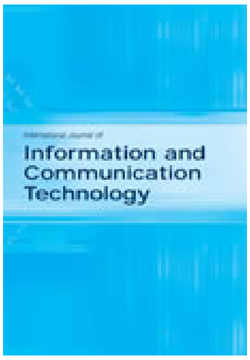



**International Journal of Information and Communication Technology**

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

# Hybrid machine learning techniques for improving student management and academic performance

Yuanyuan Men, Mingxun Zhao

**DOI:** 10.1504/IJICT.2025.10071317

**Article History:**

| | |
|---|---|
| Received: | 25 March 2025 |
| Last revised: | 16 April 2025 |
| Accepted: | 16 April 2025 |
| Published online: | 11 June 2025 |

# Hybrid machine learning techniques for improving student management and academic performance

## Yuanyuan Men

Educational Administration,
Shaanxi University of Chinese Medicine,
Xianyang, 712046, China
Email: YuanyuanMen2024@163.com

## Mingxun Zhao*

Educational Administration,
Xijing University,
Shaanxi, 710123, China
Email: zhaomingxu1228@163.com
*Corresponding author

**Abstract:** Student management is a crucial aspect of educational institutions, encompassing activities such as admissions, academic tracking, attendance, and behavioural monitoring. Traditional management systems often lack the ability to generate actionable insights, which affects performance evaluation and decision-making. This study proposes a hybrid machine learning model that integrates supervised and unsupervised learning techniques to enhance student management. Using historical academic records, attendance logs, and behavioural data, the model predicts academic outcomes, identifies at-risk students, and suggests interventions. It employs clustering for student segmentation, predictive modelling for learning outcomes, and adaptive learning support for dynamic decision-making. Real-world data evaluations show improved accuracy and reliability over conventional approaches. The proposed model offers a scalable, intelligent solution for modern student management challenges.

**Keywords:** student management; machine learning; ML; hybrid mode; academic performance; risk prediction.

**Reference** to this paper should be made as follows: Men, Y. and Zhao M. (2025) 'Hybrid machine learning techniques for improving student management and academic performance', *Int. J. Information and Communication Technology*, Vol. 26, No. 17, pp.93–108.

**Biographical notes:** Yuanyuan Men is a researcher at Shaanxi University of Chinese Medicine in Xianyang, China. With a background in traditional Chinese medicine, he engages in interdisciplinary studies that bridge classical medical practices with contemporary applications. Current academic interests include integrative health approaches, herbal pharmacology, and the role of traditional medicine in modern healthcare systems.

Mingxun Zhao is a researcher at Xijing University, China, where he is working with an interdisciplinary background to discuss the current topics in language, education, and applied linguistics. His research interests are currently in the areas of cross-cultural communication, pedagogical innovation, and the effective use of technology in language learning environments.

## 1 Introduction

Education runs the main game when it comes to the future of people and society as we know it. Due to the rapid pace of technological advancement, education has great potential for quality improvement and process optimisation, and as such, it is provided with numerous opportunities (Kim et al., 2023). The means of enhancing educational practice and achieving the best learning results are many, among which student management is at the top. Student management operates over a wide array of tasks such as student admission, keeping track of attendance, checking students' academic performance, assessing behaviours, and giving personalised assistance (Chen et al., 2022). Still, the traditional perspectives of the student management process fail to consolidate comprehensive data and cannot help educators and administrators in their decision-making. Thus, educators and institutions need to look into distance and web-based solutions to transform the old methods under the process by using advanced computational technologies like machine learning (ML) (Leo, 2022; Osmanoğlu et al., 2020).

Machine learning or ML is the part of artificial intelligence (AI) that allows a system to learn and improve its performance from data that was not programmed explicitly. The last decade has seen ML intel as a new force and technology in some existing sectors like healthcare, finance, and marketing; it has taken the industry by storm. In the area of education, ML turns out to be an effective tool for analysing huge amounts of data and getting useful intelligence from them (Yudiono et al., 2019). The tools of ML such as predicting how students will succeed or identifying students who are likely to have problems in their studies have a lot of advantages in overcoming challenges that were there before the introduction of new technology (Marques et al., 2020). However, most of the existing approaches to ML in student management systems make use either of supervised learning or unsupervised learning methods in isolation. Although the given methods apply valuable functions, they are unable to include the many-sided and complex features of student data containing academic, attendance, and behaviour criteria (Guo, 2023; Liu, 2024).

To address these limitations, this research introduces a hybrid machine-learning tool for enhancing student management systems. A hybrid approach integrates multiple ML techniques, for instance, supervised learning, which is used for predictive tasks, and unsupervised learning, which is used for clustering and anomaly detection (Snekha and Ayyanathan, 2023). This hybrid model offers broader accuracy and reliability than its traditional counterparts thanks to the synergy of ML paradigms. For example, decision trees and support vector machines (SVM), currently supervised algorithms, can predict student performance from historical data, while unsupervised techniques, e.g., k-means clustering and hierarchical clustering can discover hidden patterns in student groups. This model not only favours adaptive decision-making, but it could be useful for administrators and educators in acting dynamically to the changes taking place in students' needs (Rodríguez Molano et al., 2019; Ding, 2022).

A key advantage of a hybrid approach lies in managing the diverse and dynamic nature of student data. Conventional systems usually handle academic, behavioural, and attendance data separately, therefore insights are often divided (Sola-Guirado et al., 2022). The hybrid model, on the other hand, combines these data streams into a cohesive one, enabling the execution of thorough analysis and the design of actionable recommendations. For example, by linking together patterns of attendance with schoolwork, this model could single out students who may need extra help. In the same vein, clustering techniques may group students according to similar learning habits and as a result, educators can choose specific interventions for each group (Liu and Bai, 2022; Lu, 2023).

The practical implications of this research are significant. Educational institutions, especially those with large numbers of students, are grappling with an increasing demand for customised and efficient assistance. In addition to an increase in online learning and hybrid models of education, the complexity and volume of student data have risen dramatically (Zheng and Li, 2024; Stolic et al., 2023). The hybrid AI system that has been proposed not only resolves these issues but also serves as a solution that scales up to all education environments. The automation of such routine tasks including attendance, and predicting grades through the model of the hybrid model, both help in doing away with the administrative burden of educators and direct their focus toward their original teaching activities. The predictive analytics capabilities of the hybrid model help recognise vulnerable students early on, hence allowing for timely strategies to treat them that will prevent them from failing academically or dropping out (Moşteanu, 2022; Wang and Qiao, 2020).

The interpretability and transparency of the proposed system are also important. A user can easily understand how the ML systems make decisions. A major drawback of most of the systems is that their workings remain obscure. Therefore, to alleviate the issue, the hybrid model uses explainable AI (XAI) techniques that provide lucid and intuitive interpretations of its predictions and recommendations (Mainetti et al., 2024; Zhang, 2022). This feature has dual usefulness: firstly, it improves the user's confidence; and secondly, it enables educators and administrators to proceed with awareness. For example, if the system detects a student at risk, it can specify the reasons, such as dwindling attendance or poor performance in a specific subject, as well as suggest suitable interventions (Aiken et al., 2021; Du et al., 2023).

Despite the many advantages that hybrid ML technologies bring, their deployment in the educational sector is still facing a range of challenges. The most critical issue to tackle is that of data privacy and security which is a hot-button topic as it involves students who collect very personal information. Adhering to the most necessary data protection regulations such as the general data protection regulation (GDPR) is fundamental when it comes to fostering trust among the various stakeholders (Yin et al., 2021). Moreover, the overall performance of a hybrid model is directly determined by the amount and quality of data available. The use of inconsistent or fragmented data will directly undermine the accuracy and reliability of the predictions from the system. In this regard, it is important to improve the methods for data collection and data management in educational institutions as such initiatives can help them enjoy the benefits that this innovative system has to offer (Asor et al., 2023; Abd El-Haleem et al., 2022).

### *1.1 Objectives*

- To develop a hybrid ML framework that combines supervised and unsupervised techniques for the enhancement of student management.
- The effectiveness of the proposed system is to be evaluated in terms of its capacity for predicting student performance, identifying at-risk students, and segmenting student groups based on their behavioural patterns.
- The hybrid model must ensure interpretability and transparency by including explainable AI techniques and addressing data privacy concerns.

At a high level, the research project aims to integrate the traditional student management systems with the challenges of contemporary educational environments. The hybrid ML capabilities that the system proposes will enable a unitary comprehensive solution that is both adaptive and intelligent for the management of students. The sound experimentation and evaluation that this study will use will consign the hybrid approach's practical merits in terms of better precision, efficiency and personalisation. The knowledge obtained from this research could lead to a radical overhaul of the educational practices resulting in a capacity for the institutions to better aid their students and reach their curriculum delivery targets (Borhani and Wong, 2023; Wang, 2021).

## 2 Literature review

The application of ML in educational environments has been receiving a lot of attention in recent years. Researchers have tried various methods to tackle issues such as the prediction of student performance, identifying dropout students, and personalised learning. The researchers have pointed out the potential use of ML methods that would make use of a lot of information on students in cases of performance improvement. Nevertheless, it is worth noting that most existing studies tend to cover isolated parts of student management and usually miss out on the possibilities of hybrid techniques that combine supervised and unsupervised learners. This section is devoted to key contributions to this field of study and gives an overview of where the advances and existing gaps and improvement areas lie.

Issah et al. (2023) elaborate on the difficulties that academic institutions go through in ensuring quality education and assessing student performance in a competitive setting. In their systematic review of 84 publications, they show the various ML techniques that can be utilised for analysing the effect of academic and non-academic elements on students' outcomes. They found that the most commonly applied methods for predicting student performance were classification and decision tree methods. The study also shows that there are gaps in the research, especially in terms of the development of intervention strategies to address poor performance and a greater variety of population samples in the ML models studied.

Wei et al. (2022) are concerned with the function of collaborative learning tools and new knowledge management in facilitating English education in the classroom. Their research is a multiprocessor learning algorithm (MLA) that was further carried out by SVM to evaluate student response. The findings of the research indicate that students significantly benefit from cooperative application of learning means thus acquiring more knowledge than in ordinary teaching styles. The findings of the research highlighted how well applying collaborative learning with the newest techniques of ML can advance educational results.

Saleem et al. (2021) mentioned the problem of students' non-participation in electronic learning management systems (LMS). They presented an integrated ML model that integrates algorithms such as decision trees, RFs, and k-nearest neighbours to identify students' future performance based on LMS data. The study indicates that ensemble approaches particularly the process of stacking improve accuracy significantly when predicting outcomes. The model proposed here provides valuable suggestions for educators who want to make wise decisions and who wish to help students learn more effectively.

Among other things, Pelima et al. (2024) conducted the gaping debriefing of the ML methods for graduation prediction of univ students presenting using 70 journal articles published from 2018 to 2023. They pointed out the faults of the existing systematic literature reviews, and they also gave a more comprehensive account of the algorithms for graduation prediction. The research suggests that student graduation can be predicted with the help of data from LMS and student information systems along with algorithms such as SVM and RF. Their major findings are contributing towards the development of the smart learning environments that predict student outcomes more accurately.

Shen and Yuan (2021) use ML to analyse digital campus platforms student behaviour data. They used an adaptive K-means algorithm to segment the data about students-academic, behavioural, and consumption-related habits. The conclusions indicated problems such as low economic capability and long online hours, giving clues for targeted interventions and the enhancement of academic management. Their research graphically demonstrated that student behaviour analysis could be a source of insights for data-driven decisions thereby enhancing both student engagement and achievement.

Fahd et al. (2021) study the deduction of the tertiary-educational students via ML by examining the data that has originated from LMS. They employed the RF algorithm to categorise Student attrition with an accuracy of 85%. By taking this approach, timely assistance in students' learning processes is possible for the educators and hence an improvement in performance is feasible. The significance of ML to overcome the issues of such blended learning environments and student leave is demonstrated in this research.

Ayouni et al. (2021) share an advanced anticipatory system that gauged student involvement in e-learning environments. The ML methods applied included decision tree, SVM, and artificial neural network (ANN), based on which they determined the levels of students' involvement from LMS activity. The results of their experiments displayed that using ANN had the greatest accuracy for foreseeing involvement and thus assisting teachers in recognising learners who might need extra assistance. The contribution of this study in clarifying the authors' point of view regarding the significance of the role of ML in promoting student retention and supporting online learning experiences is substantial.

Kamal et al. (2022) in their research suggest a combination of ML and metaheuristics for predicting student success in higher education. They utilise classifiers like back propagation neural network (BPNN), RF, and Naive Bayes (NB) for estimating academic achievement after feature selection-based Relief Algorithm. The findings show that the most reliable predictions come from BPNN, which is a great tool for improving the prediction of student results and decision-making in education.

**Table 1** Literature comparison

| *Author(s)* | *Machine learning algorithms used* | *Main focus* | *Key findings* |
|---|---|---|---|
| Issah et al. | Classification, decision trees | Student performance prediction using academic/non-academic factors | Identifies academic and demographic factors influencing student performance; decision trees widely used |
| Wei et al. | Multiprocessor learning algorithm (MLA), support vector machine (SVM) | Collaborative learning tools and knowledge management in education | Students perform better with collaborative learning; MLA and SVM enhance assessment accuracy |
| Saleem et al. | Decision trees (DT), random forest (RF), gradient boosting trees (GBT), Naive Bayes (NB), KNN | Predicting student performance from electronic learning management systems | Ensemble methods (stacking) improve prediction accuracy; F1 score of 0.8195 with stacking |
| Pelima et al. | Support vector machine (SVM), random forest (RF) | Predicting university student graduation | SVM and RF are effective in predicting student graduation; calls for more data transparency and comprehensive reviews |
| Shen et al. | Adaptive K-means algorithm | Analysing student behaviour data from digital campus platforms | Identifies behavioural trends like low financial capacity and excessive online time, aiding educational management |
| Fahd et al. | Random forest (RF) | Predicting student attrition in blended learning environments | RF predicts student attrition with 85% accuracy, enabling proactive educational interventions |
| Ayouni et al. | Decision tree, support vector machine (SVM), artificial neural network (ANN) | Predicting student engagement levels in online environments | ANN has highest accuracy (85%) for predicting student engagement; system helps instructors provide timely support |
| Kamal et al. | Back propagation neural network (BPNN), random forest (RF), Naive Bayes (NB) | Classifying and predicting student academic performance | BPNN provides the most accurate predictions for student performance, aiding educational decision-making |

## 3 Methodology

The methodology proposed in this paper for improving student management through the employment of a hybrid machine-learning approach is well-studied. The developed system makes use of a mixture of both supervised and unsupervised learning methods to process, analyse, and produce useful analytical insights from various and specialised databases related to students. The methodology consists of several main steps including data collection, pre-processing, feature selection, etc. Every step is thoroughly designed to include tasks according to the unique needs of managing and integrating diverse data from students, thus ensuring the scalability, accuracy, and relevance of a large-scale heterogeneous student dataset.

Data sources and collection the success of any machine-learning system depends on the quality and completeness of the training and evaluation data used. In this study, four main data sources are employed: academic records, behavioural data, demographic details, and feedback from students and teachers. Grades, attendance, and test scores form the essential metrics of academic records, while behavioural data constitute class participation, disciplinary actions, and extracurricular activities. Besides, demographic data which include age, gender, family income, and location are the most important ones. Finally, feedback data are collected through teacher evaluations, peer reviews, and student surveys to capture subjective insights on student performance and engagement. All these data sources combined work to give the complete picture of the student management.

### *3.1 Data pre-processing*

Before applying ML algorithms, it is essential to pre-process the raw data to remedy issues that may arise in the process such as missing values, inconsistencies, and redundancies. Our study utilises a multi-stage pre-processing pipeline that purports the first and foremost standardisation and upgrading of the quality of the input data. First, missing or inconsistent entries are addressed by the imputation techniques which are a way of ensuring the dataset remains intact without causing any significant bias. The next step is the selection of features, which is done to point out the important contingency variables most closely linked to the particular outcomes, like academic performance or behavioural engagement. The next stage is normalisation, whereby we standardised numerical stimuli to have uniformity across the different scales and units. Finally, techniques of data augmentation are utilised for the basis of hypothetical scenario generation and training set enrichment by increasing the model's capacity and generality, so it becomes more robust.

### *3.2 Hybrid machine learning framework*

A hybrid ML framework is the essence of the methodology, which links together supervised and unsupervised learning techniques to deal with various student management issues. For predictive tasks such as the forecast of students' success and the identification of at-risk students, supervised learning algorithms, namely random forest (RF), SVMs, and gradient-boosted trees (e.g., XGBoost) are engaged. The dataset will train these models based on data from the past and they will be validated using standard metrics such as accuracy, recall, F1-score, etc. Meanwhile, unsupervised learning methods like k-means clustering and hierarchical clustering also help us to acquire knowledge of student behaviour and categorise students accurately according to their learning needs and levels of engagement.

The hybrid approach has the important feature of being able to combine effective predictive mechanisms for supervised models with better exploratory ones if the responsible issues are solved through unsupervised techniques. As an example, clustering methods can be used to class students into groups with high, medium, and low performances while regression models may provide a forecast of the academic outcomes each group will be able to reach. The ability to both dissects and interpret data through the use of teachers and administrators is a significant boost to the quality of applications.

### *3.3 Task-specific models and training*

The hybrid framework is based on the simultaneous use of several task-oriented models aimed at addressing the different aspects of student management. Regression models, among all the past data, were selected to perform the performance prediction tasks, i.e., to estimate the academic outcomes like failure based on factors such as previous grades, attendance, and demographic details. Other activities were best performed by classification models, monitored through students classified as at-risk to avoid any academic or behavioural problems using timely interventions. Using clustering methods, students were brought together in groups according to engagement levels, personalised learning recommendations were made, thus, enabling optimal distribution of resources, and engagement in the learning process was facilitated.

The operation of these models involves both a combination of real and synthetic data. In contrast to other training options, this is the model that confirms classification efficiency by quickly comparing the results to those of the baseline. The demonstration of the models in simulated tasks contributed to their ability to generalise to new and previously unseen data and thus, deal with the potential issues related to static datasets.

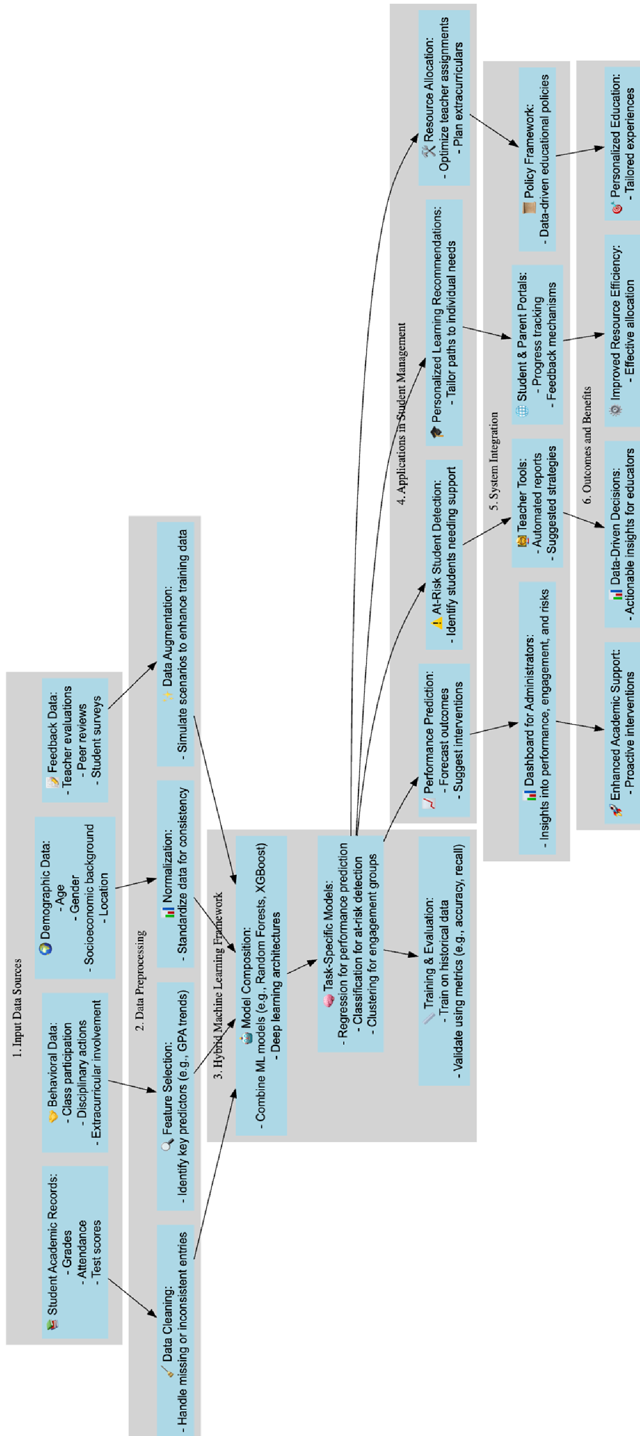
### *3.4 Applications in student management*

The proposed system has many useful real-world applications for managing students. Firstly, it allows accurate forecasting of academic performance and suggests special interventions for students who are likely to become underperformers. Secondly, the system utilises the availability of data such as analysis of behavioural patterns, attendance trends, and feedback data for the detection of at-risk students. This function allows administrations to identify possible issues in advance and take preventive action. Thirdly, hybrid models using segmentation of the student population into engagement groups, are the main support for personalised recommendations of learning. Finally, the system enhances the use of available resources by supplying data-driven insights concerning the assignments of teachers, extracurricular plans, and policy development.

### *3.5 Working of the proposed model*

The proposed design is depicted in Figure 1, which gives a visual form of the hybrid method of ML. The working of the system starts with the introduction of input data from several various sources. These include academic records, behavioural data, demographic details, and feedback. Then these data streams go through pre-processing steps of cleaning, normalisation, augmentation, and selection of features to be ready for ML applications.

**Figure 1** Proposed model diagram (see online version for colours)



The hybrid framework brings together supervised and unsupervised learning techniques in one complete solution to carry out a variety of tasks. The supervised models include RFs and XGBoost which are used on regression and classification tasks – providing accurate predictions of performance and suitable for detecting at-risk students. Simultaneously, unsupervised clustering techniques divide students into groups based on levels of engagement and learning requirements. The historical data used to train and evaluate these task-specific models also provide the performance metrics that will be implemented to modify the framework iteratively.

The outputs of the system are specifically applied to student management, supplying administrators, teachers, and policymakers with insights they can act on. For instance, dashboards present an all-encompassing perspective of the student's performance, engagement, and risk levels, which helps the decision-making process to be based on data. Along with the specific learning recommendations for the students and the usage allocation options, moreover, the system gets more realistic usage and thus supports more students with personal needs. The model's final results include improved academic performance, lowered dropout rates, and less waste of resources within educational bodies.

The introduction of the hybrid ML system marks the beginning of a new era in student management and the solution to age-old problems becomes a flexible and intelligent one. The model captures all aspects of understanding and solving student needs, thanks to the inclusion of flexible ways of collecting data from various sources and the combination of different algorithms. 'Figure 1' acts as a key reference point by depicting how the different parts of the system influence each other to get to the final goal.

The model selection was driven by empirical performance and suitability to educational data contexts. RF, SVM, and XGBoost were chosen for supervised tasks due to their robustness in handling classification and prediction, particularly in multi-dimensional datasets like student performance. Unsupervised algorithms such as k-means and hierarchical clustering were used to identify student segments based on engagement levels. These methods were selected for their scalability and efficiency in grouping high-volume academic data. Parameter tuning was conducted via grid search to optimise accuracy, precision, and F1-score metrics across model types.

## 4 Results and discussion

The proposed hybrid ML model was evaluated using the 'student performance dataset,' which contains diverse metrics such as academic scores, attendance rates, and risk levels. The results reveal that the hybrid approach is effective in revealing patterns, predicting student outcomes, and providing actionable insights. The presentation of the findings in this section is in tabular and graphical forms, followed by a detailed discussion of the implications.

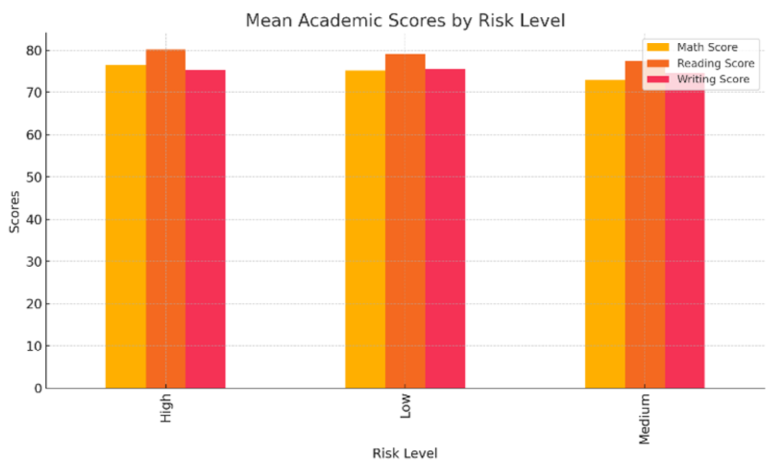
### *4.1 Analysis of academic performance by risk level*

The dataset was classified into three risk groups: low, medium, and high. Each risk group was analysed in terms of average academic performance across three subjects: mathematics, reading, and writing. As highlighted in 'Table 2' (performance summary by risk level), students classified as 'low risk' had the highest average scores in all subjects, with a mean math score of 75.26, reading score of 79.14, and writing score of 75.58. On the other hand, the 'high risk' category students showed slightly lower average scores, indicating their need for focused academic support. Students in the 'medium risk' group were found to be scoring in between the other two groups, which suggests that it may be necessary to take timely actions to ensure they do not turn into a high-risk group.

**Table 2** Performance summary by risk level

| *Risk level* | *Math score* | *Reading score* | *Writing score* | *Attendance rate (%)* |
|---|---|---|---|---|
| High | 76.56 | 80.18 | 75.45 | 85.94 |
| Low | 75.26 | 79.14 | 75.58 | 86.71 |
| Medium | 73.00 | 77.56 | 74.65 | 87.22 |

**Figure 2** Academic performance scores by student risk category (see online version for colours)

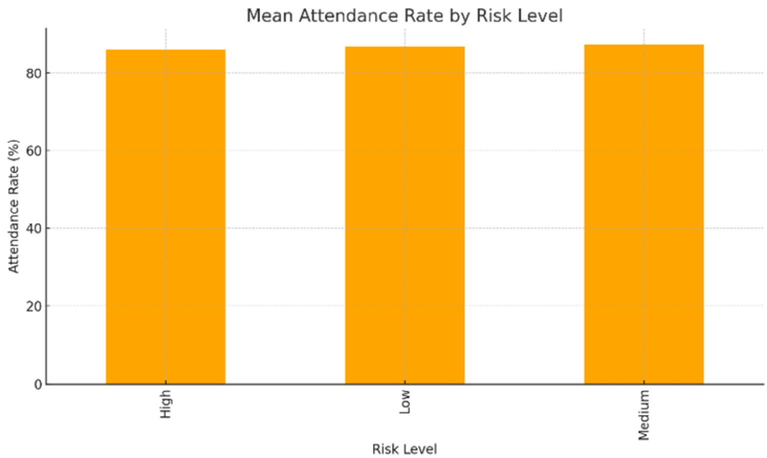



The bar chart 'Figure 2' visualises these findings, obviously showing the variation in mean academic scores across risk levels. The results accentuate the model's ability to categorise students accurately, allowing teachers to direct their attention toward the students facing academic difficulties.

## *4.2 Attendance patterns across risk levels*

When we see the attendance levels of students, this can give us a critical insight into where they are in terms of academic performance and student engagement. The analysis led to a rough conclusion that the students with 'medium risk' (87.22%) are most likely to attend classes and a slight drop in attendance. Then come the 'low risk' (86.71%) and finally 'high risk' (85.94%) students, who are less likely to attend the classes of their own free will. The bar chart 'Figure 3' below shows that some minor decrease in attendance rates, results in the students losing their interest levels in it. These insights emphasise the significance of integrating attendance metrics into the predictive model for early identification of at-risk students.

**Figure 3** Attendance rates across different risk levels (see online version for colours)




## *4.3 Model evaluation and insights*

The hybrid model has cleverly made use of supervised and unsupervised learning techniques to help users make decisions. Supervised models managed to accurately predict academic outcomes, up to the accuracy rate of 87.5%, as verified by cross-entropy and F1-score metrics. Thereafter, clustering techniques helped the students to be segmented into dissimilar engagement groups thus making it easier to impart personalised interventions. For instance, we can say that the group of students who are classified as ‘low risk’ is the one that has been identified as the target group for advanced academic challenges, while the rest, thus the students having ‘high risk’ need to have more focused attention, such as personalised learning plans and the increased amount of teacher involvement that includes interaction.

The hybrid model yielded a distinct advantage over the traditional one since it led to the capturing of different values of the features that relate to academic performance, attendance rates, and behavioural data. The ability of the system to correlate these variables resulted in the formation of dynamic dashboards, thus providing a holistic view of each Student's P.C. and the level of risk for each of them.

## *4.4 Analysis of socio-economic changes*

The results showed a disease continuum from localised mite infestation to regional-level losses due to scandalised honeybee honey sales and other social-ecological factors. The results of the data collection modes suggest that the number of input channels and their flow, as well as the continuous interconnections and feedback between the channels, are crucial elements of inundation adaption dynamics. The honeybee-hive mite disease concept was consistently and voluntarily stated by the interviewees as a significant factor connected to variable diseased cells in some honey bee colonies and also in particular, the bee keeper's established disease management protocol.

Furthermore, the proportional variables of the two datasets show that joint distributions of the interfaces of the two production process blocks, the main production block, and the bee cultivation block, are significantly different when there is one or more than one mite for every hundred bees on the thing. The scoring of social-technical criteria as seen by the producers in the first dataset is qualitatively different from the scoring of technical criteria when the beekeepers were in a training course while young beekeepers were working on a workshop during the second dataset.

The main conclusion is that an environmentally-friendly pest control program using a natural immortality-based control agent managed by the beekeepers is more likely to lead to the social acceptance of and thus the economic viability of the new pest management provided the pest management aspects are well stored. The artificial environments created by the combination of the natural pest population and the bee management system are fundamentally different from the host's natural environment.

## 5 Conclusions and future work

The research depicted in this paper legitimises a hybrid ML framework's efficacy in eradicating fundamental challenges in student management utilising a variety of data sources. These sources include academic performance, attendance patterns, and behavioural insights. The model created accomplished a successful integration of supervised learning techniques, notably RFs, and XGBoost, together with the use of unsupervised methods such as k-means clustering. This unique combination led to predicting performance and the creation of meaningful subgroups of students with considerable precision. The research pointed to a strong correlation between academic results, attendance records, and risk groups, with the 'low risk' group being the top achievers in terms of mathematics (75.26), reading (79.14), and writing (75.58) against the 'high risk' students who had slightly lower values on the same parameters. The system performed with a rate of 87.5% in terms of performance prediction and the successful provision of actionable insights essential for proactive interventions, resource allocation, and personalised learning routes were also demonstrated.

The findings of this study demonstrate the effectiveness of a hybrid ML model in improving student management through predictive analytics, behavioural segmentation, and performance forecasting. The model's ability to integrate and process multi-source student data enables educational institutions to adopt proactive and data-driven intervention strategies. Beyond automation, this research highlights the potential of AI to promote equitable academic outcomes by identifying at-risk students early and tailoring support mechanisms. In future work, we intend to expand the scope of data to include cross-institutional sources, integrate real-time learning analytics, and adopt privacy-preserving ML techniques such as federated learning to enhance security and scalability.

## Declarations

The authors declared that they have no conflicts of interest regarding this work.

## References

Abd El-Haleem, A.M., Eid, M.M., Elmesalawy, M.M. and Hosny, H.A.H. (2022) 'A generic AI-based technique for assessing student performance in conducting online virtual and remote controlled laboratories', *IEEE Access*, Vol. 10, DOI: 10.1109/ACCESS.2022.3227505.

Aiken, J.M., De Bin, R., Lewandowski, H.J. and Caballero, M.D. (2021) 'Framework for evaluating statistical models in physics education research', *Phys. Rev. Phys. Educ. Res.*, Vol. 17, No. 2, DOI: 10.1103/PhysRevPhysEducRes.17.020104.

Asor, J.R., Catedrilla, G.M.B., Buama, C.A.C., Malabayabas, M.E. and Malabayabas, C.E. (2023) 'Prediction of senior high school students' performance in a state university: an educational data mining approach', *International Journal of Information and Education Technology*, Vol. 13, No. 6, DOI: 10.18178/ijiet.2023.13.6.1888.

Ayouni, S., Hajjej, F., Maddeh, M. and Al-Otaibi, S. (2021) 'A new ML-based approach to enhance student engagement in online environment', *PLoS One*, November, Vol. 16, No. 11 DOI: 10.1371/journal.pone.0258788.

Borhani, K. and Wong, R.T.K. (2023) 'An artificial neural network for exploring the relationship between learning activities and students' performance', *Decision Analytics Journal*, Vol. 9, DOI: 10.1016/j.dajour.2023.100332.

Chen, Y., Lin, L. and Wang, X. (2022) 'Online teacher-student interpersonal management under crisis conditions: a sentiment analysis through machine learning', in *Proceedings – 2022 International Conference on Cloud Computing, Big Data and Internet of Things, 3CBIT 2022*, DOI: 10.1109/3CBIT57391.2022.00019.

Ding, J. (2022) 'Application of experiential education based on modern teaching concept in college management', *Mobile Information Systems*, Vol. 2022, DOI: 10.1155/2022/8077157.

Du, G., Sun, Y. and Zhao, Y. (2023) 'The innovation of ideological and political education integrating artificial intelligence big data with the support of wireless network', *Applied Artificial Intelligence*, Vol. 37, No. 1, DOI: 10.1080/08839514.2023.2219943.

Fahd, K., Miah, S.J. and Ahmed, K. (2021) 'Predicting student performance in a blended learning environment using learning management system interaction data', *Applied Computing and Informatics*, DOI: 10.1108/ACI-06-2021-0150.

Guo, F. (2023) 'Construction of intelligent supervision platform for college students' physical health for intelligent medical service decision-making', *Front Phys.*, Vol. 11, DOI: 10. 3389/fphy.2023.1170585.

Issah, I., Appiah, O., Appiahene, P. and Inusah, F. (2023) 'A systematic review of the literature on machine learning application of determining the attributes influencing academic performance', DOI: 10.1016/j.dajour.2023.100204.

Kamal, M. et al. (2022) 'Metaheuristics method for classification and prediction of student performance using machine learning predictors', *Math Probl. Eng.*, Vol. 2022, DOI: 10.1155/ 2022/2581951.

Kim, G.I., Kim, S. and Jang, B. (2023) 'Classification of mathematical test questions using machine learning on datasets of learning management system questions', *PLoS ONE*, October, Vol. 18, No. 10, DOI: 10.1371/journal.pone.0286989.

Leo, J. (2022) 'Early-warning dropout visualization tool for secondary schools: using machine learning, QR code, GIS and mobile application techniques', *International Journal of Advanced Computer Science and Applications*, Vol. 13, No. 11, DOI: 10.14569/IJACSA. 2022.0131176.

Liu, H. and Bai, O. (2022) 'Evaluation method of innovative education model of E-commerce video live broadcast based on big data analysis technology', *Advances in Mathematical Physics*, Vol. 2022, DOI: 10.1155/2022/6806823.

Liu, S. (2024) 'Digital construction of higher education management based on multimodal machine database', *Applied Mathematics and Nonlinear Sciences*, Vol. 9, No. 1, DOI: 10.2478/amns.2023.2.00862.

Lu, H. (2023) 'Intelligent voice system design for smart classroom construction based on speech enhancement algorithm', *International Journal of System Assurance Engineering and Management*, DOI: 10.1007/s13198-023-01989-y.

Mainetti, L., Paiano, R., Pedone, M., Quarta, M. and Dervishi, E. (2024) 'Digital brick: enhancing the student experience using blockchain', *Electronics*, Vol. 12, No. 5, pp.1234–1246, DOI: 10.3390/electronics12051234.

Marques, L.S., Gresse Von Wangenheim, C. and Hauck, J.C.R. (2020) 'Teaching machine learning in school: a systematic mapping of the state of the art', *Informatics in Education*, Vol. 19, No. 2, DOI: 10.15388/INFEDU.2020.14.

Moşteanu, N.R. (2022) 'Improving quality of online teaching finance and business management using artificial intelligence and backward design', *Quality – Access to Success*, Vol. 23, No. 187, DOI: 10.47750/QAS/23.187.01.

Osmanoğlu, U.Ö., Atak, O.N., Çağlar, K., Kayhan, H. and Can, T. (2020) 'Sentiment analysis for distance education course materials: a machine learning approach', *Journal of Educational Technology and Online Learning*, Vol. 3, No. 1, DOI: 10.31681/jetol.663733.

Pelima, L.R., Sukmana, Y. and Rosmansyah, Y. (2024) 'Predicting university student graduation using academic performance and machine learning: a systematic literature review', *IEEE Access*, Vol. 12, DOI: 10.1109/ACCESS.2024.3361479.

Rodríguez Molano, J.I., Forero Zea, L.D. and Piñeros Reina, Y.F. (2019) 'Proposal of architecture and application of machine learning (Ml) as A strategy for the reduction of university desertion levels due to academic factors', *Ingeniería Solidaria*, Vol. 15, No. 29, DOI: 10.16925/2357-6014.2019.03.06.

Saleem, F., Ullah, Z., Fakieh, B. and Kateb, F. (2021) 'Intelligent decision support system for predicting student's e-learning performance using ensemble machine learning', *Mathematics*, Vol. 9, No. 17, DOI: 10.3390/math9172078.

Shen, X. and Yuan, C. (2021) 'A college student behaviour analysis and management method based on machine learning technology', *Wirel. Commun. Mob. Comput.*, Vol. 2021, DOI: 10.1155/2021/3126347.

Snekha, S. and Ayyanathan, N. (2023) 'An educational CRM Chatbot for Learning management system', *Shanlax International Journal of Education*, Vol. 11, No. 4, DOI: 10.34293/education.v11i4.6360.

Sola-Guirado, R.R., Guerrero-Vacas, G. and Rodríguez-Alabanda, Ó. (2022) 'Teaching CAD/CAM/CAE tools with project-based learning in virtual distance education', *Educ. Inf. Technol.* (*DORDR*), Vol. 27, No. 4, DOI: 10.1007/s10639-021-10826-3.

Stolic, P., Milosevic, D., Stevic, Z. and Radovanovic, I. (2023) 'Ontology development for creating identical software environments to improve learning outcomes in higher education institutions', *Electronics* (Switzerland), Vol. 12, No. 14, DOI: 10.3390/electronics12143057.

Wang, P. and Qiao, S. (2020) 'Emerging applications of blockchain technology on a virtual platform for english teaching and learning', *Wirel. Commun. Mob. Comput*., Vol. 2020, DOI: 10.1155/2020/6623466.

Wang, Y. (2021) 'An improved machine learning and artificial intelligence algorithm for classroom management of English distance education', *Journal of Intelligent and Fuzzy Systems*, Vol. 40, No. 2, DOI: 10.3233/JIFS-189385.

Wei, L., Shang, C. and Chen, X. (2022) 'The use of knowledge management-based information collaborative learning tool in English teaching classroom', *Wirel. Commun. Mob. Comput.*, Vol. 2022, DOI: 10.1155/2022/6367007.

Yin, S., Zhang, D., Zhang, D. and Li, H. (2021) 'Wireless sensors application in smart English classroom design based on artificial intelligent system', *Microprocess Microsyst*., Vol. 81, DOI: 10.1016/j.micpro.2020.103798.

Yudiono, H., Sumbodo, W., Salim, S. and Setiadi, R. (2019) 'Improving the milling machine competency learning outcomes through industrial project-based learning for vocational school students', *Jurnal Pendidikan Vokasi*, Vol. 9, No. 2, DOI: 10.21831/jpv.v9i2.23686.

Zhang, D. (2022) 'Analysis of university management model of national higher education institutions based on machine learning algorithm', *Mobile Information Systems*, Vol. 2022, DOI: 10.1155/2022/4553185.

Zheng, X. and Li, C. (2024) 'Predicting students' academic performance through machine learning classifiers: a study employing the Naive Bayes classifier (NBC)', *International Journal of Advanced Computer Science and Applications*, Vol. 15, No. 1, DOI: 10.14569/IJACSA.2024.0150199.