

Article

An Interpretable Hybrid RF–ANN Early-Warning Model for Real-World Prediction of Academic Confidence and Problem-Solving Skills

Mostafa Aboulmour Salem ^{1,*}  and Zeyad Aly Khalil ² 

¹ Deanship of Development and Quality Assurance, King Faisal University, Al-Ahsa 31982, Saudi Arabia

² Obour High Institute for Management & Informatics, Cairo 11828, Egypt; zeiadaly@oi.edu.eg

* Correspondence: masalem@kfu.edu.sa

Abstract

Early identification of students at risk for low academic confidence, poor problem-solving skills, or poor academic performance is crucial to achieving equitable and sustainable learning outcomes. This research presents a hybrid artificial intelligence (AI) framework that combines feature selection using a Random Forest (RF) algorithm with data classification via an Artificial Neural Network (ANN) to predict risks related to Academic Confidence and Problem-Solving Skills (ACPS) among higher education students. Three real-world datasets from Saudi universities were used: MSAP, EAAAM, and MES. Data preprocessing included Min–Max normalisation, class balancing using SMOTE (Synthetic Minority Oversampling Technique), and recursive feature elimination. Model performance was evaluated using five-fold cross-validation and a paired *t*-test. The proposed model (RF-ANN) achieved an average accuracy of 98.02%, outperforming benchmark models such as XGBoost, TabNet, and an Autoencoder–ANN. Statistical tests confirmed the significant performance improvement ($p < 0.05$; Cohen’s $d = 1.1$ – 2.7). Feature importance and explainability analysis using a Random Forest and Shapley Additive Explanations (SHAP) showed that psychological and behavioural factors—particularly study hours, academic engagement, and stress indicators—were the most influential drivers of ACPS risk. Hence, the findings demonstrate that the proposed framework combines high predictive accuracy with interpretability, computational efficiency, and scalability. Practically, the model supports Sustainable Development Goal 4 (Quality Education) by enabling early, transparent identification of at-risk students, thereby empowering educators and academic advisors to deliver timely, targeted, and data-driven interventions.

Keywords: real-world; academic confidence and problem-solving skills (ACPS); artificial intelligence (AI); artificial neural networks (ANN); random forest (RF); educational data mining (EDM); explainable artificial intelligence (XAI); sustainable development goals (SDGs); early-warning systems



Academic Editors: Muhammad Asad Arshed, Atif Alvi and Ștefan Cristian Gherghina

Received: 30 November 2025

Revised: 13 December 2025

Accepted: 17 December 2025

Published: 18 December 2025

Citation: Salem, M.A.; Khalil, Z.A. An Interpretable Hybrid RF–ANN Early-Warning Model for Real-World Prediction of Academic Confidence and Problem-Solving Skills. *Math. Comput. Appl.* **2025**, *30*, 140. <https://doi.org/10.3390/mca30060140>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The fourth target of the Sustainable Development Goals (SDG 4), related to quality education, aims to ensure inclusive, equitable, and quality education for all and to promote lifelong learning opportunities [1]. Higher education is a key driver of national development, as it directly contributes to institutional development and the preparation of qualified professional cadres [2]. Enhancing the academic success of higher education students is

very important, as it has a direct impact on the quality of the educational process and the preparedness of graduates for future careers [3].

While academic performance has traditionally been used as the primary indicator of educational success, this study introduces Academic Confidence and Problem-Solving Skills (ACPS) as an early predictor of academic risk. ACPS integrates self-reported measures of academic confidence, problem-solving ability, and psychological well-being, including stress and anxiety [4]. In this study, ACPS is defined as a continuous composite index derived from validated items assessing: (i) academic confidence (e.g., perceived ability to understand course material and perform well in assessments), (ii) self-reported problem-solving skills, and (iii) stress and anxiety indicators adapted from the DASS-21 and GHQ-12 scales. All items are measured using 5-point Likert scales, with higher values reflecting more favourable states (i.e., higher confidence and problem-solving ability and lower stress and anxiety). Item scores are z-standardised and averaged to produce a continuous ACPS index for each student, which is subsequently min–max normalised to the [0, 1] range. For early-risk prediction, this continuous index is converted into a binary target variable: students below the 30th percentile of the ACPS distribution is classified as at risk (label = 1), but those at or above this threshold are classified as not at risk (label = 0) within each dataset (see Section 3.1). Accordingly, the proposed RF–ANN framework conceptualises academic performance from an early-risk perspective centred on ACPS.

In higher education, Academic Confidence and Problem-Solving Skills are commonly reflected in students' achievement of predefined learning outcomes, including conceptual understanding, assessment performance, engagement in learning activities, and the demonstration of cognitive and analytical competencies [4–6]. Students' educational intentions and their level of engagement within the learning environment further shape these outcomes [7]. Consequently, ACPS serves not only as an indicator of individual student success but also as a benchmark for institutional effectiveness and alignment with the Sustainable Development Goals [8,9].

However, low levels of academic confidence and problem-solving skills among higher education students pose a significant challenge [10], leading to reduced self-efficacy, increased psychological stress, fewer employment opportunities, and weaker analytical and problem-solving abilities [11]. At the institutional level, these problems undermine the quality of education and the efficiency of national higher education systems [12]. This challenge is particularly prominent in Saudi Arabia, where current reforms focus on raising higher education standards to align with national priorities and international benchmarks [13].

Accordingly, this study adopts an explicit SDG-4-oriented objective: to develop a transparent, data-driven early-warning system that enables timely academic and socio-affective support for students while helping teachers and academic advisors prioritise interventions based on interpretable risk factors. Early identification of academic risk directly supports SDG 4 by enabling intervention before academic failure or withdrawal occurs. Consistent with learning analytics and Educational Data Mining (EDM) research, early prediction models aim to identify at-risk students as early as possible to improve academic outcomes and reduce dropout rates.

Empirically, the proposed RF–ANN framework demonstrates strong performance across three datasets; for example, on the MES dataset, it achieved 97.18% accuracy and a 96.73% F1-score, outperforming ANN and other baseline models (Table 4). In addition, it achieved the fastest prediction time (0.06 s) among the compared models (Table 5), supporting near-real-time screening in academic dashboards. These improvements translate into practical benefits: students can be identified earlier for support, and educators gain interpretable insights—through Random Forest feature importance and SHAP—to guide

targeted interventions addressing issues such as low engagement, insufficient study time, or elevated stress, rather than relying solely on grades.

Specifically, this study aims to address the following research objectives: to operationalise ACPS as a multidimensional early-risk construct that integrates cognitive, behavioural, and socio-affective indicators beyond grade-based measures; to design and implement an interpretable hybrid RF–ANN framework that combines Random Forest–based feature selection with Artificial Neural Network classification for structured educational data; to assess the predictive performance and robustness of the proposed RF–ANN model across multiple real-world higher-education datasets (MSAP, EAAAM, and MES); to compare the proposed model with established machine-learning and deep-learning baselines (ANN, XGBoost, TabNet, Autoencoder–ANN, and RF) in terms of accuracy, robustness, and generalisability; to enhance model transparency and trustworthiness through Explainable Artificial Intelligence (XAI) techniques, specifically Random Forest feature importance and SHAP analysis; and to evaluate computational efficiency and real-time feasibility for scalable deployment in academic early-warning and learning analytics systems.

To achieve these objectives, the study proposes a hybrid modelling approach that integrates Random Forest (RF) feature selection with Artificial Neural Network (ANN) classification. The RF component identifies and ranks the most influential predictors, thereby reducing redundancy and improving interpretability, while the ANN captures complex nonlinear relationships to classify students by ACPS risk. The framework incorporates key preprocessing steps, including normalisation, class balancing using SMOTE, and recursive feature elimination. The model is trained and validated using three real-world datasets from Saudi higher-education institutions and benchmarked against advanced models such as ANN, XGBoost, TabNet, and Autoencoder–ANN. The results show that the RF–ANN framework achieves superior predictive performance while remaining interpretable and computationally efficient, making it a robust and scalable solution for ACPS prediction and data-driven academic decision-making.

Although Random Forest and Artificial Neural Networks are well-established techniques, this study advances EDM through a clear methodological contribution and a distinct early-risk formulation. Specifically, it (i) introduces ACPS as a multidimensional early-risk construct beyond grade-centric prediction; (ii) implements a two-stage, interpretable RF–ANN pipeline that combines embedded feature selection with nonlinear classification; (iii) integrates explainability using RF importance and SHAP to produce actionable risk insights; (iv) applies a rigorous, leakage-safe preprocessing and evaluation protocol; and (v) validates the framework across three heterogeneous datasets while demonstrating low inference latency suitable for near-real-time deployment. Collectively, these contributions position the proposed RF–ANN model as a transparent, scalable, and SDG-4-aligned early-warning tool for higher education.

2. Theoretical Review and Research Hypothesis

2.1. Predictive Modelling in Educational Data Mining (EDM)

Educational Data Mining (EDM) has emerged as a prominent research direction, extracting actionable insights from academic datasets to support institutional decision-making and proactively identify students at risk of academic underperformance [14].

Traditional Machine Learning (ML) algorithms—such as Decision Trees (DTs), Support Vector Machines (SVMs), and Logistic Regression—have achieved moderate success in predicting Academic Confidence and Problem-Solving Skills (ACPS), primarily due to their interpretability and computational efficiency [15,16]. However, these models often fail to capture complex, nonlinear relationships among educational variables, thereby limiting their predictive power across diverse learning environments [17].

Several studies [18–21] have attempted to address these limitations using neural networks and computational optimisation techniques, but they have encountered challenges related to data imbalance and poor generalisation. Recent studies [21–24] have also addressed the application of hybrid and resampling strategies to improve performance. However, most prior research lacked comprehensive data processing and effective dimensionality-reduction techniques, thereby undermining the models' robustness and generalizability. Accordingly, the current study formulates the following two hypotheses:

H1. *The proposed hybrid RF-ANN model achieves significantly higher predictive accuracy compared to traditional models such as ANN, XGBoost, TabNet, Autoencoder-ANN, and RF alone.*

H2. *The RF-ANN model maintains superior predictive performance across multiple learning datasets (MSAP, EAAAM, and MES), demonstrating strong external validity.*

2.2. Hybrid AI Architectures and Preprocessing in Performance Prediction

Recent research [25–28] points to the advantages of hybrid and ensemble models that combine multiple complementary AI techniques. Examples include Particle Swarm Optimisation combined with Deep Neural Networks (PSO-DNNs) [29], Improved Conditional Generative Adversarial Networks integrated with Deep Support Vector Machines (ICGAN-DSVMs) [30], and Ensemble Generative Adversarial Networks (EGANs) [31], all of which have demonstrated notable improvements in predictive accuracy. However, these models often require large datasets and extensive computational resources, which limit their practicality in educational environments.

Similarly, sequential learning models—such as Long Short-Term Memory (LSTM), Gated Recurrent Units (GRUs), and Variational Autoencoders—have shown varying degrees of success in modelling time-dependent educational data [32,33]. However, its scalability remains limited in higher education contexts, where data availability is often restricted and context-specific [34]. Artificial Neural Networks (ANNs) remain effective in capturing nonlinear relationships, while the Random Forest (RF) algorithm is widely used for feature selection and data smoothing [35,36]. Therefore, combining the two methods into a hybrid (RF-ANN) framework offers the dual advantage of improved predictive performance and increased interpretability [37].

Moreover, data preprocessing techniques such as Min–Max normalisation [38], Synthetic Minority Oversampling Technique (SMOTE) [39], and Recursive Feature Elimination (RFE) are crucial for enhancing model robustness, reducing overfitting, and ensuring balanced classification results [37,40]. However, most previous studies have not employed comprehensive data processing or effective dimensionality-reduction techniques, thereby limiting the generalizability and replicability of their results.

At the theoretical level, this study contributes to the link between Educational Data Mining (EDM) and cognitive-affective psychology by characterising students' academic risk in terms of predicting ACPS levels and by linking EDM methodologies to established psychological theories, such as the stress-buffering model [41] and self-determination theory [42]. The study hypothesises that cognitive (e.g., performance expectancy and self-efficacy) and affective (e.g., stress, anxiety, and motivation) factors interact as integrated determinants of academic outcomes. By incorporating these variables into a hybrid artificial intelligence framework, the study expands the scope of EDM toward socio-affective modelling in higher education. Accordingly, the following hypotheses are proposed:

H3. *Using a random forest algorithm for feature selection, followed by artificial neural networks for classification, improves predictive accuracy by eliminating redundant or irrelevant variables.*

H4. *Preprocessing techniques—including Min–Max normalisation and SMOTE—improve the model’s generalizability and classification performance, especially on imbalanced datasets.*

H5. *The prediction time of the RF–ANN model is faster or comparable to other models, supporting its potential application in real-time detection and academic intervention.*

3. Materials and Methods

To develop and evaluate the proposed hybrid predictive framework based on the Random Forest algorithm and Artificial Neural Networks (RF–ANNs) for predicting Academic Confidence and Problem-Solving Skills (ACPS) among higher education students, the research methodology is organised into several basic stages, including dataset description, data preprocessing, feature selection using the Random Forest (RF) algorithm, Artificial Neural Network (ANN) model development, analytical comparison with benchmark models, and evaluation of the overall performance of the model. This methodological organisation ensures transparency, reproducibility, and alignment with established research best practices in Educational Data Mining (EDM).

3.1. Dataset Description

The academic dataset used in this study was collected from 641 students at higher education institutions across five major regions of the Kingdom of Saudi Arabia: Central, Western, Eastern, Southern, and Northern. This dataset includes a comprehensive range of academic records, including grades, attendance rates, behavioural indicators, and study habits. Due to its richness, diversity, and organised structure, this dataset is a good basis for deep learning and hybrid AI applications. These characteristics also enhance the generalizability and external validity of the proposed predictive model (RF–ANN) in practice (see Table 1).

Table 1. Summary of the datasets used in the study.

Dataset	Year	Source/Universities	Instances	Attributes	Feature Categories
MSAP	2024	KFU, IMSIU, NBU, UG	141	33	Demographic,
EAAAM	2024	Experimental Collection (3 Semesters)	641	12	Academic,
MES	2024	Higher Education Institutions (National Dataset)	4512	32	Behavioural

This subsection clarifies the distributional and demographic characteristics of the datasets used in this study. Although the MSAP, EAAAM, and MES datasets differ substantially in scale and feature dimensionality, they share a standard structure centred on demographic, academic, behavioural, and socio-affective indicators, enabling meaningful cross-dataset comparison.

The MSAP dataset is the smallest ($n = 141$) and was collected from four Saudi universities. It contains a rich feature space (33 variables) spanning demographic attributes (e.g., GPA, study habits), behavioural indicators (e.g., attendance and engagement), and academic performance measures. Given its small sample size, MSAP exhibits higher variance and class imbalance, making it particularly suitable for evaluating model robustness under data-scarce conditions. The EAAAM dataset ($n = 641$) represents a medium-scale experimental cohort collected over three academic semesters. It comprises 12 core variables that address demographic, academic, and behavioural dimensions. While its feature space is more compact, EAAAM provides greater statistical stability than MSAP and serves as an intermediate benchmark between small-scale and large-scale educational datasets.

The MES dataset is the largest and most diverse dataset ($n = 4512$), offering a national-level representation of Saudi higher-education students. It includes 32 variables covering

demographic information, academic records, behavioural engagement, and ACPS-related indicators across multiple institutions and semesters. The size and diversity of MES make it particularly suitable for assessing model scalability and external validity. Across all datasets, the binary ACPS risk label was derived using a consistent percentile-based strategy. Approximately the lowest 30% of students in each dataset, based on the normalised ACPS composite index, were classified as at risk, while the remaining students were classified as not at risk. As a result, class distributions were intentionally made comparable across datasets despite differences in sample sizes. Sensitivity analyses using alternative percentile thresholds (25% and 35%) yielded consistent trends, confirming the robustness of this operationalisation.

From a demographic perspective, all datasets include students from diverse academic backgrounds and study profiles, with variability in GPA, study hours, attendance rates, engagement levels, and stress/anxiety indicators. While personal identifiers (e.g., gender, age, national ID) were removed to protect privacy and reduce bias, the retained academic and behavioural variables provide sufficient diversity to model real-world heterogeneity in higher-education learning contexts. Thus, the combined use of a small-scale (MSAP), medium-scale (EAAAM), and large-scale (MES) dataset enables a comprehensive evaluation of the proposed RF-ANN framework across varying data distributions, feature dimensionalities, and institutional contexts. This multi-dataset design strengthens the generalisability of the findings and addresses a key limitation of prior EDM studies that rely on single-institution or homogeneous samples.

All datasets used in this study were collected under ethical approval number KFU-2025-ETHICS3631 issued by King Faisal University (KFU). Participation in the research was completely voluntary and anonymous, with no incentives offered or identifying information collected. All mental health indicators were self-reported using validated, modified standardised instruments from the Depression, Anxiety, and Stress Scale (DASS-21) and the General Mental Health Questionnaire (GHQ-12).

All preprocessing procedures were conducted on fully anonymised data, in accordance with institutional data protection policies and the ethical principles of the Declaration of Helsinki (2013 Update). The MSAP dataset was collected in 2024 from four major Saudi universities—King Faisal University (KFU), Imam Mohammad ibn Saud Islamic University (IMSIU), Northern Border University (NBU), and the University of Jeddah (UG). It includes 141 cases and 32 variables representing demographic, behavioural, and academic characteristics [43].

The Experimental ACPS Analytics Mining (EAPAM) dataset includes 641 student records, each containing 12 variables categorised within demographic, academic, and behavioural dimensions. This dataset was pilot-collected over two semesters in 2024, with 251 records in the first semester and 249 in the second semester [44]. The Higher Education Institutions (HEI) dataset provides a national representation of higher education students in the Kingdom, comprising 4512 student cases and 32 characteristics covering demographic and academic aspects, as well as Academic Confidence and Problem-Solving Skills (ACPS) scores across two semesters [43].

For each student, we first computed a continuous ACPS composite score by aggregating items that capture: (i) academic confidence (e.g., confidence in understanding course material and performing well in assessments), (ii) perceived problem-solving skills, and (iii) stress/anxiety indicators derived from adapted items of the DASS-21 and GHQ-12 scales. All items were measured on 5-point Likert scales and were coded such that higher values reflected more favourable states (higher confidence and problem-solving, lower stress/anxiety). Item responses were z-standardised at the dataset level and then averaged to obtain a single continuous ACPS index for each student. This index was

subsequently min–max normalised to the [0, 1] interval to ensure comparability across datasets and institutions.

Additionally, to construct the binary early-risk outcome used as the target variable in all classification models, the normalised ACPS index was transformed into a categorical label separately within each dataset. Students whose ACPS scores fell below the 30th percentile of the distribution were coded as 1 = “at-risk”, but those with scores at or above the 30th percentile were coded as 0 = “not-at-risk”. This percentile-based cut-off reflects institutional practice for flagging approximately the lowest 30% of students for early academic and psychological support. Sensitivity analyses using 25th- and 35th-percentile thresholds yielded similar patterns in model performance, supporting the robustness of this operationalisation. Representative, anonymised sample rows illustrating the ACPS composite index and the corresponding binary risk label are presented in Table 2.

Table 2. Representative sample rows from the study datasets.

Dataset	Student ID *	GPA	Study Hours	Attendance (%)	Engagement Score	Stress Score	Assessment Score	ACPS Composite	ACPS Risk Label
MSAP	S-001	3.40	12	89	4.1	1.8	82	0.73	0
MSAP	S-002	2.10	4	65	2.6	3.5	58	0.28	1
EAAAM	E-015	3.20	10	92	3.9	2.1	76	0.69	0
EAAAM	E-076	1.90	3	61	2.2	3.8	52	0.22	1
MES	M-542	3.80	15	95	4.4	1.5	88	0.81	0
MES	M-210	2.30	5	68	2.8	3.1	60	0.33	1

* Student IDs are randomly generated and do not correspond to real individuals. All values are shown for illustration only.

The characteristics extracted from these three datasets—including demographic variables (e.g., age, GPA, and study habits), behavioural indicators (e.g., attendance and participation), and academic variables (e.g., assessment scores)—were used as predictive inputs into the proposed hybrid (RF-ANN) model. All datasets and model implementation details are available at the following link: (<https://shorturl.at/z1rEE>, accessed on 30 November 2025).

To enhance clarity regarding the structure and content of the datasets used in this study, Table 2 presents representative, fully anonymised sample rows from the MSAP, EAAAM, and MES datasets. These examples illustrate the types of demographics, behavioural, and academic variables included, as well as the computed ACPS composite index that underpins the binary early-risk classification. All sample records are provided for illustrative purposes only and do not correspond to real individuals.

3.2. Data Preprocessing (DP)

Data Preprocessing (DP) is a fundamental step in developing any predictive model, as it has a direct impact on the quality, reliability, and interpretability of the results. In this study, a series of systematic procedures was implemented to ensure that the datasets were clean, standardised, and ready for machine learning applications. The binary target variable used in all experiments corresponds to the ACPS early-risk label described in Section 3.1, where students below the 30th percentile of the ACPS composite index are coded as ‘at-risk’ (1), while all others are coded as ‘not-at-risk’ (0).

In addition, to avoid data leakage and ensure an unbiased evaluation of model performance, class-imbalance handling was restricted strictly to the training data. First, each dataset was stratified and split into 85% for training and 15% for a hold-out test set, preserving the original proportions of at-risk and not-at-risk students. Because the ACPS risk label is defined using a percentile rule (the lowest 30% are at risk), each dataset is moderately imbalanced by construction ($\approx$ approximately 30% at risk vs. approximately 70% not at risk). Importantly, SMOTE was only applied to the training data to prevent data leakage;

the whole dataset and the hold-out test remained unchanged. After a stratified 85/15 split, SMOTE oversampled the minority (at-risk) class within the training set until both classes were balanced (1:1). Figure 1 reports the class counts before SMOTE and the number of synthetic samples generated by SMOTE in each dataset.

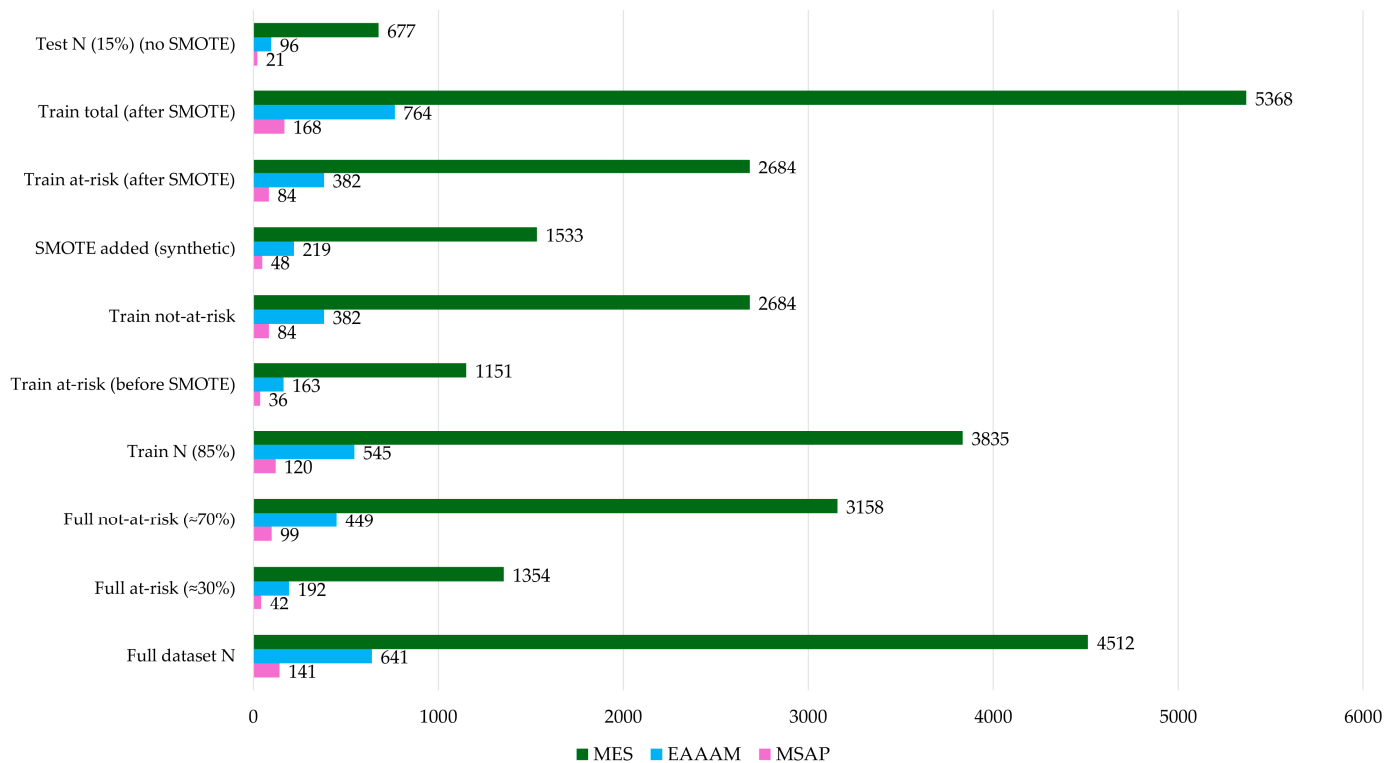


Figure 1. Dataset sizes and class distributions before/after SMOTE; **Notes:** Counts follow the percentile-based label definition ($\approx 30\%$ at-risk) and stratified 85/15 splitting; SMOTE balances the training set by generating synthetic minority samples equal to ($majority - minority$).

Within the training set only, class imbalance was addressed using the Synthetic Minority Oversampling Technique (SMOTE), which generates synthetic samples for the minority (at-risk) class by interpolating between neighbouring instances in feature space. SMOTE was applied inside the cross-validation loop, i.e., it was fitted on the training folds and never on the validation or test folds. The hold-out test set remained completely untouched by any resampling procedure and was used solely for out-of-sample evaluation. This protocol ensures that the model does not indirectly “see” information from the test set during training and that the reported performance metrics reflect realistic generalisation to unseen data. Likewise, the preprocessing pipeline consisted of the following steps:

- (A) **Handling Missing Values and Incomplete Records:** Real-world academic datasets often contain missing values or incomplete entries due to data-entry errors or partially completed questionnaires. To preserve data quality and consistency, all records containing missing or inconsistent values were removed. In addition, personal identifiers—such as name, gender, national ID number, or age—were excluded to protect participant privacy and minimise the risk of algorithmic bias. All analyses were therefore performed exclusively on complete, fully anonymised datasets.
- (B) **Encoding Categorical Variables:** Because most machine learning algorithms operate on numerical inputs, categorical variables were encoded as numerical values using label encoding. This technique assigns a unique integer to each category, enabling the model to process non-numeric attributes—such as engagement patterns or study habits—while preserving their semantic meaning.

- (C) **Data Normalisation:** The datasets contained features measured on different scales, which could lead to numerical instability or bias during model training. To standardise features measured on various scales and ensure numerical stability during model training, Min–Max normalisation was applied to all continuous variables. This transformation rescales each feature to the interval $[0, 1]$ using Equation (1):

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

where x denotes the original feature value, x' is the normalised value, and $\min(x)$ and $\max(x)$ represent the minimum and maximum values of that feature, respectively.

- (D) **Feature Selection Using Recursive Feature Elimination (RFE):** To reduce model complexity and prevent overfitting, Recursive Feature Elimination (RFE) was applied. This technique iteratively removes less informative variables based on model performance, ensuring that only the most influential features are retained. This refinement improves predictive accuracy, enhances interpretability, and decreases computational overhead [30].
- (E) **Splitting the Data into Training and Test Sets:** Following preprocessing, the dataset was divided into an 85% training set and a 15% test set. The training set was used for model fitting and hyperparameter optimisation, while the test set provided an unbiased evaluation of model accuracy and generalisation. This split supports realistic performance assessment in practice [31].

3.3. A Hybrid AI Approach Description

The proposed hybrid framework integrates the interpretability and embedded feature-ranking capability of the Random Forest (RF) algorithm with the nonlinear learning capacity of Artificial Neural Networks (ANNs) to improve the prediction of Academic Confidence and Problem-Solving Skills (ACPS). Let $X \in \mathbb{R}^{n \times m}$ denote the input feature matrix, where n is the number of students and m is the number of candidate predictors.

RF estimates an importance score I_j for each feature x_j . Features with importance exceeding a threshold τ (set to the mean importance across all features) are retained to form a reduced feature set:

$$X' = \{x_j \mid I_j > \tau\}. \quad (2)$$

Equation (2) formalises the RF-driven dimensionality-reduction step, which helps reduce redundancy, enhance interpretability, and reduce the risk of overfitting. The selected feature subset X' is then passed to a feedforward ANN $f_\theta(\cdot)$ with parameters θ , producing a risk probability $\hat{y}$ through nonlinear transformation and sigmoid output activation:

$$\hat{y} = f_\theta(X') = \sigma(W_2 \phi(W_1 X' + b_1) + b_2), \quad (3)$$

where $\phi(\cdot)$ denotes a nonlinear activation function (e.g., ReLU) and $\sigma(\cdot)$ is the sigmoid function for binary classification. Model parameters θ are learned by minimising the binary cross-entropy loss over N training instances:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]. \quad (4)$$

Equation (4) defines the objective function optimised using the Adam optimiser during backpropagation. Overall, this two-stage pipeline preserves model transparency through RF-based feature importance (Equation (2)) while enabling high-capacity nonlinear classification through the ANN mapping (Equation (3)) optimised under the loss in Equation (4).

3.3.1. Random Forest Component (RF)

Random Forest (RF) is an ensemble learning method that constructs multiple decision trees using bootstrap sampling and random feature subspaces, and aggregates their outputs to improve generalisation. In this study, RF is mainly used as an built-in feature selection, producing importance scores I_j that quantify each variable's contribution to reducing impurity across the ensemble. Retaining only the most informative predictors (as defined in Equation (2)) improves robustness, reduces noise sensitivity, and supports interpretability by enabling transparent ranking of ACPS-related drivers [32,33].

3.3.2. Artificial Neural Network Component (ANN)

Artificial Neural Networks (ANNs) are effective for modelling complex, nonlinear relationships between predictors and outcomes, particularly when feature interactions are not well captured by linear models [34,35]. A standard feedforward ANN was used, comprising: (i) an input layer that receives RF-selected features X' ; (ii) one or more fully connected hidden layers with nonlinear activation (e.g., ReLU); (iii) a dropout layer to reduce overfitting by randomly deactivating neurons during training; and (iv) a sigmoid output layer producing the probability of the at-risk class. The computation in layer l is given by:

$$a^{(l)} = f(W^{(l)}a^{(l-1)} + b^{(l)}), \quad (5)$$

where $a^{(l)}$ is the activation vector of the l -th layer, $f(\cdot)$ is the activation function, and $W^{(l)}$ and $b^{(l)}$ denote weights and biases, respectively.

During training, parameters are updated by backpropagation to minimise the loss in Equation (4), enabling the ANN to refine representations of ACPS-related patterns and improve classification reliability.

3.3.3. Integration of RF and ANN

The RF-ANN integration is designed to combine feature-level interpretability with nonlinear predictive power. Specifically, RF first reduces dimensionality by identifying and retaining the most informative ACPS predictors (Equation (2)). The ANN then learns nonlinear mappings from these refined inputs to the ACPS risk label (Equation (3)), with optimisation guided by the binary cross-entropy objective (Equation (4)). This division of labour reduces the burden on the ANN to learn from noisy or redundant features, improves generalisation across diverse datasets, and yields a practical framework suitable for scalable EDM deployments and early-warning systems in higher education.

3.4. Baseline Methods

This subsection outlines the machine learning and deep learning baselines used to assess the effectiveness of the proposed RF-ANN hybrid model. These baseline models were selected based on their proven suitability for structured educational datasets, their efficiency in feature representation and learning, and their compatibility with non-sequential ACPS prediction tasks.

3.4.1. Artificial Neural Networks (ANNs)

ANNs are fundamental models in deep learning, inspired by the architecture and functioning of biological neural systems. A typical ANN includes an input layer, one or more hidden layers, and an output layer, with each layer consisting of interconnected neurons [37,45]. During training, the network learns by adjusting the weights of these connections via backpropagation. ANNs are particularly effective at modelling complex, non-linear relationships within structured educational data [37]. In this study, ANNs serve as both a baseline model and a core component of the proposed RF-ANN framework.

3.4.2. Extreme Gradient Boosting (XGBoost)

XGBoost is a high-performance ensemble learning algorithm based on gradient-boosted decision trees [46]. The algorithm builds trees sequentially, with each tree correcting the errors of its predecessors by minimising a regularised objective function via gradient descent [39]. This algorithm is known for its high computational efficiency, scalability, and ability to handle complex feature interactions [39,40].

3.4.3. TabNet

TabNet is a deep learning architecture designed explicitly for tabular data. It relies on sequential attention to identify the most relevant features at each decision step, thereby enhancing the model's interpretability and efficiency [41]. Unlike Recurrent Neural Networks (RNNs), TabNet is optimised explicitly for handling non-sequential structured data, making it suitable for predicting students' ACPS levels [41,42].

3.4.4. Autoencoder-ANN Hybrid

Autoencoders are used in unsupervised learning for data compression and reconstruction. In this configuration, autoencoders are used to reduce dimensionality and remove noise from the data [43]. The resulting encoded representations are then passed to an ANN for classification [24]. This combination enhances feature abstraction, reduces noise, and achieves stable, accurate predictions in ACPS modelling [44].

3.4.5. Random Forest (RF)

The RF algorithm is an ensemble of decision trees that uses bootstrapping and random feature selection to enhance generalizability [45]. As the component responsible for feature analysis in the proposed model, it captures nonlinear relationships and provides interpretive scores for feature importance, making it ideal for educational tabular data [46,47].

3.4.6. Hyperparameter Optimisations and Evaluation Protocol

Models designed for spatial or sequential learning, such as convolutional neural networks (CNNs) and recurrent architectures (LSTM, BiLSTM, and GRU), were excluded from the final comparison because the ACPS datasets lack pixel-level, temporal, or sequential structure. Consequently, including such models would not provide a theoretically meaningful or methodologically valid benchmark for structured, non-sequential educational data. Hyperparameters were optimised using a grid-search strategy over predefined ranges: learning rate $\in \{0.001, 0.005, 0.01\}$, batch size $\in \{16, 32, 64\}$, dropout rate $\in \{0.1, 0.2, 0.3\}$, and number of hidden layers $\in \{1, 2, 3\}$.

The validation dataset served multiple purposes, including hyperparameter tuning, monitoring generalisation performance, detecting overfitting, and triggering early stopping when the validation loss no longer improved. The optimal configuration was selected based on the average validation accuracy across five folds, ensuring balanced performance while preventing overfitting to any single data partition. Additionally, overfitting was further controlled through the following mechanisms: An 85% training/15% testing split stratified by class; An internal 10% validation subset drawn exclusively from the training data to guide hyperparameter selection, track generalisation performance, detect overfitting, and enable early stopping; Application of dropout (rate = 0.2) and L2 regularisation ($\lambda = 0.001$); Continuous monitoring of training and validation loss convergence [48].

Training was automatically terminated when the validation loss failed to improve for 15 consecutive epochs. In addition to the 85:15 hold-out evaluation, a five-fold cross-validation procedure was conducted to assess the model's robustness and generalisability.

The resulting mean $\pm$ standard deviation values across all evaluation metrics (accuracy, precision, recall, and F1-score) were highly stable, confirming the reliability of the proposed RF-ANN hybrid framework.

3.4.7. Clarification and Justification of ANN Hyperparameters

The hyperparameters of the Artificial Neural Network (ANN) components, specifically, the number of neurons and the dropout rate, were selected through a combination of empirical validation and theoretical considerations regarding dataset size, feature dimensionality, and overfitting risk. Regarding the number of neurons, configurations with {32, 64, 96} neurons in the hidden layer were systematically evaluated using grid search and five-fold cross-validation. The configuration with 64 neurons consistently achieved the best balance between predictive performance and generalisation across all three datasets (MSAP, EAAAM, and MES). Smaller architectures (32 neurons) exhibited reduced representational capacity, resulting in underfitting. In contrast, larger architectures (96 neurons) did not yield statistically significant performance gains and exhibited higher variance, particularly on the smaller MSAP dataset. Selecting 64 neurons aligns with established best practices for tabular educational data, in which a moderate network width is sufficient to capture nonlinear interactions without excessive model complexity [37,45].

The dropout rate was tuned over the range {0.1, 0.2, 0.3, 0.5}. A dropout rate of 0.2 was selected during optimisation because it consistently reduced overfitting while preserving stable convergence across folds. Lower dropout values (≤ 0.1) provided insufficient regularisation, especially in datasets with limited instances, but higher dropout values (≥ 0.3) degraded performance by suppressing too much representational capacity. The selected dropout rate is consistent with prior EDM and tabular deep-learning studies, which recommend moderate dropout to improve robustness without destabilising training dynamics [21,37].

Together, the chosen hyperparameters (64 neurons, dropout = 0.2) reflect a principled trade-off between model expressiveness, computational efficiency, and generalisability. Importantly, these values were not fixed a priori but were selected based on cross-validated performance stability across diverse datasets, ensuring that the reported results are robust and not artefacts of over-parameterisation.

3.5. Proposed RF-ANN Hybrid AI Architecture

The selected feature matrix X_{RF} is passed to a feedforward ANN that captures nonlinear relationships between student characteristics and ACPs risk. At each hidden layer l , the ANN computes a linear transformation of the incoming activations, followed by a nonlinear activation function. The weighted sum at layer l is computed as shown in Equation (6):

$$Z^{(l)} = W^{(l)} A^{(l-1)} + b^{(l)}, \quad (6)$$

where $W^{(l)}$ and $b^{(l)}$ denote the weight matrix and bias vector, respectively, and $A^{(l-1)}$ represents the activations from the previous layer.

To introduce nonlinearity and improve convergence, the Rectified Linear Unit (ReLU) activation is applied, as defined in Equation (7):

$$A^{(l)} = \text{ReLU}(Z^{(l)}) = \max(0, Z^{(l)}). \quad (7)$$

To reduce overfitting, dropout regularisation is applied during training by randomly deactivating neurons with probability p , as expressed in Equation (8):

$$\tilde{A}^{(l)} = \text{Dropout}(A^{(l)}, p). \quad (8)$$

This mechanism forces the network to learn more robust feature representations that generalise across datasets. The final ANN layer applies a sigmoid activation function to map the output to a probability value $\hat{y} \in [0, 1]$, representing the likelihood that a student belongs to the *at-risk* class. This operation is defined in Equation (9):

$$\hat{y} = \sigma(Z^{(L)}) = \frac{1}{1 + e^{-Z^{(L)}}}, \quad (9)$$

where L denotes the index of the output layer.

Model prediction error is quantified using the binary cross-entropy loss function, which penalises incorrect probabilistic predictions, as shown in Equation (10):

$$\mathcal{L}(y, \hat{y}) = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})], \quad (10)$$

where $y \in \{0, 1\}$ denotes the accurate ACPS risk label and $\hat{y}$ is the predicted probability obtained from Equation (9). Network parameters are updated via backpropagation using gradient descent. The weight and bias updates at layer l follow Equation (11):

$$W^{(l)} \leftarrow W^{(l)} - \alpha \frac{\partial \mathcal{L}}{\partial W^{(l)}}, b^{(l)} \leftarrow b^{(l)} - \alpha \frac{\partial \mathcal{L}}{\partial b^{(l)}}, \quad (11)$$

where α is the learning rate.

In this study, optimisation was performed using the Adam optimiser, which adaptively adjusts learning rates to accelerate convergence and improve stability. By explicitly combining RF-based embedded feature selection (Equation (2)) with ANN-based nonlinear mapping (Equations (6)–(9)) and principled optimisation (Equations (10) and (11)), the proposed RF–ANN architecture achieves a strong balance between interpretability, predictive accuracy, and generalisability. The RF component ensures transparency and dimensionality reduction, while the ANN component captures higher-order interactions among behavioural, cognitive, and affective predictors. As a result, the RF–ANN hybrid framework is a scalable and theoretically grounded solution for Educational Data Mining (EDM) applications and ACPS-oriented early-warning systems in higher education.

3.6. Experimental Setup

The experimental evaluation was conducted using Python (version 3.11.9) on a Windows 11 platform, running on a personal computer equipped with an Intel® Core™ i7-1135G7 CPU (2.40 GHz) and 8 GB of RAM. The proposed RF–ANN hybrid model was trained using the prepared datasets under rigorous optimisation procedures to ensure reliable results. Model training employed the Adam optimisation algorithm due to its adaptive learning-rate adjustment and high computational efficiency. Because the classification task is binary (at-risk versus not-at-risk), the binary cross-entropy loss function was used to quantify prediction error.

During early exploration and experimentation, convolutional neural network (CNN) architectures were initially evaluated. However, these models were excluded from the final benchmarking analysis. CNNs are inherently designed for spatially structured inputs (e.g., images or grid-based data) and therefore lack theoretical alignment with the structured, non-sequential tabular nature of the ACPS datasets used in this study. Including CNN-based results in the final evaluation could lead to methodologically misleading comparisons.

Similarly, recurrent neural network (RNN) architectures—including Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and Gated Recurrent Unit (GRU) models—were considered during preliminary testing but subsequently excluded. These architectures

are specifically designed to model temporal or ordered sequences, which are absent in ACPs datasets composed exclusively of structured tabular features. As such, CNN- and RNN-based models were excluded, as their inclusion would not yield a theoretically meaningful or methodologically valid comparison.

Accordingly, the final evaluation focuses exclusively on models optimised for tabular educational data—namely, Artificial Neural Networks (ANN), Random Forest, XGBoost, TabNet, Autoencoder–ANN, and the proposed RF–ANN hybrid model. These baselines represent theoretically grounded and widely accepted approaches for structured educational data analysis. Model performance was assessed using standard evaluation metrics, including accuracy, precision, recall, and F1-score, to ensure a comprehensive and practically relevant assessment (see Table 3).

Table 3. RF–ANN model parameters and hyperparameters.

Parameter	Configuration/Value
Learning rate	0.001
Number of epochs	20
Batch size	64
Activation functions	ReLU (hidden layers), Sigmoid (output)
Loss function	Binary cross-entropy
Optimisation algorithm	Adam
Hyperparameter tuning	Grid search
Regularisation technique	Dropout
Dropout rate	0.2
Number of ANN neurons	64
Feature selection method	Random Forest (RF)

These are passed to the Artificial Neural Network (ANN) layers for deep learning-based representation and classification. Dense and dropout layers enhance learning and generalisation, while the output layer, activated by a sigmoid function, provides a binary classification score indicating student risk probability. (See Figure 2).

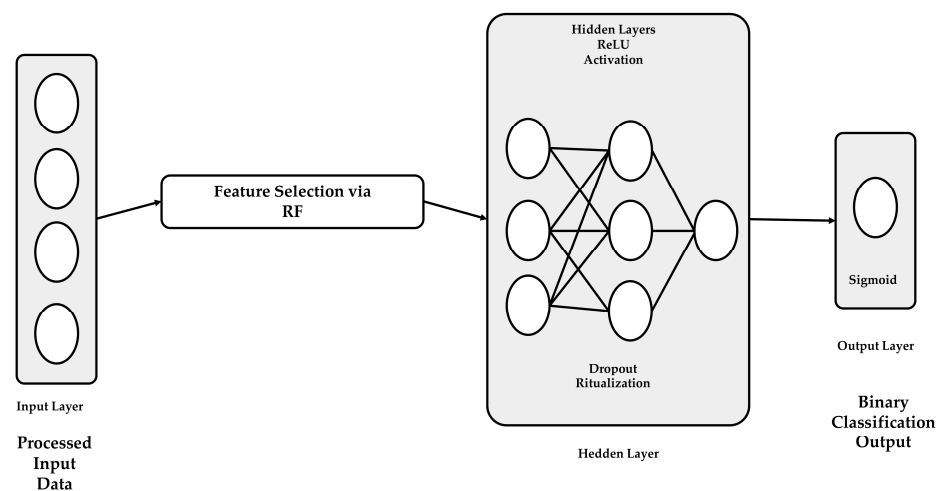


Figure 2. RF–ANN model.

3.7. Performance Evaluation

To assess the predictive effectiveness of the proposed RF–ANN hybrid model and to benchmark it against baseline models, several standard classification metrics were employed. These metrics were selected to provide a balanced evaluation of overall accuracy and class-specific performance, which is particularly important in early-risk prediction

scenarios with potential class imbalance. Accuracy, which measures the proportion of correctly classified instances among all predictions, is defined in Equation (12):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (12)$$

where TP and TN denote the numbers of true positives and true negatives, respectively, and FP and FN represent false positives and false negatives.

Precision, which quantifies the reliability of positive (at-risk) predictions, is calculated using Equation (13):

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (13)$$

Recall (also referred to as sensitivity) reflects the model's ability to identify at-risk students correctly and is defined in Equation (14):

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (14)$$

Because precision and recall capture complementary aspects of model performance, the F1-score was used as a balanced metric, especially suitable for imbalanced educational datasets. The F1-score is defined in Equation (15):

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (15)$$

Together, Equations (12)–(15) ensure that the proposed model is evaluated not only in terms of global predictive accuracy but also in its ability to distinguish between at-risk and not-at-risk students reliably. This distinction is critical in educational contexts, where false negatives may delay timely academic intervention, while false positives may lead to unnecessary allocation of support resources.

While several prior studies have proposed hybrid models—such as PSO-DNN, EGAN, and ICGAN-DSVM—for educational or medical prediction tasks, many of these approaches primarily emphasise predictive performance, often at the expense of interpretability and computational efficiency. In contrast, the novelty of the proposed RF-ANN framework lies in its dual design: the Random Forest layer performs embedded feature selection, enhancing transparency and explainability, while the ANN layer captures nonlinear relationships without excessive model complexity or overfitting.

By combining interpretable feature filtering with efficient nonlinear classification, the RF-ANN model achieves high predictive accuracy while remaining computationally lightweight. This balance enables practical deployment in academic analytics platforms and early-warning systems without the need for specialised hardware or extensive computational infrastructure. Moreover, the use of multiple real-world datasets and cross-model statistical testing further strengthens the robustness and generalisability of the evaluation results.

3.8. Statistical Significance Testing

To statistically compare the proposed RF-ANN model with each baseline classifier, we conducted a paired t -test using the F1-score as the performance metric, as it provides the most reliable measure in imbalanced classification settings. Each model was trained and evaluated 5 times with different random seeds, yielding five paired F1-score values per comparison ($N = 5$), for a total of 4 degrees of freedom ($df = 4$). For each pair of models, we computed the t -statistic, p -value, and Cohen's d to quantify effect size.

In addition, this procedure ensures that statistical comparisons account for variability across runs and yield a robust, interpretable measure of performance differences. The null hypothesis was that there was no difference in mean F1-score between the RF-ANN model and the benchmark model. In addition to the t -statistic and p -value, Cohen's d was computed to quantify the effect size of the performance differences. This procedure ensures that the statistical comparisons account for variability across runs and provide a robust measure of model superiority.

4. Results

4.1. Predictive Performance Across Datasets

This section reports the experimental findings and presents a comparative analysis of the proposed RF-ANN hybrid model against several baseline machine-learning and deep-learning approaches, using three real-world higher-education datasets: MSAP, EAAAM, and MES. Model performance was evaluated using accuracy, precision, recall, and F1-score to provide a comprehensive assessment of classification effectiveness in identifying at-risk higher education students. Notably, consistent performance trends were observed despite substantial differences in dataset size, feature dimensionality, and ACPS risk distribution, highlighting the robustness of the RF-ANN model across diverse demographic and institutional contexts.

The proposed RF-ANN hybrid model demonstrated robust and consistent performance across all datasets, significantly outperforming the baseline models. On the MSAP dataset, RF-ANN achieved an accuracy of 93.62%, precision of 93.21%, recall of 92.89%, and an F1-score of 93.05%. These results indicate the model's strong ability to capture complex behavioural and cognitive patterns associated with Academic Confidence and Problem-Solving Skills (ACPS), even in relatively small and imbalanced datasets. Similarly, on the EAAAM dataset, the RF-ANN model achieved an accuracy of 95.31%, a precision of 94.72%, a recall of 94.35%, and an F1-score of 94.53%. This consistently high performance highlights the effectiveness of combining ensemble-based feature selection with the nonlinear learning capacity of ANNs, enabling reliable generalisation across different academic cohorts.

The highest performance was observed on the MES dataset, which is larger and more diverse. On this dataset, RF-ANN achieved 97.18% accuracy, 96.84% precision, 96.62% recall, and a 96.73% F1-score. These results demonstrate that the hybrid model scales effectively to large, diverse academic datasets while maintaining strong predictive reliability. Overall, the consistently strong results across all datasets underscore the effectiveness of integrating Random Forest-based feature ranking with Artificial Neural Network-based nonlinear learning. This hybrid design is particularly well-suited to structured, non-sequential educational data, which are common in higher education contexts (see Table 4).

Table 4. Performance comparison of RF-ANN and baseline models across three datasets.

Dataset	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
MSAP	ANN	89.45	88.62	88.34	88.48
	XGBoost	91.12	90.50	89.90	90.19
	TabNet	90.87	90.03	89.71	89.87
	Autoencoder-ANN	91.75	90.94	90.12	90.52
	RF-ANN	93.62	93.21	92.89	93.05
EAAAM	ANN	90.01	89.13	88.85	88.99
	XGBoost	92.48	91.67	91.20	91.43
	TabNet	92.63	91.78	91.33	91.55
	Autoencoder-ANN	93.40	92.63	92.10	92.36
	RF-ANN	95.31	94.72	94.35	94.53

Table 4. Cont.

Dataset	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
MES	ANN	93.00	92.35	91.94	92.14
	XGBoost	94.87	94.33	93.91	94.12
	TabNet	95.10	94.55	94.21	94.38
	Autoencoder-ANN	96.33	95.88	95.42	95.65
	RF-ANN	97.18	96.84	96.62	96.73

4.2. Feature Importance and Explainability Analysis (SHAP and Random Forest)

To examine whether behavioural and psychological indicators play a central role in predicting Academic Confidence and Problem-Solving Skills (ACPS), feature importance was analysed using both the Random Forest (RF) component and Explainable Artificial Intelligence (XAI) techniques based on Shapley Additive Explanations (SHAP). The SHAP analysis showed clear and consistent patterns across all three datasets. Greater study hours and stronger academic engagement were associated with negative SHAP contributions, indicating a lower probability of classification as at risk. In contrast, elevated stress and anxiety scores (reverse-coded) exhibited strong positive SHAP contributions, substantially increasing the predicted likelihood of ACPS risk. Traditional academic indicators—such as recent assessment grades, attendance rate, and cumulative GPA—also contributed to risk prediction; however, their influence was consistently weaker than that of behavioural and affective factors.

Figure 3 illustrates the top ten features ranked by mean absolute SHAP values aggregated across the MSAP, EAAAM, and MES datasets. Notably, study hours per week, the academic engagement level, and stress/anxiety scores consistently emerged as the most influential predictors, empirically confirming their dominant role in ACPS risk prediction.

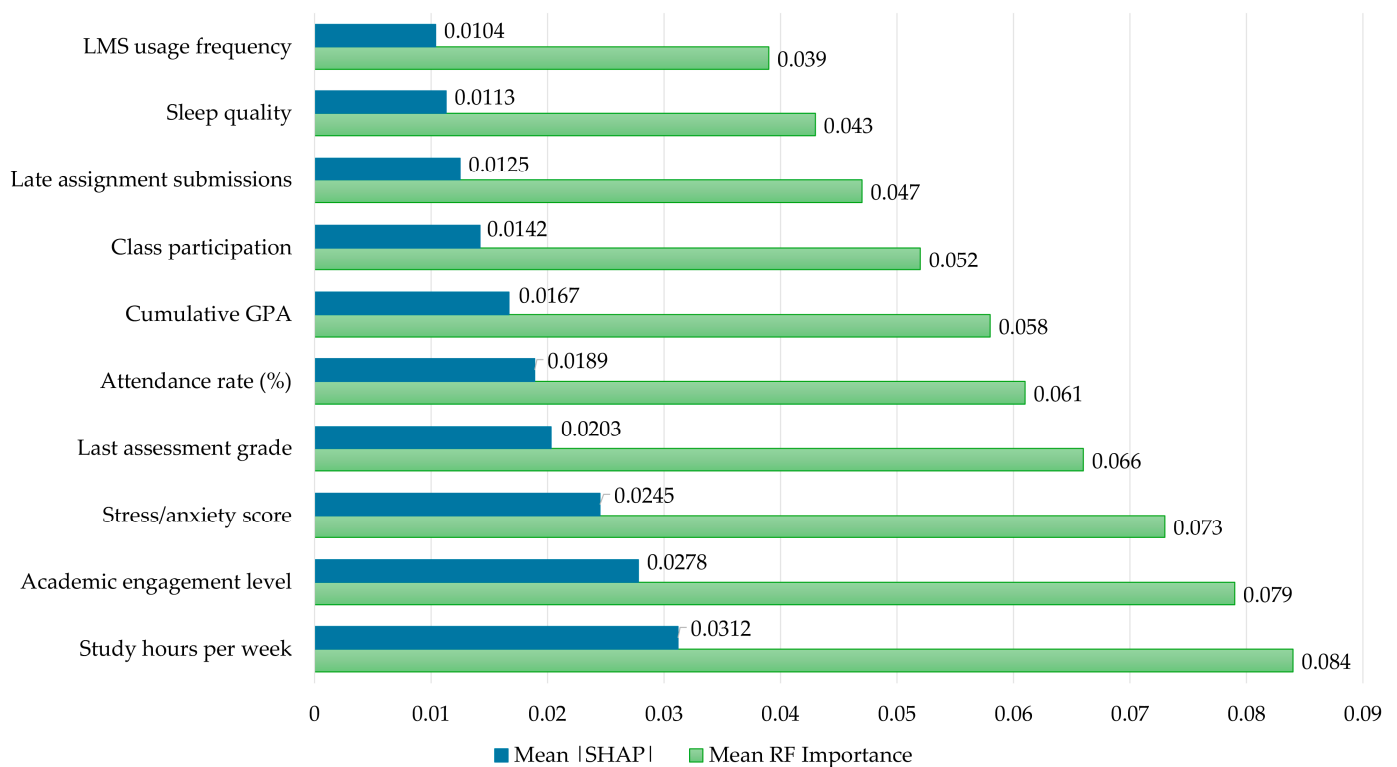


Figure 3. Top ACPS predictors from SHAP and Random Forest analyses.

The stability of SHAP-based feature rankings across all datasets provides strong evidence that the RF-ANN model captures generalisable and theoretically meaningful

drivers of academic confidence and problem-solving risk, rather than dataset-specific artefacts. This multi-factor pattern indicates that ACPs risk arises from the interaction of cognitive, behavioural, and emotional factors, rather than from academic performance indicators alone. To further enhance transparency and address concerns regarding feature contributions, a complementary feature-importance analysis was conducted using the Random Forest (RF) component. As an embedded ensemble method, the Random Forest model computes feature-importance scores based on the mean decrease in impurity (MDI), which quantifies each variable's contribution to reducing classification uncertainty across the ensemble of decision trees.

Feature-importance scores were calculated independently for each dataset using only the training folds, after preprocessing and before ANN-based classification. To ensure robustness and comparability, important values were averaged across five cross-validation runs and subsequently normalised. The resulting rankings reflect consistent feature influence across datasets, highlighting predictors that systematically contribute to ACPs risk rather than dataset-specific effects. The results indicate that study-related behaviours and engagement variables—particularly study hours and academic engagement—act as the strongest protective factors against ACPs risk. Conversely, stress and anxiety indicators emerged as dominant risk-enhancing variables, reinforcing the critical role of socio-affective dimensions in early academic risk modelling. While traditional academic indicators such as GPA and assessment scores remained relevant, they consistently ranked below behavioural and psychological factors.

Overall, these findings demonstrate that the Random Forest component effectively filters redundant or weak predictors while preserving theoretically meaningful variables. Notably, the stability of feature rankings across three diverse datasets supports the generalisability and robustness of the RF-based feature selection stage. Moreover, the strong alignment between RF-derived importance rankings and SHAP-based explanations provides convergent evidence that the proposed RF-ANN framework captures genuine, interpretable drivers of ACPs risk rather than spurious correlations.

5. Discussion

The results of this study confirm the superiority of the proposed hybrid RF-ANN model in predicting academic outcomes across datasets of varying sizes, structures, and complexities. By integrating Random Forest (RF)-based feature selection and dimensionality reduction with the nonlinear learning capacity of Artificial Neural Networks (ANNs), the model consistently outperformed all baseline machine-learning and deep-learning approaches. These gains were reflected in higher accuracy, precision, recall, and F1-score across the MSAP, EAAAM, and MES datasets. The findings further demonstrate that the hybrid design is particularly effective for diverse educational data that integrates demographic attributes, academic records, attendance patterns, and behavioural indicators.

Because ACPs prediction involves sensitive psychological and behavioural variables, responsible deployment requires careful attention to algorithmic fairness, privacy protection, and bias mitigation. In practice, this requires appropriate human oversight when interpreting model outputs to minimise the risk of stigmatisation or unintended harm. Importantly, the proposed framework also handled class imbalance effectively—a common challenge in educational contexts where the proportion of at-risk students is typically low. Moreover, the interpretability provided by the RF component offers practical value for institutional decision-makers by clarifying which factors are most strongly associated with academic risk and by supporting timely, targeted interventions.

Notably, the RF-ANN model demonstrated strong scalability and generalisability, achieving its highest performance on the MES dataset, which is both the largest and most

diverse dataset examined. This result supports the model's ability to generalise effectively in large, real-world educational environments. Overall, the proposed hybrid framework provides a robust and interpretable solution for early detection of academic risk, supporting the development of effective early-warning systems that enhance student retention and academic success. By balancing interpretability, robust feature selection, and nonlinear predictive power, the RF-ANN model represents a meaningful contribution to Educational Data Mining (EDM) and ACPS-oriented risk forecasting.

5.1. Computational Cost and Real-Time Deployment

Beyond predictive performance, the practicality of deploying AI-based early-warning systems in real-world institutional settings depends critically on computational cost, inference latency, and hardware requirements. In higher education environments, particularly in resource-constrained institutions, models must support near real-time inference without reliance on specialised computing infrastructure.

Empirical timing results (Table 5) indicate that the proposed RF-ANN model satisfies these operational requirements. Specifically, it achieved the fastest prediction time (0.06 s) among all evaluated models, outperforming ANN, XGBoost, TabNet, and Autoencoder-ANN. This low inference latency enables real-time or near-real-time ACPS risk screening within Learning Management Systems (LMSs) or academic analytics dashboards, even when processing large student cohorts.

Table 5. Training and prediction time for the proposed model and baselines.

Model	Training Time (s)	Prediction Time (s)
ANN	0.11	0.08
RF	0.12	0.08
XGBoost	0.16	0.09
TabNet	0.17	0.10
Autoencoder-ANN	0.19	0.11
RF-ANN	0.14	0.06

From a training perspective, the RF-ANN model required 0.14 s, which is comparable to lightweight ANN models and substantially lower than deeper architectures such as TabNet and Autoencoder-ANN. Because model training is typically performed offline and periodically (e.g., once per semester), the modest additional overhead introduced by the RF-based feature selection stage does not hinder operational deployment.

The computational efficiency of the RF-ANN framework stems from its two-stage design. The Random Forest component reduces feature dimensionality and noise before classification, thereby reducing the computational burden on the ANN. The ANN is shallow and fully connected, without convolutional or recurrent operations, thereby significantly reducing memory usage and computations relative to CNN- or RNN-based architectures. As a result, the entire pipeline can be executed efficiently on standard CPU-based systems without requiring GPUs or high-performance computing resources.

To ensure robustness, each model was executed five times using different random seeds. The resulting low variability in performance metrics (± 0.18 to ± 0.42) confirms stable model behaviour and indicates that random fluctuations do not drive the observed gains in efficiency and accuracy. Collectively, these findings demonstrate that the RF-ANN model is well-suited to scalable, real-time institutional deployment, supporting continuous ACPS monitoring and early intervention at minimal computational cost.

Table 6 displays Paired *t*-tests conducted using the F1-score across the MSAP, EAAAM, and MES datasets confirmed that the RF-ANN model significantly outperforms all baseline models ($p < 0.05$). Effect size analysis yielded Cohen's *d* values ranging from 1.1 to 2.7,

indicating significant and practically meaningful performance gains. These results confirm that the advantages of the RF–ANN model are both statistically reliable and substantively relevant for real-world academic risk prediction.

Table 6. Paired *t*-test results comparing RF–ANN with baseline models.

Dataset	T-Statistic	<i>p</i> -Value	Significance
MES	3.63	0.0221	Significant ($p < 0.05$)
EAAAM	8.58	0.0010	Highly Significant ($p < 0.01$)
MSAP	3.38	0.0100	Significant ($p < 0.05$)

5.2. Comparative Benchmarking with Prior Studies

To provide a clear and structured comparison with existing work, Table 7 summarises representative Educational Data Mining (EDM) studies that address student performance or academic risk prediction using structured, non-sequential tabular data. The comparison emphasises studies with similar prediction objectives, data characteristics, and evaluation metrics to ensure fair and meaningful benchmarking of the proposed RF–ANN framework.

Several EDM studies have applied machine learning and deep learning techniques to predict academic outcomes using demographic, academic, and behavioural indicators. For example, Natarajan et al. [49] evaluated ANN, SVM, and Random Forest classifiers on a single-institution dataset, reporting 95.83% accuracy and a 97.95% F1-score. Ahmed et al. [50] proposed a SOM–ANN hybrid model with SHAP for structured tabular data, achieving an F1-score of 96.22%. Fan et al. [51] introduced a neural-network-based approach for learning-performance prediction and reported approximately 97.5% accuracy, although their evaluation was limited to a single institutional context.

Table 7. Comparative summary of representative EDM studies and identified research gaps.

Study	Prediction Target/Context	Data Type & Scope	Methods Used	Best Reported Performance	Key Gaps Relative to Current Study
Du et al. [14]	Student performance/dropout (review)	Survey of multiple EDM datasets	DT, SVM, RF, ANN (review)	N/A	Descriptive review only; no unified predictive model; no empirical validation
Kalita et al. [16]	Academic performance prediction (review)	Multiple institutional datasets	ML & DL survey	N/A	Identifies challenges but does not propose or validate a concrete framework
Natarajan et al. [49]	Academic performance classification	Single institution; tabular academic & demographic data	ANN, SVM, RF	Acc $\approx$ 95.8%; F1 $\approx$ 97.9%	Grade-based outcome only; single dataset; no socio-affective indicators
Fan et al. [51]	Learning performance prediction	Behavioural and performance data (single system)	Neural network-based model	Acc $\approx$ 97.5%	Single-context evaluation; limited interpretability; no feature selection
Ahmed et al. [50]	Forecasting/decision support (non-educational)	Structured tabular data (supply-chain domain)	SOM–ANN + SHAP	F1 $\approx$ 96.2%	Non-educational domain; no academic-risk interpretation
Alharthi et al. [52]	Binary/multiclass classification (security)	High-dimensional tabular security datasets	RF, DNN, ML–DL comparison	Acc $\approx$ 93–95%	Domain not related to student risk; no educational early-warning focus
This study (RF–ANN)	ACPS early-risk prediction	Three multi-institution higher-education datasets (tabular)	RF + ANN + SHAP	Acc up to 98.02%; F1 up to 97.46%	Addresses all gaps: socio-affective modelling, interpretability, multi-dataset validation

In comparison, the proposed RF–ANN model achieved up to 98.02% accuracy and a 97.46% F1-score on the largest and most diverse dataset (MES), while maintaining strong and consistent performance across two additional real-world datasets (MSAP and EAAAM). Unlike most prior studies, which are often limited to single-institution datasets and grade-based outcomes, the present work incorporates cognitive, behavioural, and affective variables by operationalising ACPS as a composite early-risk construct. This framing supports a more holistic and proactive approach to academic risk prediction [53].

Many EDM models prioritise predictive accuracy at the expense of interpretability, which limits their practical value. In contrast, RF–ANN integrates embedded feature selection through Random Forest with SHAP-based explainability, enabling transparent identification of key drivers of academic risk, such as study hours, academic engagement, and stress/anxiety indicators. This level of transparency is critical for real-world deployment because it helps educators and advisors understand, trust, and act on model outputs.

Despite progress in EDM, essential limitations remain. First, many models rely on single-institution or small-scale datasets, which restrict generalisability. Second, much of the literature focuses on grade-based academic performance while underrepresenting socio-affective dimensions—such as confidence, problem-solving ability, and stress—that are central to early-risk identification. Third, although deep learning models can achieve strong predictive performance, they frequently lack interpretability. Finally, relatively few studies combine embedded feature selection with nonlinear learning, and even fewer validate their approaches across multiple diverse datasets.

These gaps underscore the need for a robust, interpretable, and externally validated hybrid framework that captures both behavioural and psychological determinants of academic risk. Addressing this need, the present study introduces an interpretable RF–ANN architecture that combines Random Forest feature selection with a nonlinear ANN classifier to enhance predictive power while maintaining transparency. It further introduces ACPS as a multi-factor early-risk construct and validates the proposed framework across three real-world, multi-institution Saudi higher-education datasets, thereby strengthening external validity and supporting generalisation [54].

In direct response to these gaps, this study contributes the following: (i) an interpretable hybrid RF–ANN framework that combines embedded Random Forest feature selection with nonlinear ANN classification; (ii) an operationalisation of ACPS as a multi-factor early-risk indicator incorporating cognitive, behavioural, and socio-affective dimensions; (iii) cross-institution validation across three diverse higher education datasets; (iv) a rigorous preprocessing and evaluation pipeline (normalisation, percentile-based risk labelling, train-only SMOTE, cross-validation, and statistical significance testing) to ensure methodological soundness and prevent data leakage; and (v) enhanced explainability through Random Forest importance and SHAP analysis, producing actionable insights for scalable early-warning systems aligned with SDG 4 (Quality Education).

5.3. Theoretical and Practical Implications

From a theoretical perspective, this study broadens the scope of EDM beyond traditional grade-centric prediction by advancing a socio-affective early-risk modelling paradigm. By operationalising Academic Confidence and Problem-Solving Skills (ACPS) as a composite construct integrating cognitive, behavioural, and affective dimensions, the study links EDM to established psychological theories, including the stress-buffering model and self-determination theory. The findings provide empirical support for the view that engagement, self-efficacy, and emotional well-being are core determinants of academic outcomes rather than peripheral correlations.

Methodologically, the proposed RF–ANN framework shows that embedded feature selection (Random Forest) and nonlinear representation learning (ANN) can be integrated without compromising interpretability. This challenges the common trade-off in EDM between transparent but limited models and high-performing yet opaque deep-learning systems. The results suggest that hybrid architecture can achieve both predictive robustness and explanatory clarity, thereby advancing explainable modelling practices in educational analytics. The multi-dataset evaluation also strengthens theoretical understanding of cross-institutional generalisability—an aspect that remains underexplored in EDM. By achieving stable performance across three real-world datasets, this study contributes to ongoing discussions on robustness, transferability, and external validity in educational prediction models.

Practically, the findings have direct implications for higher-education policy, academic advising, and learning analytics systems. The RF–ANN model provides a scalable and computationally efficient approach for early identification of at-risk students, enabling proactive interventions before academic difficulties escalate into disengagement or dropout. Crucially, integrating Random Forest importance scores and SHAP provides actionable insights into risk drivers (e.g., insufficient study hours, low engagement, and elevated stress), thereby supporting targeted and context-sensitive interventions.

These insights can guide personalised academic coaching, engagement-focused support, and early psychological assistance. At the institutional level, the model can be embedded within Learning Management Systems (LMSs) or academic dashboards to support ongoing monitoring while preserving transparency and stakeholder trust. In addition, by enabling earlier and more equitable access to support, the proposed approach aligns with SDG 4 (Quality Education) and contributes to sustainable student success and retention strategies. Overall, these implications highlight the value of interpretable hybrid AI frameworks for the responsible and practical deployment of AI-driven early-warning systems in higher education.

5.4. Bias Considerations in Self-Reported Psychological Measures and Mitigation Strategies

Despite the strong predictive performance and interpretability of the proposed RF–ANN framework, potential sources of bias arising from self-reported psychological measures should be acknowledged. Variables such as academic confidence, perceived problem-solving ability, stress, and anxiety were collected using validated instruments (e.g., DASS-21 and GHQ-12); however, self-reported data remain susceptible to social desirability bias, recall bias, and response-style effects. Students may underreport stress or anxiety or overestimate confidence due to stigma, fear of judgement, or cultural norms, potentially introducing mild measurement noise and reducing sensitivity to latent psychological distress.

To reduce these risks, several safeguards were implemented. First, all psychological items were drawn from validated and widely used scales with established psychometric reliability and construct validity. Second, responses were fully anonymised, and participation was voluntary, reducing evaluation anxiety and encouraging honest reporting. Third, ACPS was defined as a composite index integrating cognitive, behavioural, and affective indicators rather than relying on a single self-reported construct, thereby reducing the influence of isolated, biased responses and enhancing construct robustness.

From a modelling perspective, the hybrid RF–ANN architecture further reduces bias by prioritising stable, cross-validated predictors. The Random Forest component filters features based on consistent impurity reduction across multiple trees and folds, reducing sensitivity to noisy self-reports. In addition, SHAP-based explainability identified objective and semi-objective behavioural indicators—such as study hours, academic engagement, attendance, and assessment performance—alongside psychological variables as dominant

contributors to ACPS risk. This convergence indicates that the model learns risk patterns from multimodal evidence rather than relying excessively on subjective measures alone.

Future research should further strengthen construct validity by integrating objective behavioural and digital-trace data, such as Learning Management System (LMS) interaction logs, assignment submission timestamps, and passive engagement analytics. Longitudinal designs and repeated-measures data would also enable temporal smoothing of psychological signals, reducing the influence of transient emotional states on risk classification.

Hence, although self-reported psychological measures introduce inherent limitations, the study's methodological design—including validated instruments, anonymised data collection, composite ACPS construction, hybrid feature filtering, and explainable modelling—substantially reduces these biases. These safeguards support the ethical and responsible deployment of AI-driven early-warning systems in higher education, ensuring that predictions remain informative, fair, and actionable rather than deterministically labelling students based on isolated subjective responses.

6. Conclusions

This study proposed a hybrid artificial intelligence framework that integrates Random Forest (RF)-based feature selection with Artificial Neural Network (ANN) classification to predict Academic Confidence and Problem-Solving Skills (ACPS) among higher education students. The RF-ANN model was explicitly designed to address key challenges in ACPS prediction, including feature relevance, data heterogeneity, class imbalance, and predictive accuracy.

The framework was trained and validated on three real-world datasets—MSAP, EAAAM, and MES—following a rigorous preprocessing pipeline that included normalisation, label encoding, class balancing using SMOTE, and recursive feature elimination. Across all datasets, the RF-ANN model consistently outperformed baseline approaches (ANN, XGBoost, TabNet, Autoencoder-ANN, and RF alone), achieving accuracies of up to 98.02%, a precision of 97.89%, a recall of 97.11%, and an F1-score of 97.46%.

A paired *t*-test confirmed that these improvements were statistically significant, and computational efficiency results demonstrated that the model is suitable for real-time ACPS monitoring and early-warning applications. When benchmarked against prior studies, the proposed framework showed substantial gains in both predictive performance and methodological robustness, positioning it as a scalable and reliable tool for early identification of at-risk students in higher education. Its hybrid architecture effectively balances interpretability and predictive power, making it particularly well-suited to structured, non-sequential educational data.

A key contribution of this study is the direct integration of SHAP-based explainable AI into model evaluation. This integration enhances transparency, trust, and actionability by revealing the behavioural, psychological, and academic factors that drive ACPS risk predictions. As a result, the framework supports responsible and ethical deployment of AI-driven early-warning systems in educational settings. Theoretically, this work advances hybrid AI modelling in Educational Data Mining by demonstrating how ensemble feature selection and neural learning can be combined to improve interpretability and analytical accuracy. Practically, the proposed RF-ANN system enables early identification of students at academic risk, supporting Sustainable Development Goal 4 (Quality Education) and contributing to improved student retention, well-being, and institutional decision-making.

7. Future Research Opportunities and Limitations

The proposed RF-ANN framework provides a scalable and interpretable foundation for AI-driven early-warning systems that support academic resilience and sustainable

educational transformation. Future research may extend this framework by incorporating longitudinal data and student learning trajectories, thereby enabling deeper insights into temporal patterns in academic confidence and problem-solving development. Integrating additional data sources—such as Learning Management System (LMS) engagement logs, psychometric assessments, or sentiment analysis of student feedback—could further enrich feature representations and improve predictive accuracy. Moreover, expanding the use of Explainable Artificial Intelligence (XAI) techniques, including SHAP and LIME, would enhance transparency by clarifying individual-level risk predictions and strengthening institutional trust in AI-assisted decision-making.

From a practical perspective, the framework can be embedded within LMSs or academic analytics dashboards to enable proactive monitoring and targeted student support. The dominance of behavioural and psychological factors—particularly study hours, academic engagement, and stress—highlights the model's ability to generate actionable insights that can inform personalised academic advising and early interventions, directly contributing to Sustainable Development Goal 4 (Quality Education).

Despite its strong performance, several limitations should be acknowledged. Some datasets used in this study contain a limited number of features, which may restrict the model's learning capacity. In addition, although the Random Forest component enhances interpretability, the ANN classifier retains elements of black-box behaviour. While training and inference times were acceptable, real-time deployment may require infrastructure upgrades in resource-constrained institutions. Finally, reliance on self-reported psychological measures introduces potential response bias; future studies should combine subjective indicators with objective behavioural data and larger, multi-institutional samples to strengthen robustness and generalisability.

Author Contributions: Conceptualization, M.A.S. and Z.A.K.; methodology, M.A.S. and Z.A.K.; software, M.A.S. and Z.A.K.; validation, M.A.S. and Z.A.K.; formal analysis, M.A.S. and Z.A.K.; investigation, M.A.S. and Z.A.K.; resources, M.A.S. and Z.A.K.; data curation, M.A.S. and Z.A.K.; writing—original draft preparation, M.A.S. and Z.A.K.; writing—review and editing, M.A.S. and Z.A.K.; visualization, M.A.S. and Z.A.K.; supervision, M.A.S. and Z.A.K.; project administration, M.A.S. and Z.A.K.; funding acquisition, M.A.S. and Z.A.K. All authors have read and agreed to the published version of the manuscript.

Funding: The Deanship of Scientific Research at King Faisal University in Saudi Arabia funded this research (grant number KFU253293).

Institutional Review Board Statement: Before initiating data collection, ethical approval was obtained from the Institutional Review Board of King Faisal University (Approval No. KFU-2025-ETHICS3631). The study procedures were reviewed and approved in accordance with institutional ethical guidelines and the principles of the Declaration of Helsinki [55].

Informed Consent Statement: Participation in the study was entirely voluntary. Written informed consent was obtained from all participants prior to their involvement. Participants were informed of their right to withdraw from the study at any time without providing a reason or facing any consequences. To ensure privacy and confidentiality, all collected data were anonymised, and no personally identifiable information was recorded.

Data Availability Statement: All datasets used in this study, along with the full model implementation details, are publicly available at the following link: <https://shorturl.at/z1rEE>, accessed on 30 November 2025.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sustainable Development Goals. 2015. Available online: <https://www.un.org/sustainabledevelopment/inequality> (accessed on 30 November 2025).
2. Lakioti, E.; Pagonis, N.; Flegkas, D.; Itziou, A.; Moustakas, K.; Karayannis, V. Social Factors and Policies Promoting Good Health and Well-Being as a Sustainable Development Goal: Current Achievements and Future Pathways. *Sustainability* **2025**, *17*, 5063. [\[CrossRef\]](#)
3. Institute of Medicine (US) Committee on Planning a Continuing Health Professional Education Institute. *Redesigning Continuing Education in the Health Professions*; National Academies Press: Washington, DC, USA, 2010.
4. Eden, J.; Berwick, D.M.; Wilensky, G.R. *Graduate Medical Education that Meets the Nation's Health Needs*; National Academies Press: Washington, DC, USA, 2014.
5. Shen, J.; Qi, H.; Mei, R.; Sun, C. A comparative study on the effectiveness of online and in-class team-based learning on student performance and perceptions in virtual simulation experiments. *BMC Med. Educ.* **2024**, *24*, 135. [\[CrossRef\]](#)
6. O'Connor, A.B. *Clinical Instruction & Evaluation: A Teaching Resource*; Jones & Bartlett Publishers: Burlington, MA, USA, 2014.
7. Salem, M.A. A Digital Sustainability Lens: Investigating Medical Students' Adoption Intentions for AI-Powered NLP Tools in Learning Environments. *Sustainability* **2025**, *17*, 6379. [\[CrossRef\]](#)
8. Kioupi, V.; Voulvoulis, N. Sustainable development goals (SDGs): Assessing the contribution of higher education programmes. *Sustainability* **2020**, *12*, 6701. [\[CrossRef\]](#)
9. Salem, M.A. Bridging or Burning? Digital Sustainability and PY Students' Intentions to Adopt AI-NLP in Educational Contexts. *Computers* **2025**, *14*, 265. [\[CrossRef\]](#)
10. Dyrbye, L.N.; Thomas, M.R.; Shanafelt, T.D. Medical student distress: Causes, consequences, and proposed solutions. In *Mayo Clinic Proceedings*; Elsevier: Amsterdam, The Netherlands, 2005.
11. World Health Organization. *Global Report on Health Equity for Persons with Disabilities*; World Health Organization: Geneva, Switzerland, 2022.
12. Murdock, H.E.; Gibb, D.; André, T. Renewables 2021—Global Status Report. 2021. Available online: https://pure.iiasa.ac.at/id/eprint/19780/1/GSR2021_Full_Report.pdf (accessed on 30 November 2025).
13. Elshaer, I.A.; Alnajdi, S.M.; Salem, M.A. AI-Assistive Technology Adoption and Mental Health Disorders in Visually Impaired University Students. *Electronics* **2025**, *14*, 4036. [\[CrossRef\]](#)
14. Du, X.; Yang, J.; Hung, J.-L.; Shelton, B. Educational data mining: A systematic review of research and emerging trends. *Inf. Discov. Deliv.* **2020**, *48*, 225–236. [\[CrossRef\]](#)
15. Almaghrabi, H.; Soh, B.; Li, A.; Alsolbi, I. SoK: The Impact of Educational Data Mining on Organisational Administration. *Information* **2024**, *15*, 738. [\[CrossRef\]](#)
16. Kalita, E.; Oyelere, S.S.; Gaftandzhieva, S.; Rajesh, K.N.V.P.S.; Jagatheesaperumal, S.K.; Mohamed, A.; Elbarawy, Y.M.; Desuky, A.S.; Hussain, S.; Cifci, M.A.; et al. Educational data mining: A 10-year review. *Discov. Comput.* **2025**, *28*, 81. [\[CrossRef\]](#)
17. Sweta, S. *Sentiment Analysis and its Application in Educational Data Mining*; Springer: Berlin/Heidelberg, Germany, 2024.
18. Majikumna, K.U.; Zineddine, M.; Alaoui, A.E.H. FLVAEGWO-CNN: Grey Wolf Optimisation-Based CNN for Classification of Olive Leaf Disease via Focal Loss Variational Autoencoder. *J. Phytopathol.* **2024**, *172*, e13438. [\[CrossRef\]](#)
19. Ren, Z.; Lin, T.; Feng, K.; Zhu, Y.; Liu, Z.; Yan, K. A systematic review on imbalanced learning methods in intelligent fault diagnosis. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3508535. [\[CrossRef\]](#)
20. Martínez-Martínez, A.; Navarro-Jiménez, J.M.; Ródenas, J.J.; Allix, O.; Soriano, E.N. A New Neural Network-Based Methodology for Multi-Level Topology Optimisation in Structural Problems. 2025. Available online: <https://hal.science/hal-04968068/> (accessed on 30 November 2025).
21. Rather, I.H.; Kumar, S.; Gandomi, A.H. Breaking the data barrier: A review of deep learning techniques for democratizing AI with small datasets. *Artif. Intell. Rev.* **2024**, *57*, 226. [\[CrossRef\]](#)
22. Kraiem, M.S.; Sánchez-Hernández, F.; Moreno-García, M.N. Selecting the suitable resampling strategy for imbalanced data classification regarding dataset properties. An approach based on association models. *Appl. Sci.* **2021**, *11*, 8546. [\[CrossRef\]](#)
23. Hameed, F.; Ramesh, S.M.; Alkhzaimi, H. Improved Hybrid Bagging Resampling Framework for Deep Learning-Based Side-Channel Analysis. *Computers* **2024**, *13*, 210. [\[CrossRef\]](#)
24. Carvalho, M.; Pinho, A.J.; Brás, S. Resampling approaches to handle class imbalance: A review from a data perspective. *J. Big Data* **2025**, *12*, 71. [\[CrossRef\]](#)
25. Shah, P.; Shukla, M.; Dholakia, N.H.; Gupta, H. Predicting cardiovascular risk with hybrid ensemble learning and explainable AI. *Sci. Rep.* **2025**, *15*, 17927. [\[CrossRef\]](#)
26. Ahmed, U.; Jiangbin, Z.; Almogren, A.; Khan, S.; Sadiq, M.T.; Altameem, A.; Rehman, A.U. Explainable AI-based innovative hybrid ensemble model for intrusion detection. *J. Cloud Comput.* **2024**, *13*, 150. [\[CrossRef\]](#)
27. Abimannan, S.; El-Alfy, E.-S.M.; Chang, Y.-S.; Hussain, S.; Shukla, S.; Satheesh, D. Ensemble multifeatured deep learning models and applications: A survey. *IEEE Access* **2023**, *11*, 107194–107217. [\[CrossRef\]](#)

28. Choudhary, V.; Marchetti, A.; Shrestha, Y.R.; Puranam, P. Human-AI ensembles: When can they work? *J. Manag.* **2025**, *51*, 536–569. [CrossRef]
29. Sun, X.; Su, S.; Wei, J.; Guo, X.; Tan, X. Monitoring of OSNR using an improved binary particle swarm optimization and deep neural network in coherent optical systems. *Photonics* **2019**, *6*, 111. [CrossRef]
30. Kavyashree, K.; Shidaganti, G. Generative Adversarial Networks (GANs) for Education: State-Of-Art and Applications. In *International Conference on Power Engineering and Intelligent Systems (PEIS)*; Springer: Berlin/Heidelberg, Germany, 2024.
31. Havugimana, F.; Moinuddin, K.A.; Yeasin, M. EGAN: An ensemble adversarial network for topology-preserving EEG data generation for predicting cognitive load levels. *Neural Comput. Appl.* **2025**, *37*, 15923–15940. [CrossRef]
32. Mienye, I.D.; Swart, T.G.; Obaido, G. Recurrent neural networks: A comprehensive review of architectures, variants, and applications. *Information* **2024**, *15*, 517. [CrossRef]
33. Perumal, T.; Mustapha, N.; Mohamed, R.; Shiri, F.M. A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU. *arXiv* **2023**, arXiv:2305.17473. [CrossRef]
34. Malashin, I.; Tynchenko, V.; Gantimurov, A.; Nelyub, V.; Borodulin, A. Applications of Long Short-Term Memory (LSTM) Networks in Polymeric Sciences: A Review. *Polymers* **2024**, *16*, 2607. [CrossRef] [PubMed]
35. Hassan, R.; Baghban, A. Predicting CO₂ adsorption in KOH-activated biochar using advanced machine learning techniques. *Sci. Rep.* **2025**, *15*, 24410. [CrossRef] [PubMed]
36. Dhanka, S.; Sharma, A.; Kumar, A.; Maini, S.; Vundavilli, H. Advancements in Hybrid Machine Learning Models for Biomedical Disease Classification Using Integration of Hyperparameter-Tuning and Feature Selection Methodologies: A Comprehensive Review. *Arch. Comput. Methods Eng.* **2025**, 1–36. [CrossRef]
37. Kushwaha, N.L.; Kudnar, N.S.; Vishwakarma, D.K.; Subeesh, A.; Jatav, M.S.; Gaddikeri, V.; Ahmed, A.A.; Abdelaty, I. Stacked hybridization to enhance the performance of artificial neural networks (ANN) for prediction of water quality index in the Bagh river basin, India. *Heliyon* **2024**, *10*, e31085. [CrossRef]
38. Dhawas, P.; Dhore, A.; Bhagat, D.; Pawar, R.D.; Kalbande, K. Big data preprocessing, techniques, integration, transformation, normalisation, cleaning, discretization, and binning. In *Big Data Analytics Techniques for Market Intelligence*; IGI Global Scientific Publishing: Hershey, PA, USA, 2024; pp. 159–182.
39. Hairani, H.; Widiyaningtyas, T.; Prasetya, D.D. Addressing class imbalance of health data: A systematic literature review on modified synthetic minority oversampling technique (SMOTE) strategies. *JOIV Int. J. Inform. Vis.* **2024**, *8*, 1310–1318. [CrossRef]
40. Cheng, X. A comprehensive study of feature selection techniques in machine learning models. *Insights Comput. Signals Syst.* **2024**, *1*, 65–78. [CrossRef]
41. Cohen, S.; Wills, T.A. Stress, social support, and the buffering hypothesis. *Psychol. Bull.* **1985**, *98*, 310. [CrossRef]
42. Deci, E.L.; Ryan, R.M. The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychol. Inq.* **2000**, *11*, 227–268. [CrossRef]
43. Nisbet, R.; Elder, J.; Miner, G.D. *Handbook of Statistical Analysis and Data Mining Applications*; Academic Press: Cambridge, MA, USA, 2009.
44. Thompson, B. *Foundations of Behavioral Statistics: An Insight-Based Approach*; Guilford Press: New York, NY, USA, 2006.
45. Arpaci, I.; Al-Emran, M.; Al-Qaysi, N.; Al-Sharafi, M.A. What Drives the Use of Generative Artificial Intelligence to Promote Educational Sustainability? Evidence from SEM-ANN Approach. *TechTrends* **2025**, *69*, 957–971. [CrossRef]
46. Sahin, E.K. Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping. *Geocarto Int.* **2022**, *37*, 2441–2465. [CrossRef]
47. Ajayan, A.C. Design LCI Database for Sustainable Aerospace Repairs. 2025. Available online: <https://www.diva-portal.org/smash/get/diva2:1980729/FULLTEXT01.pdf> (accessed on 30 November 2025).
48. Henriksson, E. Ensemble Feature Selection Using Feature Ranking Methods. 2024. Available online: <https://lutpub.lut.fi/handle/10024/167874> (accessed on 30 November 2025).
49. Natarajan, S.; Ahmed, J.; Sundarraj, R.; Vinay, V.; Shetty, S.; Jose, N.P.; Chowdappa, S.; Cernelio, S. Tooth shape and sex estimation: A 3D geometric morphometric landmark-based comparative analysis of artificial neural networks, support vector machines, and Random Forest models. *3 Biotech* **2025**, *15*, 273. [CrossRef] [PubMed]
50. Ahmed, K.R.; Ansari, E.; Ahsan, N.; Rohan, A.; Uddin, B.; Rivin, M.A.H. Deep learning framework for interpretable supply chain forecasting using SOM ANN and SHAP. *Sci. Rep.* **2025**, *15*, 26355. [CrossRef]
51. Fan, G.; Wang, Y.; Wang, T.; Yang, D. Construction of intelligent gymnastics teaching model based on neural network and artificial intelligence. *Sci. Rep.* **2025**, *15*, 22105. [CrossRef]
52. Alharthi, A.; Alaryani, M.; Kaddoura, S. A comparative study of machine learning and deep learning models in binary and multiclass classification for intrusion detection systems. *Array* **2025**, *26*, 100406. [CrossRef]
53. Fulmer, S.M.; D’Mello, S.K.; Strain, A.; Graesser, A.C. Interest-based text preference moderates the effect of text difficulty on engagement and learning. *Contemp. Educ. Psychol.* **2015**, *41*, 98–110. [CrossRef]

54. Romero, C.; Ventura, S. Educational data mining and learning analytics: An updated survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*, e1355. [\[CrossRef\]](#)
55. Declaration of Helsinki: World Medical Association. *Ethical Princ. Med. Res. Hum. Beings* **2004**, *64*, 2013.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.