, N. (2023). Machine learning-based prediction of mental well-being using health behavior data from university students. *Bioengineering, 10*(5), 1–13.https://doi.org/10.3390/bioengineering10050575.

Ahmed, A., Zhang, Y., Khan, S., Ali, M., & Li, J. (2022). Machine learning models to detect anxiety and depression through social media: A scoping review. *Computer Methods and Programs in Biomedicine Update, 2*, 100066. https://doi.org/10.1016/j.cmpbup.2022.100066.

Al-Ali, R., Alhumaid, K., Khalifa, M., Salloum, S. A., Shishakly, R., & Almaiah, M. A. (2024). Analyzing socio-academic factors and predictive modeling of student performance using machine learning techniques. *Emerging Science Journal, 8*(4), 1304–1319. https://doi.org/10.28991/ESJ-2024-08-04-05.

Azizah, Z., Ohyama, T., Zhao, X., Ohkawa, Y., & Mitsuishi, T. (2024). Predicting at-risk students in the early stage of a blended learning course via machine learning using limited data. *Computers and Education: Artificial Intelligence, 7*, 100261. https://doi.org/10.1016/j.caeai.2024.100261.

Barbayannis, G., Bandari, M., Zheng, X., Baquerizo, H., Pecor, K. W., & Ming, X. (2022). Academic stress and mental well-being in college students: Correlations, affected groups, and COVID-19. *Frontiers in Psychology, 13*, 886344. https://doi.org/10.3389/fpsyg.2022.886344.

Bas, G. (2021). Relation between student mental health and academic achievement revisited: A meta-analysis. In Health and academic achievement - New findings. *IntechOpen*. https://doi.org/10.5772/intechopen.95766.

Ciarrochi, J., Hayes, S. C., Oades, L. G., & Hofmann, S. G. (2022). Toward a unified framework for positive psychology interventions: Evidence-based processes of change in coaching, prevention, and training. Frontiers in Psychology, 12, 1–16. https://doi.org/10.3389/fpsyg.2021.809362.

Chaka, C. (2022). Educational data mining, student academic performance prediction, prediction methods, algorithms and tools: An overview of reviews. *Journal of E-Learning and Knowledge Society, 18*(2), 58–69. https://doi.org/10.20368/1971-8829/1135578.

Gülşen, F. U., & Şahin, E. E. (2022). Basic psychological needs, academic self-efficacy, self-leadership, career adaptability, and life-satisfaction: Data-set from Turkish university students. Data in Brief, 40, 107834. https://doi.org/10.1016/j.dib.2022.107834.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2018). *Multivariate data analysis* (8th ed.). Cengage Learning EMEA.

Hutter, F., Kotthoff, L., & Vanschoren, J. (Eds.). (2019). *Automated machine learning: Methods, systems, challenges.* Springer. https://doi.org/10.1007/978-3-030-05318-5.

Ke, L., Tong, S., Cheng, P., & Peng, K. (2024). Exploring the frontiers of LLMs in psychological applications: A comprehensive review. *arXiv*. https://arxiv.org/abs/2401.01519.

Lange, R. S. (2014). Pascarella, T., & Terenzini, P. (2005). How college affects students: A third decade of research (2nd ed.). San Francisco: Jossey-Bass [Review of the book]. Journal of Student Affairs in Africa, 2(2), 47–50. https://doi.org/10.14426/jsaa.v2i2.70.

Laverghetta, A., Nighojkar, A., Mirzakhalov, J., & Licato, J. (2022). Predicting human psychometric properties using computational language models. In Proceedings of the 2022 Conference on Computational Linguistics and Language Resources (pp. 151–169). Springer. https://doi.org/10.1007/978-3-031-04572-1_12.

Mustafa, M. B., Zaharim, N. M., Mohamed, Z., Yusof, N. M., & Rahim, R. A. (2020). The relationship between psychological well-being and university students' academic achievement. International Journal of Academic Research in Business and Social Sciences, 10(7), 518–525. https://doi.org/10.6007/ijarbss/v10-i7/7454.

Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2022). Predicting students' academic performance using machine learning techniques: A literature review. *International Journal of Business Intelligence and Data Mining, 20*(4), 456–479. https://doi.org/10.1504/IJBIDM.2022.123214.

Naveed, H., Farooq, M. U., Rehman, M. H., Khan, A., & Ahmad, M. (2023). A comprehensive overview of large language models. *arXiv*. https://arxiv.org/abs/2307.06435.

Raiaan, M. A. K., Mohiuddin, M. A., Rahman, M. M., Afser, M. A., & Karim, R. (2024). A review on large language models: Architectures, applications, taxonomies, open issues and challenges. IEEE Access, 12, 26839–26874. https://doi.org/10.1109/ACCESS.2024.3365742.

Rajkishan, S. S., Meitei, A. J., & Singh, A. (2024). Role of AI/ML in the study of mental health problems of the students: A bibliometric study. International Journal of System Assurance Engineering and Management, 15(5), 1615–1637. https://doi.org/10.1007/s13198-023-02052-6.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. Psychological Bulletin, 138(2), 353–387. https://doi.org/10.1037/a0026838.

Safa, R., Edalatpanah, S. A., & Sorourkhah, A. (2023). Predicting mental health using social media: A roadmap for future development. In Deep learning for personalized healthcare decision support (pp. 285–303). Elsevier. https://doi.org/10.1016/B978-0-443-19413-9.00014-X.

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. Psychological Bulletin, 143(6), 565–600. https://doi.org/10.1037/bul0000098.

Tabassum, R., Kashif, M. F., & Jamil, M. (2024). Impact of mental health and well-being on academic performance of university students. Journal of Social Research and Development, 5(3), 37–48. https://doi.org/10.53664/jsrd/05-03-2024-04-37-48.

Wong, B. T. M. (2017). Learning analytics in higher education: An analysis of case studies. Asian Association of Open Universities Journal, 12(1), 21–40. https://doi.org/10.1108/AAOUJ-01-2017-0009.

About the Contributor(s)

Seif Hashem Al-Azzam is a final-year graduate student pursuing a Master's degree in Software Engineering at Philadelphia University in Jordan. My research interests include artificial intelligence, educational data mining, and machine learning applications in higher education.
Email: 202211320@philadelphia.edu.jo
ORCID: https://orcid.org/0009-0000-0673-6201

Mohammad Al-Oudat is an Assistant Professor in the Department of Computer Science at Philadelphia University in Jordan. His academic work focuses on artificial intelligence, data science, and intelligent systems. He has published several research papers in reputable international journals.
Email: maloudat@philadelphia.edu.jo
ORCID: https://orcid.org/0000-0002-0553-7961