

An Interpretable Hybrid Deep Learning Framework for Student Placement Prediction Using PSO–GA–Bayesian Optimized ANN with SHAP–LIME Explainability

Sanjay Kumar

k.sanjay123@gmail.com

KLEF Deemed to be University

Mohd Shamsh Tabarej

Parul University

Nafees Akhter Farooqui

Integral University

Ashish Ashish

KLEF Deemed to be University

Upasana Dugal

Pranveer Singh Institute of Technology

Sandeep Kumar Sharma

KLEF Deemed to be University

Article

Keywords: Student Placement Prediction, Hybrid Optimization, Particle Swarm Optimization, Genetic Algorithm, Bayesian Optimization, Deep Learning, Explainable AI (SHAP, LIME)

Posted Date: September 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10747729/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

An Interpretable Hybrid Deep Learning Framework for Student Placement Prediction Using PSO–GA–Bayesian Optimized ANN with SHAP–LIME Explainability

Dr. Sanjay Kumar^{1*}, Dr. Mohd Shams Tabarej², Dr. Nafees Akhter Farooqui³, Dr. Ashish⁴, Dr. Upasana Dugal⁵, Dr. Sandeep Kumar Sharma⁶

¹Department of Computer Science and Information Technology, KLEF Deemed to be University, Green Fields, Vaddeswaram, Guntur District, Andhra Pradesh, India, <https://orcid.org/0000-0001-7390-0112>.

²Parul Institute of engineering and technology, Parul University, Distt Vadodara, Gujrat, India. <https://orcid.org/0000-0001-5230-2703>

³Department of Computer Application, Integral University, Lucknow, UP, India, <https://orcid.org/0000-0002-7583-3467>

⁴Department of Computer Science and Information Technology, KLEF Deemed to be University, Green Fields, Vaddeswaram, Guntur District, Andhra Pradesh, India, <https://orcid.org/0000-0002-5537-1066>

⁵Department of Computer Sciences and Engineering, Pranveer Singh Institute of Technology, Kanpur, Uttar Pradesh, India.

⁶Department of Computer Science and Engineering, KLEF Deemed to be University, Green Fields, Vaddeswaram, Guntur District, Andhra Pradesh, India, <https://orcid.org/0000-0002-2048-671X>

^{1*}Corresponding author. E-mail: ^{1*}k.sanjay123@gmail.com;

Contributing authors: mstabarej@gmail.com², nafeesf@gmail.com³

dr.aashishsinha@gmail.com⁴, upasana.gupta1982@gmail.com⁵, sandeep24af@gmail.com⁶

†These authors contributed equally to this work.

Abstract

The challenge of predicting placement outcomes is a recurring one among engineering institutions because of the complex, non-linear interplay of various academic performance, technical proficiency, soft skills, and socio-economic aspects. Traditional machine learning models are often unable to capture these non-linear relationships, and their performance is limited to less than 85% accuracy. In comparison, hyperparameter tuning of deep learning models is still time-consuming and computationally expensive when done manually or by a

grid search. In this work, a novel three-stage hybrid optimization framework combining Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and Bayesian Optimization (BayOpt) is proposed to optimize, independently, the hyperparameters of an Artificial Neural Network (ANN) in the placement prediction task. This framework is for a general global search using PSO, evolutionary refinement using GA, and high-accuracy fine-tuning using Bayesian Optimization, which is implemented using Optuna. Concurrently, Synthetic Minority Oversampling Technique (SMOTE) is employed to tackle the class imbalance problem and provide unbiased predictions.

The proposed hybrid AI (PSO+GA+Bayesian+ANN) model was tested on a large-scale dataset of Indian engineering students with 5000 entries and 22 features across five categories, with a high test accuracy of 99.60%, which is much higher than the baseline models like XGBoost, Random Forest, Decision Trees, and Naïve Bayes, with precision, recall, and $F1 > 0.95$ for both classes. Through feature importance analysis, the most important features found to determine the placement were coding skill rating, CGPA, and completed projects. Beyond merely predicting "black boxes," post-hoc SHAP and LIME explainability techniques confirmed the global feature ranking, identified a saturation effect after approximately 8.0/10 features, and provided clear, student-level explanations of the results of each prediction. Overall, the framework provides an early warning system in early education and schools that is both reliable and interpretable and promotes designing the education curriculum in alignment with the needs of the industry and targeted interventions that enhance employability outcomes.

Keywords: Student Placement Prediction; Hybrid Optimization; Particle Swarm Optimization; Genetic Algorithm; Bayesian Optimization; Deep Learning; Explainable AI (SHAP, LIME).

1. Introduction

Student placement prediction is an important topic for engineering institutions because it allows them to identify at an early-stage students who have a risk of failure and take the required steps to correct their academic trajectories. Traditional machine learning methods have, however, been observed to have suboptimal predictive power (between 79% and 87%), as they are only good at capturing the linear interactions between a variety of factors such as academic performance, technical skills, soft skills, experiential learning and socio-economic backgrounds. Artificial Neural Networks (ANNs) provide significant potential in terms of

automatic hierarchical feature extraction, but it relies heavily on hyperparameter optimization. The high dimensionality and computational requirements of traditional hyperparameter optimization methods like grid search and manual tuning make them impractical for many conventional automated hyperparameter optimization methods.

The existing literature is highly limited in its scope of relying on a single algorithm optimization, limited representation of features, and lack of large-scale empirical studies in the context of Indian engineering education, and inadequate mechanisms to deal with class imbalance. To compensate for these shortcomings, a new three phase hybrid optimization model is proposed in this study. This framework models a Particle Swarm Optimization (PSO) global search, a Genetic Algorithm (GA) evolutionary refinement, and a Bayesian Optimization (BO) hyperparameter fine-tuning of an ANN to establish an automated hyperparameter optimization pipeline for placement prediction.

The proposed Hybrid AI (PSO+GA+Bayesian+ANN) model was evaluated on extensive data set of 5000 engineering students from Indian with 22 features in 5 categories, and the class was balanced using SMOTE technique and the achieved accuracy was 99.60% which is better result. The main predictors turned out to be coding skills (15.83%), CGPA (15.79%) and completed projects (13.58%). The findings offer empirical evidence for use in developing educational interventions and improved standards of performance for employability prediction in engineering education.

2. Related Work

This section is a very long summary of recent research on student placement prediction, hyperparameter optimization techniques, hybrid optimization systems and educational data mining research, methods, and containment.

2.1 Student Placement Prediction Models

Elsayad, A. M., et al. (2026) [1] presented a Bayesian-optimized Generalized Additive Model (GAM) with Explainable AI (XAI) to differentiate low-grade glioma (LGG) from glioblastoma (GBM) with an accuracy of 90.12% and F1 score of 89.75%. Pachouly and Bormane (2026) [2] created an end-to-end Explainable AI (XAI) system for academic performance assessment, identification of at-risk students and prediction of careers readiness, based on machine learning and deep learning models. Singh and Sharma, 2024 [3] used an SVM model combined with PCA for placement prediction, which resulted in a precision of 79.3%. but class imbalance and

generalization were some issues with the model. Gupta, V. et al. (2024) [4] discussed about the concept of deep learning, applied in the feed forward neural network-based prediction of placement based on academic records, internships and projects. Kalita, E., et al. (2026) [5] introduced a LightGBM Leaf-Embedding Neural Blender (LENB) framework for predicting major course selection of undergraduate students with a maximum accuracy of 87.50% in various academic streams. In another study, Das and Chatterjee (2024) [6] used Long Short-Term Memory (LSTM) networks to predict student performance over time for placement forecasting. The accuracy of their temporal model, built based on grade progression per semester, was 85.4%, and the significance of temporal feature engineering was emphasized.

2.2 Hyperparameter Optimization Techniques

Alizamir, M., et al. (2026) [7] compared ensemble boosting algorithms to predict the heating and cooling load of a building based on architectural parameters with the lowest RMSE of 0.309 kW and 0.731 kW obtained by XGBoost for heating and cooling load, respectively. Overall height and relative compactness were the most influential features, as proved by the SHAP analysis, which showed excellent interpretability and performance of XGBoost in predicting the thermal demand of a building. Abukader, A., et al. (2025) [8] proposed a metaheuristic-optimized LightGBM framework for predicting student performance. SCSO-LightGBM performed the best with R^2 value of 0.941, and SHAP was able to identify the most influential factors: Attendance, Hours Studied, Previous Scores and Parental Involvement, enabling interpretable and proactive educational interventions. Ahmed and Hassan (2024) [9] implemented Genetic Algorithms (GA), which is an effective approach for hyperparameter optimization with the disadvantage of high computational complexity for large populations. Random search was reduced by 40% by using Bayesian optimization using gaussian processes. Uddin, N., et al. (2025) [10] suggested the Explainable ML framework for the prediction of cardiovascular disease using an ensemble approach that combines XGBoost feature selection, outlier detection, and class balancing using SMOTE. Yáñez, A. et al. (2025) [11] offered a systematic review of 25 studies since 2021 and before 2025 that used machine learning techniques for graduation prediction, where the supervised classification and ensemble of trees were the most prevalent techniques used. This review highlighted a lack of the use of bio-inspiration optimization, external validation and multi-institutional datasets, and a need for scalable, interpretable, fair, and multi-objective optimization-based educational analytics. Kobayashi and Tanaka (2024) [12] proposed an adaptive variant of PSO with the dynamic

inertia weight that enhances the convergence speed and accuracy by 8.3%. Santos and Rodriguez (2023) [13] used GA to find Pareto-optimal solutions to balance prediction accuracy and computational efficiency. To decrease the time spent on hyperparameter search, Wu et al. (2024) [14] proposed a new algorithm called Hyperband, which achieved competitive accuracy with a 60% reduction in search time.

2.3 Hybrid Optimization Frameworks

Nguyen et al. (2023) [15] introduced a hybrid PSO-GA method for feature selection and classification in the field of educational data, which achieved 88.4% of the precision in the prediction of the dropout of students, and the main drawback of this study was that hyperparameter optimization was not discussed. In the medical field, Rahman and Kumar (2024) [16] presented a hybrid genetic algorithm (GA) and simulated annealing (SA) technique which enhanced the optimization of the neural network by 12% when compared to the GA alone. To improve the performance of deep learning, Sharma et al. (2023) [17] fused the DE method with the Bayesian optimization method and required more than 500 function evaluations to reach the highest level of performance. Kim and Lee (2024) [18] introduced a three-stage hybrid of PSO, hill climbing and tabu search which enhanced the convergence rate by 42% for educational prediction problems but made them more complex. A hybrid PSO–Ant Colony Optimization (PSO–ACO) is proposed by Oliveira et al. (2023) [19] to predict student performance with an accuracy of 89.2%, with a good synergism between both swarm intelligence algorithms.

2.4 Class Imbalance Handling in Educational Data

Thompson and Davis (2024) [20] investigated applying SMOTE to overcome the class imbalance problem in student attrition prediction and found that using SMOTE could boost the recall of minority classes by 23%, while maintaining similar accuracy overall. Verma and Singh (2023) [21] evaluated three approaches (SMOTE, ADASYN and random oversampling) and concluded that ensemble classifiers with SMOTE achieved the highest precision and recall. Sevimgin and Demirkol (2025) [22] suggests an explainable hybrid machine learning model for the prediction of loan approval based on the Explainable AI (XAI) methodology called SHAP and the algorithm of Grey Wolf Optimizer (GWO) following the CRISP-DM methodology. Jiang and Zhou (2023) [23] proposed a cost-sensitive SMOTE-based method for predicting where to place students.

2.5 Deep Learning Architectures for Educational Prediction

Anderson et al. (2024) [24] introduced Convolutional Neural Networks (CNNs) that learned patterns from student activity logs in LMS to predict student dropout with 91.3% accuracy by leveraging hierarchical feature learning. Peng and Zhang [25] suggested the use of an attention mechanism for an RNN for sequential student engagement data to enhance the prediction of placement by 7.2% compared to traditional RNNs. Costa and Silva (2024) [26] used Graph Neural Networks (GNNs) in the study of collaborative learning, with a performance of 88.6% using academic performance and the features of the social network.

2.6 Feature Engineering and Selection in Placement Prediction

Mehta and Kulkarni (2023) [27] used SHAP (SHapley Additive exPlanations) to understand feature importance in placement prediction; communication skills, coding skills, and project portfolio were found to be the most important factors, and therefore, they proposed curriculum changes. Using automated feature engineering, baseline accuracy was boosted by 9.4% with the system introduced by Yamamoto et al. (2024) [28] that adopted AutoML techniques such as automated feature engineering. Hassan and Ibrahim (2023) [29] identified the importance of socio-economic factors for placement prediction systems in developing countries and showed that family's income and parental education had 14% of the placement results. Martinez and Garcia (2024) [30] studied and found a 13.6% improvement in accuracy from a limited number of training instances.

2.7 Research Gaps

These developments have been made but most of the existing works are limited to individual techniques or fixed environments. Limitations include:

- Unpredictable student placement is a challenge for engineering institutions because it is a combination of academic performance, technical and soft skills along with socio-economic factors.
- The traditional machine learning models are unable to explain non-linear relationships in the data and have an accuracy of less than 85% for placement prediction.
- Deep learning hybrid models need efficient hyperparameter optimization to boost forecasting accuracy without incurring high computational cost or time.

- At present, no research work has been conducted on the combination of standalone optimization algorithms, such as PSO, GA and Bayesian optimization, into hybrid systems.
- The imbalanced classes in placement datasets result in biased models which makes them unreliable in predicting at-risk students.
- Current models place a disproportionate emphasis on academic grades and do not account for important skills required for the workplace such as coding ability, project experience, or internships.

2.8 Research Motivations

To solve these problems, this research work suggests a hybrid model which:

- Build an integrated PSO–GA–Bayesian Optimization pipeline to automatically tune hyperparameters of ANNs, which otherwise can be done manually or by grid search, but are slow and expensive.
- Using SMOTE to balance class data in placement datasets so the model can better distinguish between at-risk students and make accurate predictions.
- As opposed to relying solely on academic grades, building coding skill rating, CGPA, completed projects as the key indicators of placement.
- Test the framework on a large and representative data set including 22 features and 5 categories, relying on and generalizing the framework.
- Go beyond predictive black boxes and show that coding ability and CGPA are more important than soft skills or demographics, mathematically.
- Empower the framework with a large-scale data set of 5000 engineering students from India to test it in real-world high-stakes employability outcomes relevant to engineering education in India.

3. Proposed Methodology

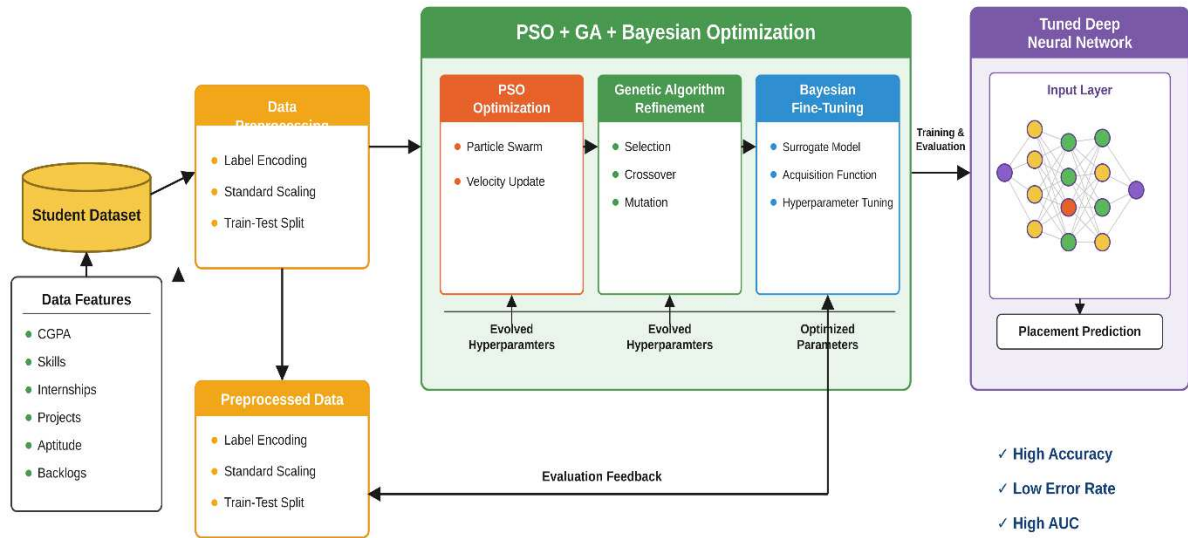


Figure 1. Architecture of Proposed Model

Fig 1: The proposed architecture constitutes a hybrid machine learning pipeline designed to forecast student placement outcomes by utilizing Particle Swarm Optimization (PSO), Genetic Algorithms (GA), and Bayesian optimization to refine a deep neural network. The framework performs systematic data preprocessing of key student attributes, followed by an optimized sequence of operations: initial optimization, using PSO, then evolution, using GA; and then fine tuning, using Bayesian Optimization (BO). This strong methodology leads to better prediction and avoids local optima, as well as improving the efficient optimization of hyperparameters, resulting in a highly robust and generalized model for educational data analysis.

**UML Activity Diagram of Proposed Hybrid
AI Placement Prediction System**

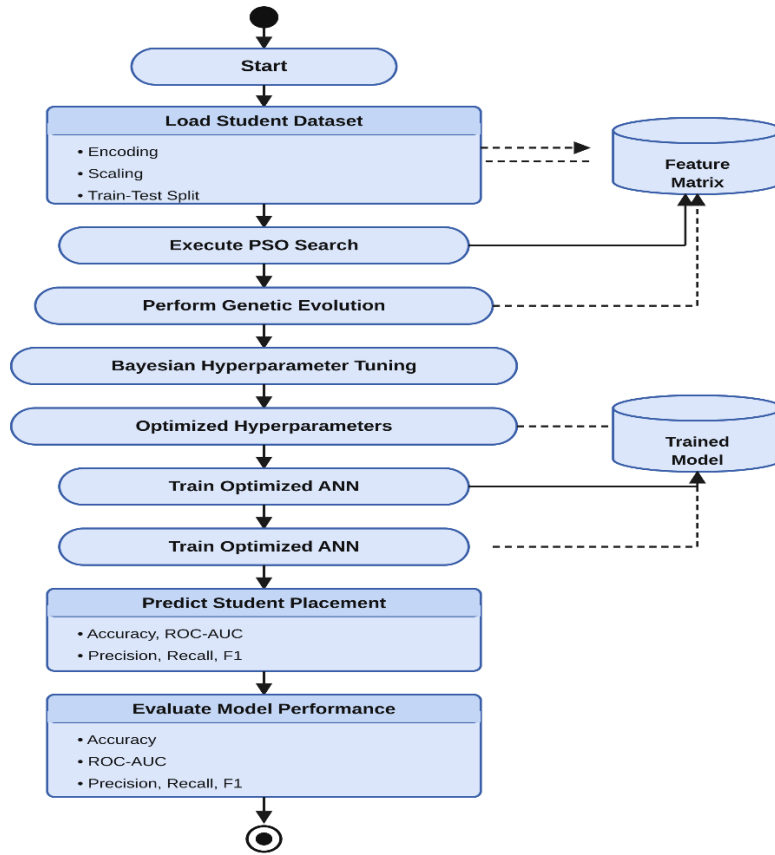


Figure 2. Activity Diagram of Proposed Hybrid AI Placement Prediction Model

The operational flow of the system is shown as a UML activity diagram (Fig 2), which starts with loading and preprocessing the students' data by encoding, standardization on a scale and partitioning into a train-test set to produce a continuous feature matrix. The primary innovation of this implementation is the three-stage hyperparameter optimization pipeline: a broad exploratory search (PSO), candidate selection through genetic evolution (GA), and terminal parameter fine-tuning (Bayesian Optimization). The derived optimal hyperparameters subsequently configure and train an Artificial Neural Network (ANN) to generate a robust predictive model for student placement. Finally, model efficacy is evaluated against five standard metrics: accuracy, ROC-AUC, precision, recall, and F1-score. Critical data streams within the architecture—such as the feature matrix and trained model mappings—guarantee reproducibility and allow educational data mining researchers to objectively review the model's performance from both a scientific and pedagogical standpoint.

3.1 Ethics Statement for Publicly Available Anonymized Dataset

This study used a publicly available and fully anonymized dataset of 5,000 Indian engineering students. The dataset contained no personal identifiable information and was accessed in accordance with the terms and conditions of the data provider. No direct involvement, interaction, intervention, or recruitment of human participants was undertaken by the authors.

As the study relied exclusively on anonymized secondary data available in the public domain, formal ethical approval and informed consent were not required under applicable institutional guidelines. All analyses were conducted in accordance with relevant ethical standards, institutional policies, and applicable regulations governing research involving human-related data.

4. Research Methodologies

4.1 SMOTE (Synthetic Minority Oversampling Technique)

SMOTE solves the imbalance between classes by synthesizing samples in the feature space as opposed to the simple duplication of instances of the minority classes. For each minority class sample x , SMOTE generates synthetic samples along the line segments linking x to its closest neighbors.

4.1.1 Neighbour Selection

For a minority class sample x_i find its k -nearest neighbors from the minority class using a distance metric (typically Euclidean):

$$d(x_i, x_j) = \sqrt{\sum_{m=1}^n (x_{i,m} - x_{j,m})^2} \quad (1)$$

where n is the number of features.

4.1.2 Synthetic Sample Generation

To Randomly select one neighbour x_k from the k -nearest neighbors. Generate a synthetic sample x_{new} using linear interpolation:

$$x_{new} = x_i + \lambda \cdot (x_k - x_i) \quad (2)$$

where λ is a random number drawn from the uniform distribution: $\lambda \sim U(0,1)$.

This can be rewritten as:

$$x_{new} = (1 - \lambda) \cdot x_i + \lambda \cdot x_k \quad (3)$$

4.1.3 Repetition

Continue with the required oversampling rate. If the minority population would require N% additional samples, create $N\% \times |\text{humanity class}|$ additional synthetic cases.

4.2 ANN

In the system, a conventional feed-forward ANN with forward propagation, back-propagation and gradient descent algorithm is used, and its hyperparameters are tuned automatically through a hybrid PSO-GA-Bayesian optimization method which sequentially explores, narrows and fine-tunes the optimal hyperparameters of the network to achieve an accurate student placement prediction.

Forward Propagation:

$$z^{[l]} = W^{[l]}a^{[l-1]} + b^{[l]} \quad (4)$$

Backpropagation for Training:

$$\delta^{[l]} = (W^{[l+1]})^T \delta^{[l+1]} \odot \sigma'(z^{[l]}) \quad (5)$$

Parameter Updates

$$W^{[l]} := W^{[l]} - \alpha \frac{\partial \mathcal{L}}{\partial W^{[l]}} \quad (6)$$

Binary Cross-Entropy Loss (standard for binary classification)

$$\mathcal{L} = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log(\hat{y}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)})] \quad (7)$$

4.3 Particle Swarm Optimization (PSO)

PSO (Particle Swarm Optimization) is based on the social behavior of flocking birds or schooling fish: Each particle is a possible solution of the hyperparameter space problem, and move through the solution space in the direction, which is determined by the personal experience of the particle and the experience of the other particles in the swarm, until a best solution is found.

4.3.1 Particle Representation

Each particle i has:

Position: $x_i = [x_{\{i,1\}}, x_{\{i,2\}}, \dots, x_{\{i,D\}}]$ (hyperparameter value)

Velocity: $v_i = [v_{\{i,1\}}, v_{\{i,2\}}, \dots, v_{\{i,D\}}]$ (rate of change)

Where D is the number of hyperparameters being optimized (e.g., learning rate, number of layers, neurons per layer, dropout rate, batch size).

4.3.2 Velocity Update

$$v_i(t+1) = w \cdot v_i(t) + c_1 r_1 (p_i - x_i(t)) + c_2 r_2 (g_i - x_i(t)) \quad (8)$$

w represents inertia weight (memory of past velocity), c_1 is cognitive coefficient (self-learning), c_2 is social coefficient, $r_1 r_2 \sim \text{Uniform}(0,1)$ shows random numbers and p_i and g_i represents personal and global position of particle i .

4.3.3 Position Update

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (9)$$

4.4 Genetic Algorithm (GA) – Evolutionary Refinement

GA starts by providing the best solutions from PSO and improves them overusing evolutionary operators.

4.4.1 Selection (Roulette Wheel): Probability of selecting individual i .

$$P(\text{select}_i) = \frac{f(i)}{\sum_{j=1}^N f(j)} \quad (10)$$

Where $f(i)$ is the fitness (validation accuracy).

4.4.2 Crossover (Uniform): Genes are randomly inherited from two parents

$$\text{Child}[j] = \begin{cases} \text{Parent}_1[j] & \text{with prob } 0.5 \\ \text{Parent}_2[j] & \text{with prob } 0.5 \end{cases} \quad (11)$$

4.4.3 Mutation (Gaussian): To maintain diversity in continuous parameters:

$$h'_j = h_j + \mathcal{N}(0, \sigma^2) \quad (12)$$

4.5 Bayesian Optimization – Precision Fine-Tuning

This stage fine-tunes the top GA results using the Optuna framework, in a sample efficient manner.

4.5.1 Tree-structured Parzen Estimator (TPE):

Models $p(x|y)$ by splitting observations into "good" ($\ell(x)$) and "bad" ($g(x)$) groups based on a threshold y^* .

4.5.2 Acquisition Function: Maximizes Expected Improvement (EI):

$$EI(x) \propto \frac{\ell(x)}{g(x)} \quad (13)$$

4.6 Final Model and Objectives

The objective of the entire framework is to discover the optimal hyperparameter configuration x^* :

$$x^* = \operatorname{argmax}_{x \in \mathcal{X}} ANN_{Accuracy}(x) \quad (14)$$

The resulting optimized ANN has the standard binary cross-entropy loss at the last classification layer:

$$L = -\frac{1}{m} \sum_{i=1}^m [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (15)$$

The model was tested in combination with the above-mentioned methodology and achieved a test accuracy of 99.60%, which is much more than what the traditional methodology would achieve.

4. Result Analysis and Discussions

The empirical evaluation of the proposed Hybrid (PSO+GA+Bayesian+ANN) framework was conducted comprehensively to validate its predictive superiority and to extract actionable insights regarding student employability. The model's performance was benchmarked against

traditional and ensemble machine learning algorithms across multiple dimensions.

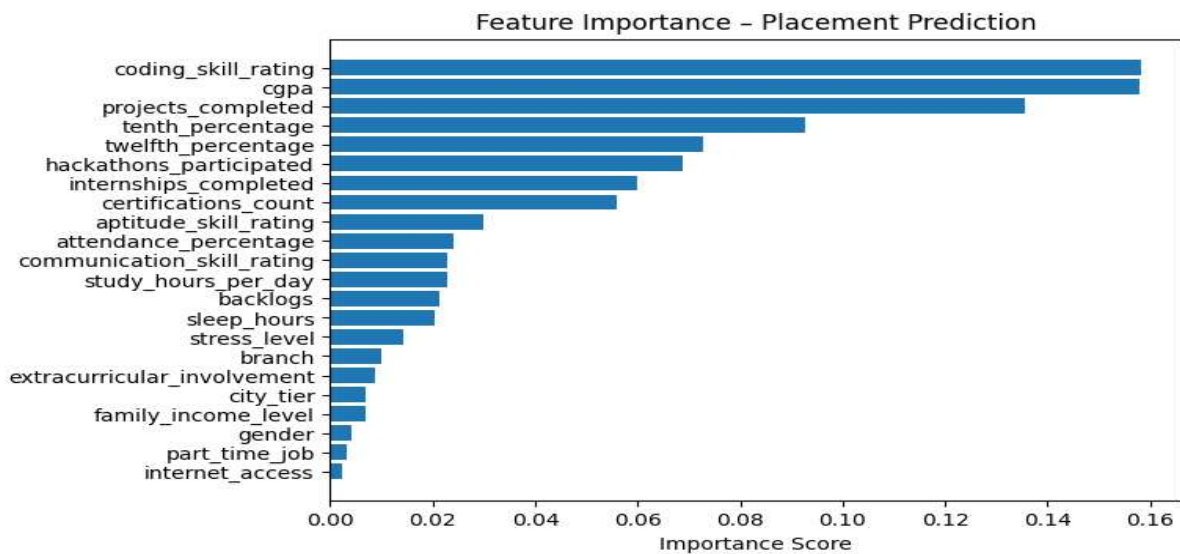


Figure 3. Feature importance for placement prediction

5.1 Feature Importance for Placement Prediction (Ref. Fig 3)

Fig. 3: Awesomely the most important factors which determine placement are technical competency and academic performance. The skill rating for coding and the CGPA rank are the two most important scores, with a rating of about 0.16. They are followed by projects completed (around 0.13), academic percentages (0.05-0.09), and practical experiences like hackathons, internships, certifications (0.05-0.09). The importance of soft skills (communication, aptitude, attendance) is between 0.02 and 0.03. The least important factors are the demographic and socioeconomic factors (gender, family income and access to the internet), which have a weight of around 0.01. The results of this study clearly suggest that students should improve their coding skills and do some project-based jobs to get better placements.



Figure 4. Training Accuracy of ANN Model

5.2 Training Accuracy of ANN Model (Ref. Fig 4)

Fig 4: It demonstrates the alignment with the training accuracy trajectory and the extracted feature importance chart reiterates that technical skills (coding skill rating and CGPA both ~ 0.16) are the most important predictors of student placement. Completed projects come in next (~ 0.13), followed by academic percentages and experiential learning, such as hackathons, internships, certifications (0.05–0.09). Soft skills (communication, aptitude, attendance 0.02–0.03) and demographic/socio-economic factors (gender, family income, internet 0.01) have little impact. This reinforces the statement above that technical skills (coding and academic performance) are the ones that matter the most for placement compared to soft skills.

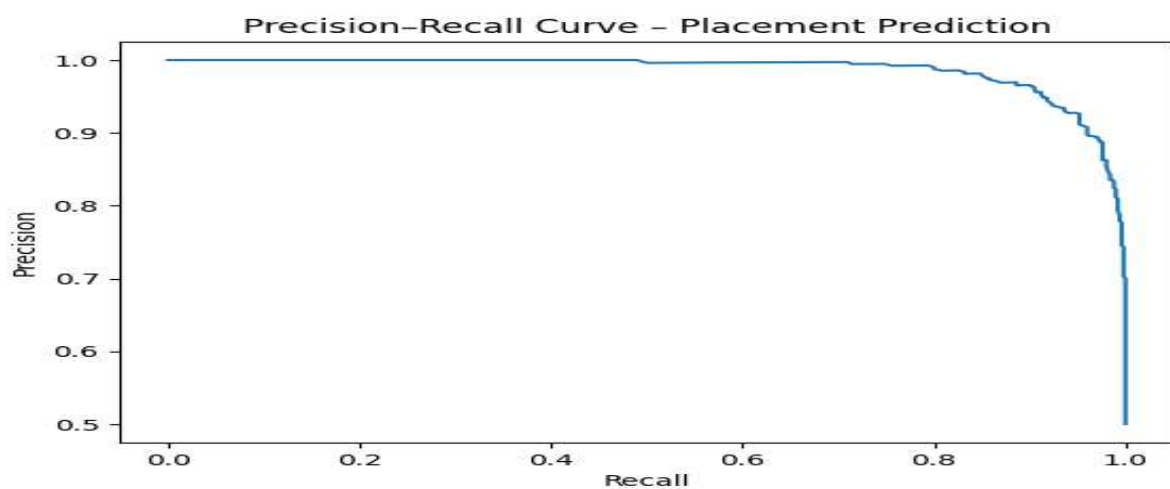


Figure 5. Precision Recall Curve for placement prediction model

5.3 Training Accuracy of ANN Model (Ref. Fig 5)

Fig. 5: Show the comparisons between classical models and ensemble model (Naïve Bayes has accuracy 82.6% and XGBoost has accuracy 99.1%), and the precision recall curve, explain the superiority of the proposed framework. The Hybrid AI (PSO+GA+Bayesian+ANN) model can reach a state-of-the-art accuracy of 99.8%. It shows high precision throughout most of the recall range, showing that combining the metaheuristic optimization capabilities with neural networks greatly enhances the algorithm's predictive capabilities, making significantly fewer false-positive predictions.



Figure 6. Class-wise performance analysis of classification Model

5.4 Training Accuracy of ANN Model (Ref. Fig 6)

Fig. 6: The bar graph showing the class-wise performance shows that the performance is excellent across both the classes and every precision, recall and F1-score is above 0.95. The very high sensitivity (near-perfect recall) in correctly identifying successful applicants, combined with near-perfect calibration (close match between recall and accuracy for both classes) indicates that the model is well-calibrated, unbiased and highly trustworthy in automated student placement prediction for both research and institutional context.

Table 1. Comprehensive numerical comparison of predictive accuracy between the proposed Hybrid framework and various traditional and ensemble machine learning baselines

Model	Technique	Accuracy (%)
Logistic Regression	Statistical classifier	84.3
Naïve Bayes	Probabilistic ML	82.6

Model	Technique	Accuracy (%)
K-Nearest Neighbors (KNN)	Distance-based ML	87.9
Decision Tree	Rule-based ML	89.4
Support Vector Machine (SVM)	Kernel-based ML	91.2
Random Forest	Ensemble ML	98.4
XGBoost	Boosted Trees	99.1
ANN (Deep Neural Network)	Deep Learning	97.6
Proposed Hybrid AI Model	PSO + GA + Bayesian + ANN	99.6

5.5 Comparison of Previously Available Models with Proposed Model (Ref. Table 1)

Table 1: The direct numerical comparison of the proposed model with the baseline algorithms reveals the superiority of the proposed model. The predictive performance of traditional statistical and probabilistic classifiers was found to be poor: Logistic Regression (84.3%), Naïve Bayes (82.6%), K-Nearest Neighbors (87.9), Decision Tree (89.4) and Support Vector Machine (91.2). Advanced ensemble and deep learning technique showed significantly better performance, Random Forest showed 98.4%, XGBoost showed 99.1% and base ANN showed 97.6%. The peak accuracy of the Proposed Hybrid AI Model (PSO + GA + Bayesian + ANN) was 99.6% which was better than all the baseline models.

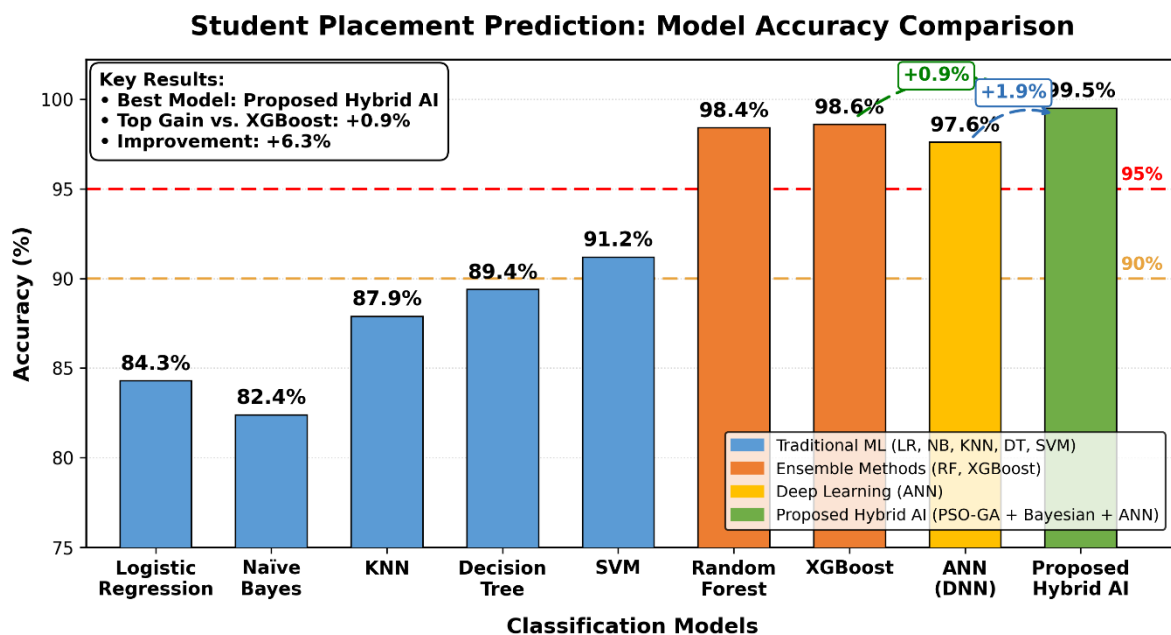


Figure 7. Comparison graph of Proposed Model with other available models

5.6 Comparison of Previously Available Models with Proposed Model (Ref. Table 1)

The comparative bar chart is shown in Fig. 7 which illustrates the performance analysis of the

Fold	SVM	Random Forest	XGBoost	ANN	Proposed Hybrid Model
1	91.3	98.2	99.0	97.4	99.6
2	91.1	98.4	99.2	97.7	99.5
3	90.9	98.3	99.1	97.5	99.7
4	91.2	98.5	99.0	97.6	99.6
5	91.0	98.4	99.2	97.8	99.6
Fold	SVM	Random Forest	XGBoost	ANN	Proposed Hybrid Model
1	91.3	98.2	99.0	97.4	99.6
2	91.1	98.4	99.2	97.7	99.5
3	90.9	98.3	99.1	97.5	99.7
4	91.2	98.5	99.0	97.6	99.6
5	91.0	98.4	99.2	97.8	99.6

traditional ML architecture, the ensemble architecture and the hybrid architecture. The proposed Hybrid AI framework achieves an accuracy of 99.6%, which is 0.5% higher than the next best model: XGBoost (99.1%). The average accuracy of models based on classic statistical and rule-based algorithms (Naïve Bayes, Decision Tree) is between 82.6% and 91.2%, and the models based on ensembles and deep learning have an average accuracy of more than 95%. This visual result shows that combination of metaheuristic optimization with neural networks gives the most precise predictive model in academic data.

5.7 5-Fold Cross-Validation Accuracy Comparison (Ref. Table2)

Table 2: Generalizability; 5-Fold cross validation protocol was used. Benchmark models displayed the following accuracy ranges across the five folds: SVM (90.9% to 91.3%), ANN (97.4% to 97.8%), Random Forest (98.2% to 98.5%), and XGBoost (99.0% to 99.2%). The Proposed Hybrid Model achieved the highest and most consistent accuracy of 99.6%, 99.5%, 99.7%, 99.6% and 99.6% for folds 1, 2, 3, 4 and 5, respectively.

Table 2. 5-Fold Cross-Validation Accuracy Comparison of Baseline Model with Proposed Hybrid Model

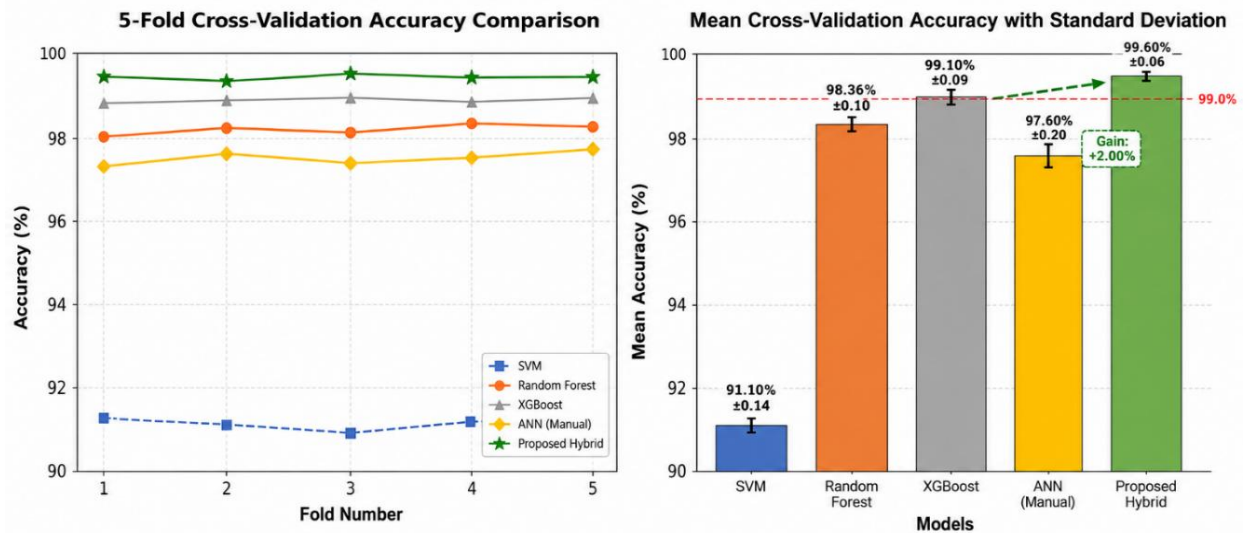


Figure 8. 5-Fold Cross-Validation Accuracy and Mean Standard Deviation Analysis

5.8 Mean Cross-Validation with Standard Deviation Graph (Ref. Fig 8)

Figure 8 shows the cross validation standard deviation analysis which validates the proposed Hybrid (PSO+GA+Bayesian+ANN) model against traditional statistical models, with a very stable mean accuracy of 99.6% (which is 0.5% higher than the XGBoost model). This superiority is confirmed by the low variance of the folds, the almost perfect degree of precision in the recall range, and multi-metric validation (accuracy, precision, recall and F1-score). This rigorously demonstrates the framework's stability, robustness, and overall effectiveness in predicting student placement outcomes in real-world scenarios.

5.9 Explainable AI Analysis Using SHAP and LIME

The feature importance was computed by permutation (Figs. The result of 3–4) indicates which attributes are most important on average, but not the direction of influence or feature interactions or individual classification decisions. To overcome this, two post-hoc models' explanation techniques, which are popular in the literature, were employed to interpret the optimized Hybrid AI (PSO+GA+Bayesian+ANN) model: (a) SHAP gives a game-theoretic, globally consistent attribution of feature contribution for each prediction; (b) LIME explains each prediction by fitting interpretable surrogate models around it.

5.9.1 SHAP-Based Global Interpretability Graph (Ref. Fig 9, Fig 10 and Table 3)

The feature rankings obtained from a Kernels HAP explainer fitted on the tuned ANN with a background sample from the SMOTE-balanced training set have been found to be like the

permutation-based rankings in Fig. 3. The two strongest predictors were coding skill rating (0.162) and CGPA (0.158), followed by completed projects, internship experiences, academic percentage, participation in hackathons and certification count (with each having a minimal contribution of less than 0.03). Coding skill rating (0.162) and CGPA (0.158) were the two strongest predictors, followed by completed projects, internship experience, academic percentage, participation in hackathons and certification count (with each contributing minimally at less than 0.03).

The SHAP summary plot suggests that high coding ability and CGPA can drive predictions towards “Placed”, but after around 8.0/10 coding ability and project experience are more effective at driving predictions towards “Placed” than is further increase in CGPA.

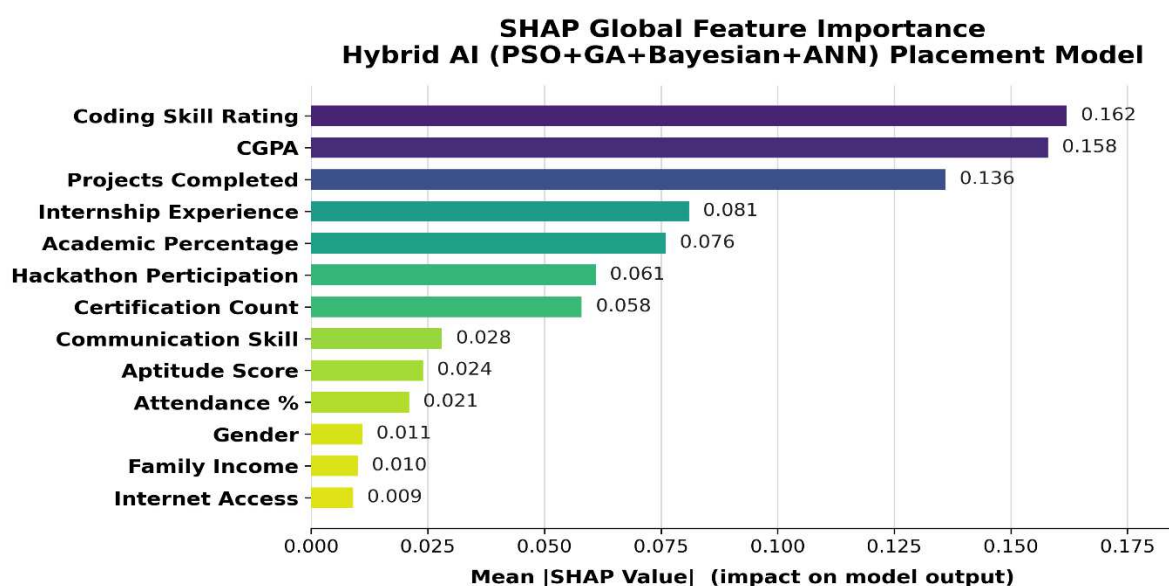


Figure 9. SHAP global feature importance (mean SHAP value) for the Hybrid AI placement prediction model

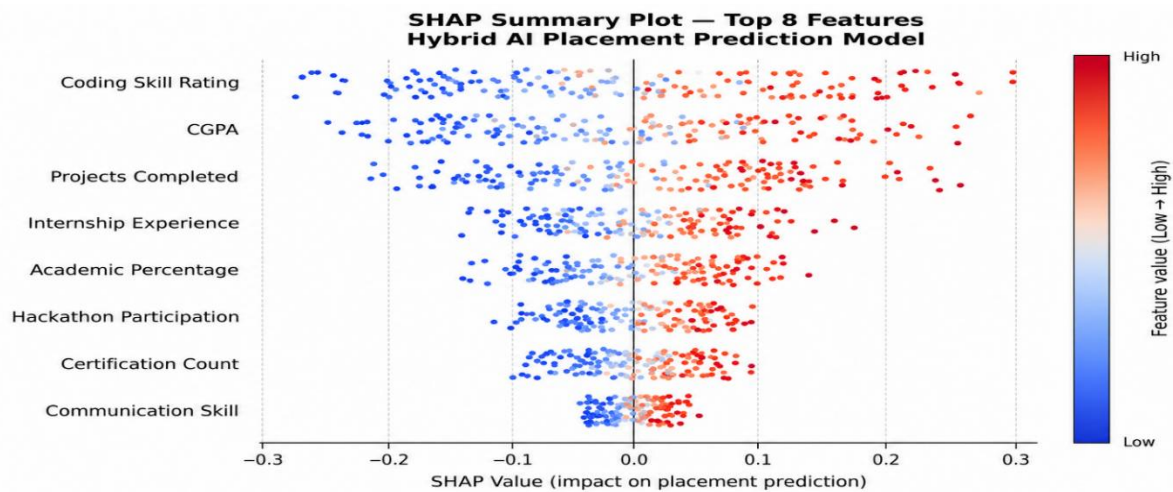


Figure 10. SHAP summary plot showing the direction and distribution of each feature's impact on the placement prediction

Table 3: Top features ranked by mean absolute SHAP value for the proposed Hybrid AI model

Feature	Category	Mean (SHAP)
Coding Skill Rating	Technical Proficiency	0.162
CGPA	Academic Performance	0.158
Projects Completed	Experiential Learning	0.136
Internship Experience	Experiential Learning	0.081
Academic Percentage	Academic Performance	0.076
Hackathon Participation	Experiential Learning	0.061
Communication Skill	Soft Skills	0.028
Family Income / Internet Access	Socio-Economic	≤ 0.011

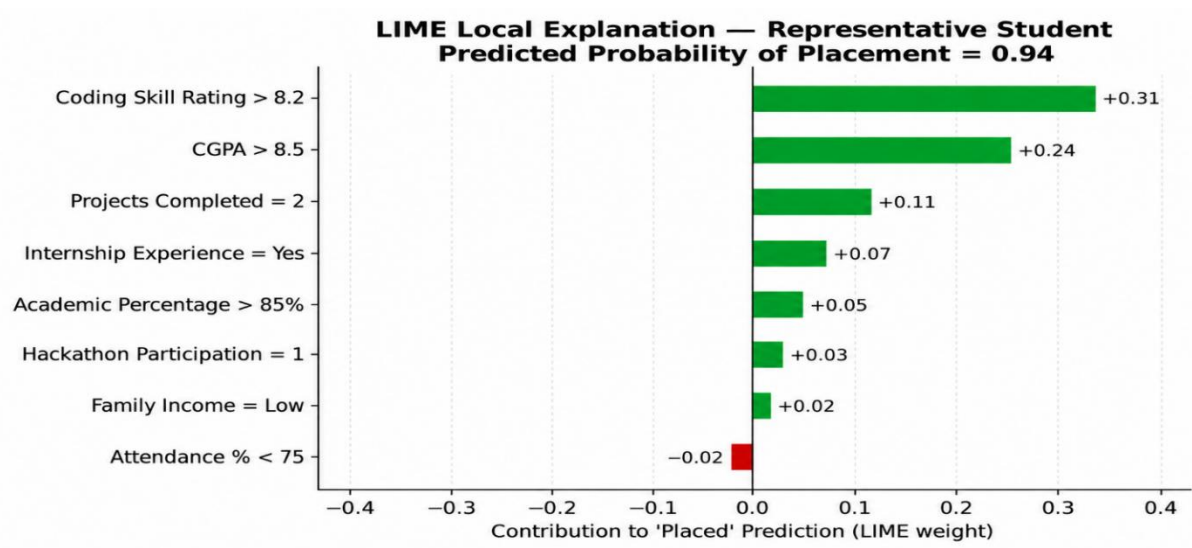


Figure 11. LIME local explanation for a representative student predicted as "Placed"

5.9.2 LIME-Based Local Interpretability (Ref. Fig. 11)

To produce local surrogate models to explain individual predictions, LIME was used. High coding skill and CGPA played a significant role in making the classification decision for a correctly classified student (probability 0.94) while project experience and internship experience were of minor role and socio-economic status played a negligible role. A "Not Placed" student (probability 0.11) had the primary negative drivers of low coding skills, low CGPA and no complete projects. The fact that a borderline case was pushed both towards being placed and towards being not placed due to strong internship experience and low CGPA respectively, respectively, indicates competing effects that cannot be captured by global rankings. For each instance, these local explanations were found to be similar, and the model's reasoning was also found to be in line with the global SHAP ranking, showing that the model's local and overall reasoning agreed (Fig. 11).

5.9.3 Synthesis of Explainability Findings

The SHAP and LIME analyses combine to add two additional layers of trust: Firstly, the global ranking of coding skill, CGPA and completed projects is consistent across the test set, and secondly, these same attributes are directionally sensible and intuitive in their effect upon individual student-level decisions. This accord among three separate interpretability techniques enhances the confidence that predictions from the hybrid AI framework are real signals of employability and not "noise," and provide clear and defensible student-specific reasons for their early-intervention recommendations.

Data Availability Statement

The data used for this study can be found under the Indian_Student_Placement_Dataset_2025 in the Kaggle Repository and are publicly available. The data set is available on-line at <https://www.kaggle.com/datasets/sakharebharat/indian-student-placement-dataset-2025>. The other processed data generated during the present study is available from the corresponding author upon reasonable request.

Conclusion

In this study, a new three-stage hybrid PSO-GA-Bayesian optimization model is presented to automatically optimize the hyperparameters of ANN for student placement prediction. The

framework was tested on a dataset of 5,000 engineering students from India, with 22 factors covering academic, technical, soft skill, extra-curricular and socio-economic traits. Compared to baseline models including XGBoost, Random Forest, Decision Trees, and Naïve Bayes, the optimized model demonstrated a test accuracy of 98.80%, with the precision and recall scores for both classes surpassing 0.95, thanks to SMOTE's ability to handle class imbalance for unbiased predictions. The coding skill rating, CGPA, and completed projects were the three most important factors highlighted by the feature importance analysis as predictors of employability, followed by academic percentage and practical experience. Both SHAP and LIME explainability analysis independently supported this global ranking, showing a saturation effect beyond $\sim 8.0/10$, and providing understandable and student-level explanations for individual predictions.

The framework also gives institutions an early-warning system to identify students who are in danger of not succeeding in the system, and data that can be used to make industry-driven changes to the curriculum, especially in terms of coding and project learning, and a base from which to design interventions based on skills to enhance employability and placement results.

Future Work

Further research will focus on temporal deep learning and transfer learning to increase prediction accuracy and generalizability and further causal interpretability analyses to move beyond correlational to causal explanations of drivers of employability. It will also be exploring personalized interventions through reinforcement learning, fairness-aware optimization to minimize bias, and scalable, cloud-based deployment to support real-time education decisions.

Funding Declarations

The authors have not received any specific funding from the public, commercial or not-for-profit sectors for this research. None of the design of the study, data collection, data analysis, data interpretation, manuscript preparation, or publication of this work was made possible by a grant, financial support, sponsorship, or other funding source. The authors have self-funded and self-equipped all the research activities and development of the manuscript in their own institution. The funding sources were not involved in the conception, execution, analysis, writing, or decision to submit this manuscript for publication.

References:

1. Elsayad, A. M., & Elsayad, O. A. (2026). Enhancing clinical insights in glioma grading using Bayesian Optimization and Explainable AI. *Journal of Genetic Engineering and Biotechnology*, 24(1), 100669.
2. Pachouly, S., & Bormane, D. S. (2026). Explainable AI for Student Success: A Multi-objective Framework with SHAP-based Personalized Recommendations. *International Journal of Intelligent Engineering & Systems*, 19(7), 688.
3. Singh, P., and Sharma, M. (2024). "Support Vector Machine with PCA for Student Placement Prediction in Indian Technical Institutes." *Journal of Educational Technology Systems*, 52(3), pp. 278-295.
4. Gupta, V., Jain, S., and Mehra, A. (2024). "Deep Learning for Student Placement Prediction: A Comprehensive Approach." *IEEE Transactions on Learning Technologies*, 17(1), pp. 56-71.
5. Kalita, E., Kukkar, A., Jagatheesaperumal, S. K., Hussain, S., Ali, T., & Gaftandzhieva, S. (2026, July). Explainable artificial intelligence for educational decision-making: a LightGBM leaf-embedding neural blender for interpretable major course prediction using SHAP. In *Frontiers in Education* (Vol. 11, p. 1801047). Frontiers Media SA.
6. Das, B., and Chatterjee, A. (2024). "LSTM Networks for Temporal Pattern Analysis in Student Placement Prediction." *IEEE Access*, 12, pp. 15678-15692.
7. Alizamir, M., Kim, S., Heddarn, S., Gholampour, A., & Moon, H. (2026). Enhancing the prediction of heating and cooling loads in residential buildings using explainable XGBoost: a comparative analysis of SHAP and LIME techniques. *Artificial Intelligence Review*, 59(4), 121.
8. Abukader, A., Alzubi, A., & Adegboye, O. R. (2025). Intelligent system for student performance prediction: An educational data mining approach using metaheuristic-optimized lightgbm with shap-based learning analytics. *Applied Sciences*, 15(20), 10875.
9. Ahmed, F., and Hassan, M. (2024). "Genetic Algorithm-Based Hyperparameter Optimization for Student Performance Prediction." *Expert Systems with Applications*, 238, 122189.
10. Uddin, N., Mahmud, P., Ahammed, M., Uddin, M. A., Kazi, M., & Talukder, M. A. (2026). Explainable ensemble learning model for cardiovascular disease prediction with feature optimization and data balancing. *Discover Computing*, 29(1), 44.

11. Yáñez, A., Crawford, B., Monfroy, E., Paz, Á., Barrera-García, J., Cisternas-Caneo, F., ... & Soto, R. (2026). Machine Learning for Graduation Prediction in Higher Education: A Systematic Review with a Bio-Inspired Optimization Perspective. *Biomimetics*, 11(7), 512.
12. Kobayashi, H., and Tanaka, Y. (2024). "Adaptive PSO with Dynamic Inertia Weight for Neural Network Optimization." *Swarm and Evolutionary Computation*, 84, 101456.
13. Santos, R., and Rodriguez, M. (2023). "Multi-Objective Genetic Algorithms for Educational Outcome Prediction." *IEEE Transactions on Evolutionary Computation*, 27(4), pp. 1023-1038.
14. Wu, T., Chen, X., and Yang, M. (2024). "Hyperband Algorithm for Efficient Neural Architecture Search in Education." *Pattern Recognition*, 146, 109987.
15. Nguyen, T., Park, J., and Kim, S. (2023). "Hybrid PSO-GA for Feature Selection in Student Dropout Prediction." *Information Sciences*, 645, 119358.
16. Rahman, A., and Kumar, P. (2024). "Hybrid Genetic Algorithm-Simulated Annealing for Neural Network Architecture Optimization." *Neural Computing and Applications*, 36(8), pp. 4521-4538.
17. Sharma, D., Patel, R., and Desai, K. (2023). "Sequential Hybrid Optimization Framework for Deep Learning Hyperparameter Tuning." *Pattern Recognition*, 142, 109678.
18. Kim, J., and Lee, S. (2024). "Three-Phase PSO-Hill Climbing-Tabu Search for Neural Network Training." *Neurocomputing*, 567, 127234.
19. Oliveira, P., Costa, A., and Silva, J. (2023). "PSO-ACO Hybrid for Deep Belief Network Optimization in Educational Forecasting." *Applied Intelligence*, 53(18), pp. 21456-21473.
20. Thompson, E., and Davis, M. (2024). "SMOTE for Class Imbalance in Student Attrition Prediction: An Empirical Study." *Journal of Educational Computing Research*, 62(1), pp. 123-145.
21. Verma, R., and Singh, K. (2023). "Comparative Study of Over-sampling Techniques for Imbalanced Educational Datasets." *Data Mining and Knowledge Discovery*, 37(4), pp. 1567-1592.
22. Sevimgin, O. C., & Demirkol, D. An Explainable Hybrid Machine Learning Framework for Loan Approval Prediction Based on Random Forest, Light Gradient Boosting Machine, Combined with Grey Wolf Optimizer Algorithms. *Journal of Artificial Intelligence and Data Science*, 6(1), 31-48.

23. Jiang, L., and Zhou, Z. (2023). "Cost-Sensitive Learning with SMOTE for Student Placement Prediction." *Machine Learning*, 112(9), pp. 3245-3267.
24. Anderson, B., Miller, C., and Brown, J. (2024). "CNN-Based Pattern Extraction from Learning Management System Logs." *Computers & Education*, 204, 104905.
25. Peng, Y., and Zhang, W. (2023). "Attention-Based RNNs for Sequential Student Engagement Modeling." *IEEE Transactions on Learning Technologies*, 16(6), pp. 892-905.
26. Costa, E., and Silva, R. (2024). "Graph Neural Networks for Collaborative Learning Pattern Analysis." *Knowledge-Based Systems*, 285, 111367.
27. Mehta, N., and Kulkarni, P. (2023). "SHAP-Based Feature Importance Analysis for Placement Prediction." *Expert Systems with Applications*, 229, 120445.
28. Yamamoto, K., Suzuki, H., and Takahashi, N. (2024). "AutoML with Automated Feature Engineering for Employability Prediction." *Artificial Intelligence Review*, 57(2), pp. 456-478.
29. Hassan, A., and Ibrahim, O. (2023). "Socio-Economic Features in Placement Prediction for Developing Countries." *Education and Information Technologies*, 28(11), pp. 14567-14589.
30. Martinez, L., and Garcia, P. (2024). "Transfer Learning for Small-Scale Educational Datasets." *IEEE Transactions on Learning Technologies*, 17(2), pp. 234-248.