

ISSN 1840-4855

e-ISSN 2233-0046

Original scientific article

<http://dx.doi.org/10.70102/afts.2026.1836.942>

COGNITIVE READINESS BASED CAMPUS PLACEMENT PREDICTION USING ANN AND SHAP

Hameeda Khatoon^{1*}, Chandra Prakash Vudatha²

^{1*}Research Scholar, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, India.

e-mail: 2102030004@kluniversity.in, orcid: <https://orcid.org/0009-0005-4199-9118>

²Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India.

e-mail: vchandrap@kluniversity.in, orcid: <https://orcid.org/0000-0002-5030-1122>

Received: June 15, 2026; Revised: July 24, 2026; Accepted: August 31, 2026; Published: September 29, 2026

SUMMARY

Placements at campus level are one of the most crucial events for engineering and computing students; however, institutes use simplistic cutoffs and subjective judgment methods to determine the placement status of their students. In this study, a classification system is proposed by introducing new operational cognitive readiness criteria which include four major components: Memory, Intelligence, Learning Ability, and Soft Skills. An ANN (Artificial Neural Network) is developed using the Kaggle dataset, containing details of 10,000 students including CGPA, Aptitude Scores, Soft Skills Scores, Internship, Project, Training, and Placement Status. ANN is chosen because it can model the non-linear relationship among these cognitive factors. The initial assessment was conducted using the unfiltered data set where the ANN provided 78.00% accuracy and ROC-AUC score of 0.86, which proves that cognitive factors play a partial role in explaining the historical placement outcomes. In order to evaluate the alignment, cognitive readiness filtering is done where historically placed students are excluded whose total cognitive score is below 70 on a 100 scale. The filtering process reduced the number of inconsistent records related to positive placement outcomes with respect to the operational cognitive paradigm to 21.59%. On the filtered ready-to-go cohort, the ANN was able to provide 94.58% accuracy, 0.90 F1 score, and 0.98 ROC-AUC value. Crucially, while the ANN achieved the highest nominal accuracy among tested models, corrected statistical tests confirmed that its performance remains statistically competitive with rather than demonstrably superior to simpler classical machine learning baselines. Kernel SHAP method was applied to interpret the model prediction and revealed Learning Ability and Intelligence as the key features, then Soft Skills and Memory. Overall, the proposed framework provides an interpretable decision-support approach that aligns with institutional placement evaluation processes.

Key words: placement prediction, graduate employability, cognitive readiness, artificial neural network, explainable AI, SHAP.

INTRODUCTION

Research Problem and Significance

For many engineering and computer science undergraduate students, placement is not just an academic milestone. It is the point where in several years' worth of learning is converted into their first job and the return on investment for the students, their families and the university. Universities publish placement statistics from time to time and incoming students utilize the statistics to gauge the quality of the programme concerned. However, in making everyday decisions, the concept of being "placement ready" is typically based on simplistic cut-off criteria, heuristics or the opinion of faculty and placement officials. These approaches work in straightforward cases but risks missing out on borderline students who could have succeeded with proper assistance or fails to highlight candidates who, although good enough in theory, struggle in assessments, technical interviews and group discussions.

Despite the growing use of machine learning in campus recruitment, an important research problem remains unresolved. Existing placement models commonly process academic performance, aptitude scores, training records and skill indicators as separate input variables. Consequently, the resulting predictions may not clearly represent a student's overall readiness for placement. Limited interpretability further reduces the practical value of such models because placement officers and academic counsellors require both a predicted outcome and an understandable explanation of the factors influencing it. Therefore, the present study addresses the need for an interpretable placement prediction framework that combines available student attributes into meaningful cognitive readiness indicators. Such a framework can support the early identification of placement-related strengths and weaknesses and facilitate more focused placement preparation. Machine-learning methods have increasingly been used to predict student placement and graduate employability from academic, aptitude, internship, project and profile attributes [1][2][3]. Related studies have examined placement outcomes across different graduate populations [4][5]. Comparative evaluations of regression, tree-based, support-vector and neural-network models show that performance varies across datasets and configurations [6][7]. Machine learning has also been used to predict graduate employment status and employment type, while a recent survey identified limited standardisation in feature selection and evaluation procedures [8] [9].

Graduate employability depends on academic attainment, professional competencies and labour-market-relevant skills [10][11]. Work-based learning, communication, teamwork and other soft skills contribute to employment readiness [12][13]. Internships are associated with improved employment opportunities and may support students' academic, professional and personal development [16][17][18]. Placement prediction should therefore represent broader readiness dimensions rather than relying only on raw academic attributes [14][15]. Explainable artificial intelligence is increasingly used to improve the transparency of educational prediction. SHAP and related approaches have been applied to student adaptability and academic-performance prediction [19][20]. Explainable frameworks have also been developed for dropout prediction and educational decision support [21][22][23]. An analysis of trends in research will help in determining important factors, methodology, and direction in data-based decision making [25]. These studies demonstrate the value of interpretable models when predictions inform student training, counselling or academic intervention.

Proposed Framework and Novel Contributions

The present study is positioned at the intersection of campus placement prediction, graduate employability and explainable artificial intelligence. Instead of using a large collection of loosely related variables, attributes obtained from a publicly available placement dataset are reorganised into four operational cognitive readiness factors: Memory, Intelligence, Learning Ability and Soft Skills. The relevant attributes include CGPA, aptitude-test scores, soft-skills ratings, projects, internships, workshops or certifications and placement training. Research trend analysis allows the identification of the influential factors as well as the emerging directions for evidence-based decision-making [27]. This representation reflects the broader understanding of employability as a combination of academic knowledge, reasoning ability, experiential learning and interpersonal competence [28].

Memory represents sustained academic performance through cumulative CGPA. Intelligence combines aptitude performance with academic achievement to represent academic and reasoning strength. Learning Ability combines projects, internships, workshops, certifications and placement training to represent active learning and practical exposure. The Soft Skills factor represents the communication and interpersonal rating available in the dataset. These factors are intended as operational indicators derived from educational records rather than direct clinical or psychometric measurements of cognition. The resulting representation describes student readiness using concepts that can be understood by faculty members, placement officers and academic counsellors. An artificial neural network is used to predict the binary placement outcome because it can model nonlinear interactions among the four cognitive readiness factors.

Initial modelling on the complete dataset produced approximately 78% accuracy. Examination of the records identified historically placed students with at least one cognitive factor below 70. This motivated a secondary readiness-filtered scenario for studying how alignment with the operational readiness criteria affects model behaviour. Exclusion does not imply that the recorded placement outcome is incorrect; it only indicates that the record does not satisfy the predefined readiness rule. Kernel SHAP is applied to explain the direction and magnitude of each factor's contribution to the predicted placement probability. This supports transparent interpretation of the ANN's outputs. The principal contribution is the integration of operational cognitive-factor derivation, readiness-based dataset analysis and global and local SHAP explanations within an ANN placement-prediction framework.

Organisation of the Paper

The remainder of the paper is organised as follows. Section 2 reviews previous studies on campus placement prediction, employability factors and explainable artificial intelligence. Section 3 describes the dataset, preprocessing procedures, cognitive factor derivation, cognitive readiness filter, ANN configuration and SHAP methodology. Section 4 presents the experimental results and comparisons with other classification models. Section 5 discusses the principal findings and their relationship with previous studies. Section 6 presents the study's limitations, practical implications and future scope. Section 7 concludes the study.

RELATED WORK

This section reviews three areas relevant to the study: machine-learning-based placement prediction, employability-readiness determinants and explainable educational prediction. These areas provide the basis for integrating operational cognitive factors, readiness filtering, ANN prediction and SHAP explanations.

Machine-Learning Approaches to Placement and Employability

Placement and employability prediction is commonly formulated as a binary classification problem using academic, aptitude, skill, training and profile attributes [1][2][3]. Recent studies have extended this approach across different graduate populations and institutional contexts [4][5]. Comparative evaluations indicate that the performance of regression, tree-based, support-vector and neural-network models depends on the dataset and selected features [6][7]. Machine learning has also been used to predict graduate employment status and employment type, although limited standardisation remains in feature selection and evaluation procedures [8] [9]. Hybrid employability frameworks have combined multi-criteria decision-making with machine learning [10]. Recent methodological research has further emphasised rigorous model evaluation, while placement-specific work has incorporated SHAP to interpret influential attributes.

Table 1 shows that existing studies primarily compare classifiers using raw academic, skill and profile attributes. Although recent approaches extend prediction across different graduate contexts and incorporate SHAP-based interpretation, the reviewed studies do not jointly integrate operational cognitive factors, readiness filtering and global and local explanations. The present framework addresses this gap through a cognitively informed ANN-SHAP workflow.

Table 1. Representative machine-learning-based placement and employability prediction studies

Ref.	Year	Context/Domain	Features (high level)	Methods/Models	Target task	Key point/limitation relative to the present study
[1]	2025	IT graduates in a low-income context	Academic, programme and background attributes	Tree-based classifiers	Employability prediction	Demonstrates strong tree-based performance but uses raw attributes without cognitive-factor construction or SHAP.
[2]	2024	Engineering graduates' placement	Academic and skill-related attributes	Multiple ML classifiers	Placed/Not Placed	Focuses on classifier comparison without readiness filtering or explainable analysis.
[3]	2020	Institutional student-placement data	Academic, aptitude and profile attributes	Multiple ML classifiers, including ANN	Placement prediction	Demonstrates the usefulness of ML but relies on low-level input attributes.
[4]	2025	Early employability of Vietnamese graduates	Academic, profile and employability-related attributes	Comparative ML models	Early employability prediction	Extends prediction to a recent graduate context but does not construct cognitive factors.
[6]	2022	IT-graduate employability	Academic and skill indicators	DT, NB, LR, RF and SVM	Employable/Non-employable	Compares classifiers without cognitive aggregation, readiness filtering or SHAP.
[8]	2025	South Korean university graduates	GPA, geographic, family-background and graduate attributes	Multiple ML models with SHAP	Employment status and type	Provides explainability but uses heterogeneous raw attributes without readiness filtering.
[10]	2025	Moroccan graduate employability	Academic, skill and decision criteria	Hybrid AHP–TOPSIS and ML	Employability prediction and ranking	Integrates decision analysis with ML but does not use cognitive-factor-based ANN–SHAP modelling.
[11]	2025	Graduate job-placement prediction	Grades, GPA, internship and extracurricular attributes	Multiple classifiers with SHAP	Job-placement prediction	Provides SHAP-based interpretation but does not integrate cognitive-factor construction with readiness filtering.

Cognitive, Soft-Skill and Experiential Determinants of Employability

Graduate employability extends beyond academic performance and depends on professional competencies, communication, teamwork, adaptability and workplace behaviour [10][11][13]. Work-based learning and reflective placements help students translate academic knowledge into employment readiness [12] [14] [15]. Internships are associated with improved employment

opportunities and graduate employability and may also support academic, professional and personal development [16][17][18]. Collectively, these findings support the inclusion of academic, aptitude, soft-skill and experiential attributes in placement prediction. However, the reviewed studies generally examine these attributes separately rather than transforming them into compact readiness dimensions. The present study addresses this limitation by deriving four operational cognitive readiness factors from observable academic, skill, training, project, workshop and internship attributes.

Explainable AI in Educational Prediction

Explainable artificial intelligence (XAI) has gained importance in educational prediction because model outputs should be interpretable when used for academic support or intervention. SHAP-based approaches have been applied to student adaptability and academic-performance prediction [19][20]. Explainable pipelines have also been developed for dropout prediction, educational decision support and student-performance analysis [21][22][23].

SHAP has recently been applied directly to graduate-employment and job-placement prediction [8]. Although these studies identify influential academic and experiential attributes, the inputs remain predominantly raw variables and are not integrated with operational cognitive-factor construction or readiness filtering. SHAP supports both global feature-importance analysis and instance-level explanation [26]. Data-driven analytical approaches can support more informed decision-making by identifying relevant factors and patterns within complex socioeconomic environments. Accordingly, Kernel SHAP is applied to the ANN to estimate the predictive contributions of Memory, Intelligence, Soft Skills and Learning Ability. These contributions explain model behaviour and are not interpreted as causal effects.

Research Gap and Contribution

As reviewed in Sections 2.1–2.3, existing placement-prediction studies primarily use raw academic, aptitude, skill and profile attributes. Employability research highlights soft skills and practical learning experiences, while recent studies demonstrate the value of SHAP-based interpretation. However, these elements remain insufficiently integrated within a cognitive-readiness-oriented placement framework. The proposed framework transforms seven observable attributes into four operational factors Memory, Intelligence, Soft Skills and Learning Ability and evaluates ANN performance before and after applying a predefined readiness rule. Global and local SHAP explanations are then used to identify overall factor importance and explain individual predictions. The novelty lies in integrating ANN and SHAP with operational cognitive-factor construction, readiness-based dataset analysis and student-level explanation rather than in either method individually. Coefficient sensitivity analysis, comparative model evaluation, cross-validation, confidence intervals and corrected statistical tests further assess the robustness of the framework [24]. Together, these elements provide an interpretable and systematically validated placement-readiness workflow.

METHODOLOGY

This study proposes a cognitively informed ANN framework for predicting the campus-placement outcomes of computer graduates. Instead of using the original academic and activity attributes independently, seven observable attributes are reorganised into four operational factors Memory, Intelligence, Soft Skills and Learning Ability. The ANN is evaluated using the complete dataset and a readiness-filtered dataset in which historically placed records failing at least one predefined readiness criterion are excluded. Global and local SHAP explanations are subsequently used to interpret the predictive contribution of each factor, consistent with recent explainable educational and placement-prediction research [19][20].

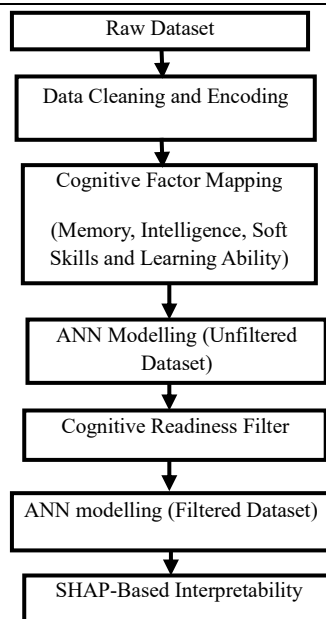


Figure 1. Research methodology workflow

Figure 1 provides an overview of the workflow: data acquisition and preprocessing; cognitive factor derivation; ANN modelling on the cognitive factors; application of the cognitive readiness filter; re-modelling on the filtered data; and SHAP-based interpretability.

Dataset Description and Preprocessing

Data Source and Sample Distribution

The experiments were conducted using the publicly available Placement Prediction Dataset obtained from Kaggle. The dataset is available at <https://www.kaggle.com/datasets/ruchikakumbhar/placement-prediction-dataset> under the Creative Commons CC0 licence. It contains 10,000 student records and 12 attributes covering academic performance, aptitude, practical learning activities, training participation and placement outcomes. The range of academic and skill-related attributes makes the dataset suitable for constructing cognitive readiness factors and evaluating machine-learning models for campus placement prediction.

PlacementStatus is the target variable and contains two classes. As summarised in table 2, the original dataset includes 4,197 students labelled Placed and 5,803 students labelled NotPlaced, representing 41.97% and 58.03% of the complete sample, respectively. The distribution shows a moderate difference between the two classes, with NotPlaced students forming the larger group. Stratified sampling was therefore used during model development to preserve the original class proportions in the training and testing subsets.

Table 2. Distribution of placement classes in the original dataset

Placement status	Number of records	Percentage (%)
Placed	4,197	41.97
Not Placed	5,803	58.03
Total	10,000	100.00

Feature Description and Selection

The original dataset contains StudentID, ten student-related attributes and PlacementStatus as the target variable. Their data types, observed ranges and descriptions are presented in table 3. Seven attributes were selected to construct the four cognitive readiness factors. CGPA and AptitudeTestScore were used to derive Memory and Intelligence, SoftSkillsRating represented Soft Skills, and Projects, Internships, Workshops/Certifications and PlacementTraining were used to derive Learning Ability. StudentID was

excluded because it functions only as an identifier. SSC_Marks and HSC_Marks were excluded because CGPA provides a more recent cumulative measure of academic performance, while their inclusion could give excessive weight to academic history. ExtracurricularActivities was excluded because it was not mapped to any of the four operational cognitive factors. PlacementStatus was retained as the target variable for binary classification.

Table 3. Description of the original dataset attributes

Attribute	Type and range	Description
StudentID	Integer: 1–10,000	Unique identifier assigned to each student record
CGPA	Numerical: 6.5–9.1	Cumulative academic performance in the degree programme
Internships	Integer: 0–2	Number of internships completed by the student
Projects	Integer: 0–3	Number of academic or technical projects completed
Workshops/Certifications	Integer: 0–3	Number of workshops attended or certifications completed
AptitudeTestScore	Numerical: 60–90	Score obtained in the aptitude assessment
SoftSkillsRating	Numerical: 3.0–4.8	Rating representing communication and interpersonal skills
ExtracurricularActivities	Categorical: Yes or No	Indicates participation in extracurricular activities
PlacementTraining	Categorical: Yes or No	Indicates participation in institutional placement training
SSC_Marks	Numerical: 55–90	Marks obtained in the secondary-school examination
HSC_Marks	Numerical: 57–88	Marks obtained in the higher-secondary examination
PlacementStatus	Categorical: Placed or NotPlaced	Recorded campus-placement outcome and target variable

Data Preprocessing

The dataset was examined for missing values, duplicate identifiers, invalid class labels and inconsistencies in column names. All 12 columns contained 10,000 non-null values; therefore, missing-value imputation was not required. Column names and class labels were standardised to maintain consistency during data processing. StudentID was removed before model development because it was used only for record identification. PlacementStatus was encoded as 1 for Placed and 0 for Not Placed. PlacementTraining was transformed to 100 for Yes and 0 for No so that it could be combined with the remaining components of the Learning Ability factor. The other selected attributes were already numerical and did not require categorical encoding. The original data were divided into stratified training and testing subsets using an 80:20 ratio and a fixed random seed of 42. This produced 8,000 records for model development and 2,000 records for final testing while preserving the original proportions of Placed and Not Placed students. After cognitive factor construction, min–max normalisation was applied to the four derived factors. The scaler was fitted using only the training data and was subsequently applied to the testing data to prevent information from the test set from influencing model development. The detailed cross-validation procedure is presented in the model-validation subsection.

Cognitive Factor Derivation

Academic performance and aptitude are relevant attributes in student-placement prediction [2][3]. Communication, interpersonal competence and other soft skills contribute to graduate employability, while internships and practical learning experiences support professional development and employment readiness [13][14][15]. Based on this literature and the available dataset attributes, seven observable attributes were transformed into four operational factors: Memory, Intelligence, Soft Skills and Learning Ability. Memory represents cumulative academic attainment; Intelligence combines aptitude and academic performance; Soft Skills represents interpersonal readiness; and Learning Ability captures participation in projects, workshops or certifications, internships and placement training.

Mathematical Notation

For the mathematical formulation, the subscript i represents an individual student record. The symbols used in cognitive factor derivation are defined in table 4.

Table 4. Mathematical notation used for cognitive factor derivation

Symbol	Meaning
C_i	CGPA of student i
A_i	Aptitude-test score of student i
S_i	Soft-skills rating of student i
P_i	Number of projects completed by student i
W_i	Number of workshops or certifications completed by student i
N_i	Number of internships completed by student i
T_i	Placement-training value: 100 for Yes and 0 for No
M_i	Memory score of student i
I_i	Intelligence score of student i
SS_i	Soft Skills score of student i
L_i	Learning Ability score of student i

Cognitive Factor Formulation

Memory represents cumulative academic performance. As CGPA is measured on a 0–10 scale, it is multiplied by 10 to express the value on a comparable scale (Equation 1):

$$M_i = C_i \times 10 \quad (1)$$

Intelligence combines aptitude-test performance with cumulative academic achievement. The converted CGPA value and aptitude-test score A_i are averaged using equation (2):

$$I_i = \frac{A_i + (C_i \times 10)}{2} \quad (2)$$

Soft Skills represents communication and interpersonal readiness. The original soft-skills rating is multiplied by 20 (Equation 3):

$$SS_i = S_i \times 20 \quad (3)$$

Learning Ability combines placement training, projects, workshops or certifications and internships. It is calculated using equation (4):

$$L_i = \frac{T_i + (30 \times P_i) + (30 \times W_i) + (50 \times N_i)}{4} \quad (4)$$

The coefficients in equations (1)–(4) are predetermined scale-conversion values, not ANN-learned weights. Multipliers of 10 and 20 convert CGPA and Soft Skills to approximately 0–100 scales. Based on the dataset ranges, three projects or workshops produce maximum scores of 90, while two internships and placement training produce maximum scores of 100. The larger internship coefficient reflects its smaller count range and direct professional exposure [14][15]. All coefficients were fixed before ANN training and were not optimised using placement outcomes.

ANN Modelling Framework

The ANN uses the four cognitive readiness factors derived in Section 3.2 as input features. The same feature representation and network configuration are applied in the unfiltered and readiness-filtered scenarios to ensure a consistent comparison.

Feature–Target Definition

After cognitive factor generation, each student i is represented by four input features: Memory, Intelligence, Soft Skills and Learning Ability. The input feature vector is defined as

$$X_i = [M_i, I_i, SS_i, L_i] \quad (5)$$

where M_i, I_i, SS_i, L_i denote the Memory, Intelligence, Soft Skills and Learning Ability scores of student i , respectively.

The placement outcome is represented as a binary target variable:

$$y_i = \begin{cases} 1, & \text{if student } i \text{ is Placed} \\ 0, & \text{if student } i \text{ is Not Placed} \end{cases} \quad (6)$$

Thus, X_i in equation (5) represents the unnormalised input feature vector for student i , while y_i in equation (6) represents the corresponding binary placement outcome. The same feature and target definitions are used in both experimental scenarios.

Feature Normalisation

Although expressed on approximately comparable scales, the four cognitive factors have different numerical ranges and distributions. Therefore, min–max normalisation is applied to each factor before ANN training. For feature j of student i , the normalised value Z_{ij} is calculated as shown in equation (7):

$$Z_{ij} = \frac{X_{ij} - X_j^{\min}}{X_j^{\max} - X_j^{\min}} \quad (7)$$

where X_{ij} is the original value of feature j for student i , $X_j^{\min}$ and $X_j^{\max}$ are the minimum and maximum training-set values of that feature, and Z_{ij} is the resulting normalised value. Here, j represents one of the four cognitive factors. The normalised values lie between 0 and 1.

The minimum and maximum values are calculated using only the training data and are subsequently applied to the validation and test data. This procedure maintains consistent feature scaling while preventing information from the validation or test sets from influencing model training.

The readiness filter described in Section 3.4 is applied to the original cognitive-factor scores before normalisation. Within each experimental scenario, the normalised feature vector $Z_i = [Z_{i_1}, Z_{i_2}, Z_{i_3}, Z_{i_4}]$ is subsequently supplied to the ANN.

ANN Architecture and Network Settings

The ANN is implemented using scikit-learn's MLPClassifier with a feed-forward architecture comprising an input layer with four neurons, one for each cognitive factor; a first hidden layer with 16 neurons and ReLU activation; a second hidden layer with 8 neurons and ReLU activation; and a single-neuron output layer with sigmoid activation, representing the predicted probability of the *Placed* class. The resulting ANN architecture is illustrated in figure 2. The network is trained using the Adam optimiser with a maximum of 500 iterations. A fixed random seed (random_state=42) is used to support reproducibility. Model performance is evaluated using accuracy, precision, recall and F1-score.

The ANN hyperparameters were examined using five-fold stratified cross-validation on the training data. The search considered hidden-layer structures of (8), (16), (16, 8), (32, 16) and (32, 16, 8) neurons; ReLU and tanh activation functions; regularisation values of 0.0001, 0.001 and 0.01; and initial learning rates of 0.001 and 0.005. F1-score was used as the selection criterion because it balances precision and recall. In the unfiltered scenario, the retained (16, 8) ReLU configuration obtained a cross-validation F1-score of 73.34%, compared with 73.85% for the highest-scoring tested architecture. In the

readiness-filtered scenario, it obtained 88.66%, compared with 88.80% for the highest-scoring tested configuration. As these differences were only 0.51 and 0.14 percentage points, respectively, the compact (16, 8) architecture was retained to limit unnecessary model complexity and maintain the same network configuration across both scenarios. The final ANN used the Adam optimiser, $\alpha = 0.0001$, $\text{learning_rate_init} = 0.001$, $\text{max_iter} = 500$ and $\text{random_state} = 42$.

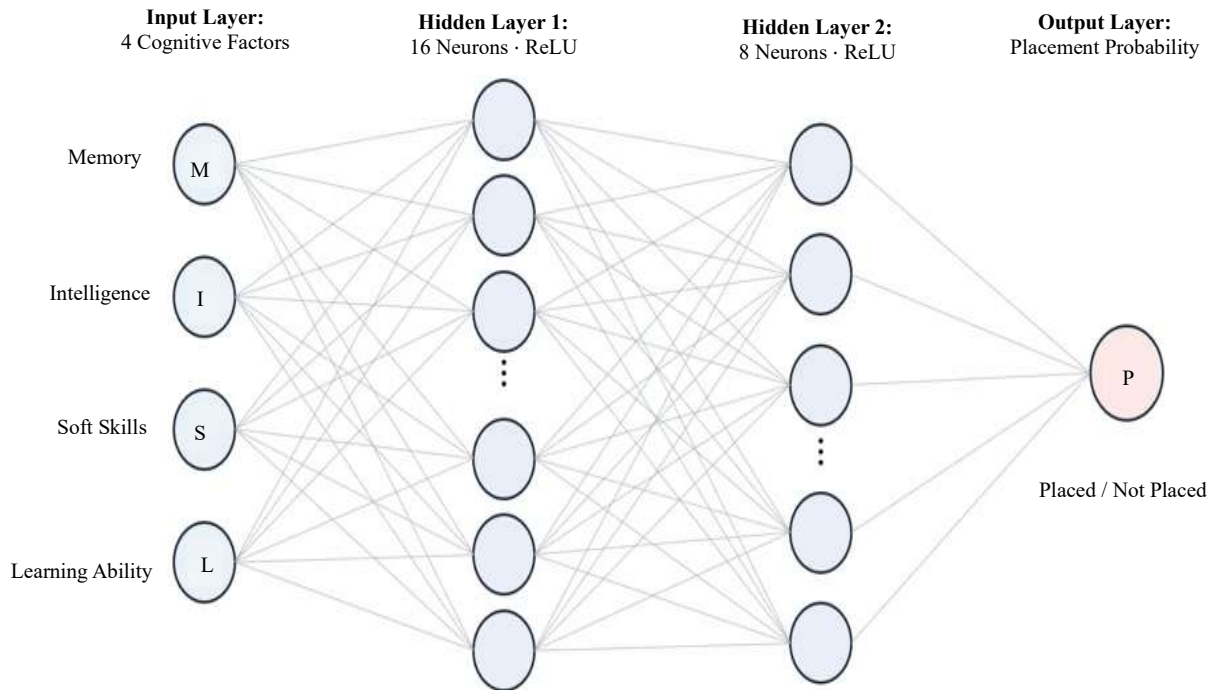


Figure 2. ANN architecture used for placement prediction

Parameter Initialization and Settings: The ANN weights for each layer l are initialized using Glorot (Xavier) normal initialization $W^{(l)} \sim \mathcal{N}(0, \sigma^2)$ with variance $\sigma^2 = \frac{2}{n_{\text{in}} + n_{\text{out}}}$, where n_{in} and n_{out} denote the number of input and output neurons for that layer, respectively. All bias vectors $b^{(l)}$ are initialized to zero. The optimization process uses the Adam optimizer with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$, combined with an initial learning rate $\alpha = 0.001$ and an L_2 regularization parameter $\alpha_{\text{reg}} = 0.0001$ to prevent overfitting. Training is conducted for a maximum of 500 iterations using a fixed random seed ($\text{random_state}=42$) for full reproducibility.

Evaluation Metrics

To quantitatively evaluate the classification performance of the models, predictions are categorized using the standard confusion matrix components (Equations 8-13): True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

1. Accuracy

Accuracy measures the proportion of overall correctly classified student placement records out of the total predictions:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

2. Precision

Precision represents the proportion of correctly predicted placed students out of all students predicted as placed:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

3. Recall (Sensitivity)

Recall evaluates the proportion of actual placed students who were correctly identified by the model:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

4. F_1 -Score

The F_1 -score is the harmonic mean of Precision and Recall, providing a balanced single-metric evaluation under class imbalance:

$$F_1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (11)$$

5. Receiver Operating Characteristic – Area Under Curve (ROC–AUC)

The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) across varying decision thresholds:

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

The ROC–AUC score represents the integral of the ROC curve, measuring the degree of separability between the *Placed* ($y = 1$) and *Not Placed* ($y = 0$) classes:

$$\text{ROC–AUC} = \int_0^1 TPR(FPR) d(FPR) = P(\hat{P}_{\text{Placed}}(x_i) > \hat{P}_{\text{Placed}}(x_j) \mid y_i = 1, y_j = 0) \quad (13)$$

where $\hat{P}_{\text{Placed}}$ is the model's predicted probability for the positive class.

Cognitive Readiness Filtering

The cognitive readiness filter is used to construct a readiness-aligned version of the dataset based on the four operational factors derived in Section 3.2. The filter is applied only to records labelled as *Placed* because the purpose is to examine whether the recorded positive placement outcomes satisfy the readiness criteria defined in this study. The exclusion of a record does not imply that its placement outcome is incorrect; it indicates only that the record does not satisfy the operational readiness threshold used in the filtered experimental scenario.

Mathematical Formulation of the Readiness Filter

A readiness threshold of $\tau = 70$ is applied uniformly to Memory, Intelligence, Soft Skills and Learning Ability. For each student i , the number of factors below the readiness threshold is calculated using equation (14):

$$WC_i = I(M_i < \tau) + I(I_i < \tau) + I(SS_i < \tau) + I(L_i < \tau) \quad (14)$$

Where $WC_i \in \{0,1,2,3,4\}$ is the weakness count for student i , and $1(\cdot)$ is an indicator function that returns 1 when the specified condition is true and 0 otherwise. A value of $WC_i = 0$ indicates that all four cognitive factors are at or above the readiness threshold $\tau = 70$, whereas $WC_i \geq 1$ indicates that at least one cognitive factor is below the readiness threshold.

The record-exclusion decision is defined in equation (15) as follows:

$$D_i = \begin{cases} 1, & \text{if } y_i = 1 \text{ and } WC_i \geq 1, \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

where $D_i = 1$ indicates that the record is excluded from the readiness-filtered dataset, while $D_i = 0$ indicates that it is retained. Accordingly, a Placed student with at least one factor below 70 is excluded. A Placed student with all four factors at or above 70 is retained, and every Not Placed student is retained irrespective of the weakness count.

Modelling Before and After the Cognitive Readiness Filter

The ANN framework described in Section 3.3 is applied in both experimental scenarios. In the unfiltered scenario, the model is trained and evaluated using all placement records. In the filtered scenario, filtering preceded the stratified split, and scaling used training data only.

All architectural, preprocessing and training settings are kept identical across the two scenarios. Therefore, the comparison isolates the effect of the change in dataset composition introduced by the cognitive readiness filter. Differences in accuracy, precision, recall and F1-score are reported to examine how excluding placed records that do not meet the operational readiness threshold affects classifier performance.

Coefficient Sensitivity Analysis

To assess the robustness of the coefficient selection, the project, workshop/certification and internship coefficients were varied simultaneously by $\pm 10\%$. The lower, proposed and higher coefficient settings were (27, 27, 45), (30, 30, 50) and (33, 33, 55), respectively. The readiness threshold, data split, preprocessing procedure, random seed and ANN configuration were held constant. The results are presented in table 5.

Table 5. Coefficient sensitivity analysis of the cognitive readiness filter

Coefficient setting (Projects, Workshops/Certifications, Internships)	Records removed	Post-filter accuracy	Post-filter F1-score
Lower: (27, 27, 45)	2,361	95.75%	91.55%
Proposed: (30, 30, 50)	2,159	94.58%	90.13%
Higher: (33, 33, 55)	1,908	92.65%	87.74%

Across the tested coefficient settings, post-filter accuracy ranged from 92.65% to 95.75%, and F1-score ranged from 87.74% to 91.55%. In every setting, post-filter accuracy remained substantially higher than the corresponding unfiltered accuracy of approximately 78%. The lower coefficients produced a stricter filter and consequently removed more placed records, whereas the higher coefficients retained more records. These results demonstrate that the principal improvement following cognitive readiness filtering is preserved under moderate coefficient variations. The sensitivity analysis is interpreted as evidence of robustness and not as proof that the proposed coefficients are mathematically optimal.

Impact and Fairness Considerations of Readiness Filtering

The readiness filter applies the same factor threshold and decision rule to all placed records without directly using demographic or personal characteristics. Because the dataset does not contain protected attributes such as gender, age, ethnicity, disability or socioeconomic status, formal group-fairness measures could not be calculated. Therefore, the filter is treated only as an experimental readiness-alignment mechanism and not as a criterion for determining student eligibility or denying placement opportunities. Evaluation using datasets containing protected attributes is identified as future work.

The proposed ANN was compared with logistic regression, RBF-SVM, random forest, Extra Trees, histogram-based gradient boosting and a deeper three-hidden-layer MLP (64, 32, 16). All models used identical cognitive-factor inputs, stratified data partitions and evaluation metrics. Hyperparameter selection was conducted using stratified five-fold cross-validation on the training data only. The comparative results are reported in Section 4.3.

The stability of the fixed ANN configuration was evaluated using stratified five-fold cross-validation conducted separately within the unfiltered and readiness-filtered scenarios. Mean performance, standard deviation and 95% confidence intervals based on the Student's *t*-distribution were calculated from the fold-wise results. Pairwise exact McNemar tests were used to compare the ANN with the alternative classifiers using identical test records within each scenario. Holm correction was applied separately within each scenario to account for multiple comparisons. A direct significance test between the unfiltered and filtered scenarios was not conducted because filtering changes the dataset composition.

SHAP analysis examined the predictive contributions of Memory, Intelligence, Soft Skills and Learning Ability to the ANN output. Consistent with previous educational prediction studies [19][22], model-agnostic Kernel SHAP was applied using a weighted k-means representation of the training data as the background distribution. Because the model contained only four inputs, all feature coalitions were evaluated without sampling-based approximation. Shapley values were calculated for the predicted probability of the Placed class. Mean absolute values across the evaluation sample provided global importance scores, while signed values showed whether each factor increased or decreased an individual prediction relative to the expected model output. Local explanations were also generated for representative correctly predicted Placed and Not Placed test records.

Input: Dataset D with 10,000 student records containing raw attributes (CGPA, Aptitude, SoftSkills, Projects, Internships, Certifications, PlacementTraining, PlacementStatus)

1: // Step 1: Cognitive Factor Derivation & Feature Mapping

3: $M_i \leftarrow CGPA_i * 10$ //Memory5: $SS_i \leftarrow SoftSkillsRating_i * 20$ // Soft Skills

7: Construct feature vector $X_i = [M_i, I_i, SS_i, L_i]$

9: // Step 2: Operational Cognitive Readiness Filtering

```

10: Set readiness threshold  $\tau \leftarrow 70$ 

11: Initialize Filtered Dataset  $D_{\text{filtered}} \leftarrow \emptyset$ 

12: For each student record  $(X_i, y_i)$  in  $D$ :

13:   Weakness Count  $WC_i \leftarrow \text{sum}(I(M_i < \tau) + I(I_i < \tau) + I(SS_i < \tau) + I(L_i < \tau))$ 

14:   If  $y_i == 1$  (Placed) AND  $WC_i \geq 1$  Then

15:     Exclude record ( $D_i = 1$ )                                // Exclude placed student with weak factors

16:   Else:

17:     Append  $(X_i, y_i)$  to  $D_{\text{filtered}}$                         // Retain ready placed & all unplaced students

18:   End If

19: End For

20: // Step 3: Model Training & Evaluation

21: Split  $D_{\text{filtered}}$  into Training (80%) and Test (20%) sets using Stratified Sampling

22: Fit Min-Max Scaler on Training set and normalize feature vectors:  $Z_{ij} = (X_{ij} - X_{j\_min}) / (X_{j\_max} - X_{j\_min})$ 

23: Train feed-forward ANN (4 -> 16 [ReLU] -> 8 [ReLU] -> 1 [Sigmoid]) using Adam optimizer (max_iter=500)

24: Evaluate metrics: Accuracy, Precision, Recall, F1-Score, and ROC-AUC on Test set

25: // Step 4: Explainability Analysis

26: Apply Kernel SHAP on trained ANN using weighted k-means background samples

27: Compute feature Shapley values  $\phi_j$  for global importance ranking and individual local student explanations

```

Algorithm 1 details a four-step pipeline that maps raw student attributes into four cognitive factors, excludes inconsistent placed records below threshold $\tau = 70$, trains an ANN, and explains predictions using Kernel SHAP.

RESULTS AND ANALYSIS

ANN Performance Before Cognitive Readiness Filtering

In the first experiment, the ANN was trained and evaluated using the complete unfiltered dataset and the four derived cognitive factors. As presented in table 6, the model achieved an accuracy of 78.00%, precision of 76.08%, recall of 69.37% and F1-score of 72.57%. The unfiltered ANN achieved a ROC-AUC of 0.863, demonstrating good discrimination between the Placed and Not Placed classes. These results establish the baseline performance against which the effect of cognitive readiness filtering is evaluated. The comparatively lower recall indicates that a proportion of placed students were not correctly identified in the unfiltered scenario.

Effect of Cognitive Readiness Filtering on the ANN

In the second experiment, the cognitive readiness filter was applied before ANN training. Placed records with at least one cognitive-factor score below 70 were excluded, whereas placed records satisfying all four readiness criteria and all Not Placed records were retained. The ANN was subsequently retrained and evaluated using the readiness-filtered dataset. Table 6 reports the numerical comparison of ANN performance before and after readiness filtering. Accuracy increased from 78.00% to 94.58%, precision from 76.08% to 85.65%, recall from 69.37% to 95.10% and F1-score from 72.57% to 90.13%. ROC-AUC also increased from 0.863 to 0.981. Figure 3 illustrates changes in the four-classification metrics. Recall showed the largest increase, indicating that more Placed students were correctly identified in the filtered cohort. These results reflect the change in dataset composition introduced by the readiness criteria.

Table 6. ANN performance before and after cognitive readiness filtering

Metric	Before filtering	After filtering
Accuracy	78.00	94.58
Precision	76.08	85.65
Recall	69.37	95.10
F1-score	72.57	90.13

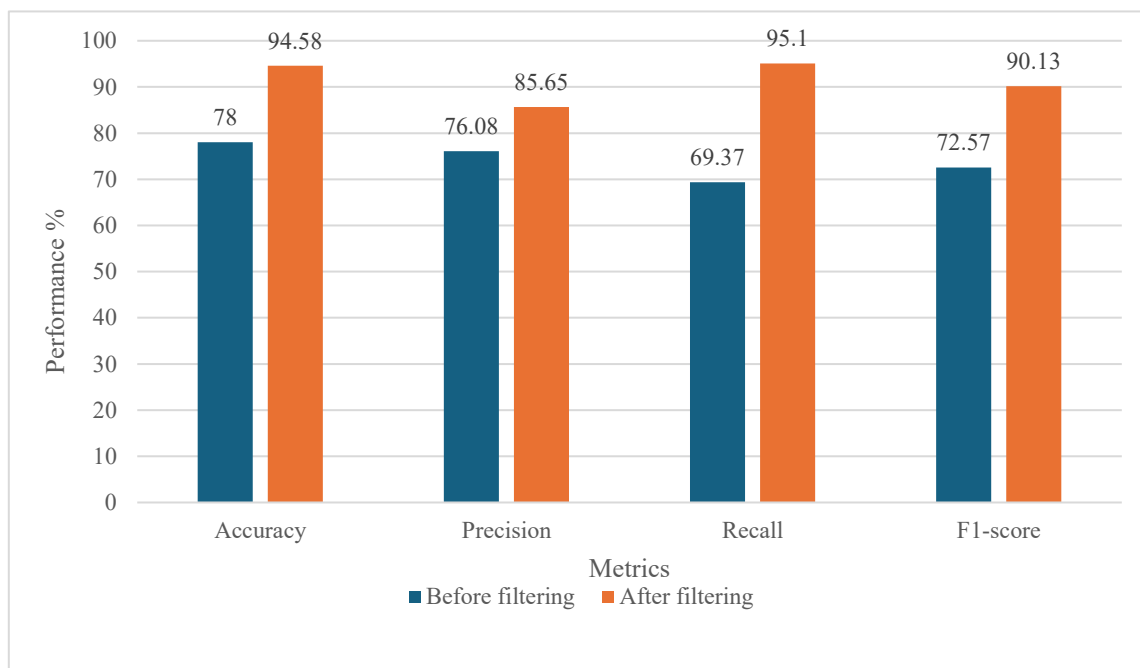


Figure 3. ANN performance before and after cognitive readiness filtering (accuracy, precision and recall and F1-Score; blue = before filtering, orange = after filtering)

Comparison with Other Classifiers

The proposed ANN was compared with conventional and advanced models using identical cognitive-factor inputs, stratified data partitions and evaluation metrics. As shown in table 7, all models improved after readiness filtering. The proposed ANN achieved the highest post-filter accuracy and F1-score, while the advanced tree ensembles and deeper MLP produced closely comparable results. This indicates that the performance improvement was mainly associated with readiness filtering rather than a specific model family.

Table 7. Performance of conventional and advanced models before and after cognitive readiness filtering

Model	Accuracy before (%)	Precision before (%)	Recall before (%)	F1-score before (%)	Accuracy after (%)	Precision after (%)	Recall after (%)	F1-score after (%)
Proposed ANN (16, 8)	78.00	76.08	69.37	72.57	94.58	85.65	95.10	90.13
Logistic Regression	77.90	74.42	72.11	73.25	93.63	87.02	88.73	87.86
RBF-SVM	78.00	76.08	69.37	72.57	94.26	84.57	95.34	89.63
Random Forest	78.05	75.58	70.44	72.92	94.46	85.27	95.10	89.92
Extra Trees	78.20	75.28	71.51	73.35	94.20	85.30	93.87	89.38
Histogram Gradient Boosting	78.05	75.45	70.68	72.98	94.39	85.40	94.61	89.77
Deep MLP (64, 32, 16)	77.55	75.79	68.30	71.85	94.39	85.71	94.12	89.72

Global SHAP analysis of Cognitive Factors

Kernel SHAP was applied to the readiness-filtered ANN to interpret predictions for the Placed class. Absolute SHAP magnitudes were averaged across students to obtain global importance scores. Table 8 reports the resulting mean |SHAP| values. Figure 4(a) illustrates the corresponding global ranking of the cognitive factors.

Table 8. Global SHAP importance of cognitive factors in the filtered ANN scenario (class placed)

Cognitive factor	Mean SHAP value
Learning Ability	0.2257
Intelligence	0.0705
Soft Skills	0.0203
Memory	0.0113

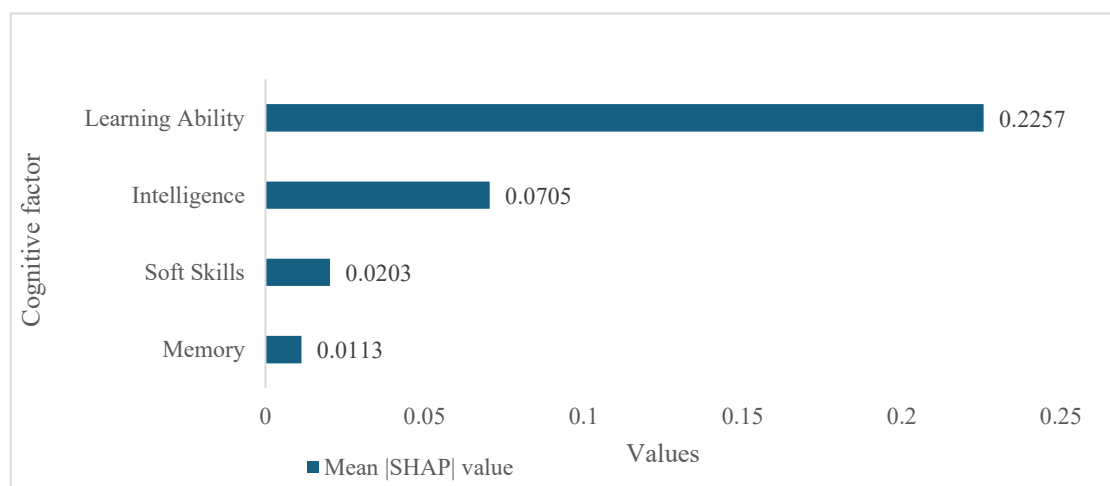


Figure 4(a). Global SHAP importance of cognitive factors in the filtered scenario

The ranking is quite distinctive. Learning Ability has by far the largest mean contribution, being over three times larger than Intelligence and much higher than the contributions of Soft Skills and Memory. This indicates that Learning Ability, derived from projects, internships, workshops or certifications and placement training, made the largest contribution to the ANN's placement predictions. Intelligence, derived from aptitude-test performance and academic achievement, ranked second, while Soft Skills and Memory showed comparatively smaller predictive contributions. Overall, the SHAP results show that

the ANN relied primarily on Learning Ability and Intelligence when predicting placement, while Soft Skills and Memory made smaller contributions. These findings provide model-level interpretability and should be understood as predictive contributions rather than evidence of causal effects on placement outcomes.

Local SHAP Explanations

Figure 4(b) provides local SHAP explanations for two correctly classified students. For the Placed student (Memory = 84.0, Intelligence = 87.0, Soft Skills = 82.0 and Learning Ability = 87.5), the respective SHAP contributions were +0.041, +0.175, -0.022 and +0.423, moving the baseline Placed probability from 0.271 to 0.888. For the Not Placed student (Memory = 73.0, Intelligence = 69.5, Soft Skills = 86.0 and Learning Ability = 45.0), the corresponding contributions were -0.004, -0.094, -0.009 and -0.164, reducing the probability from 0.271 to less than 0.001. Thus, Learning Ability and Intelligence were the dominant case-level contributors. These values describe predictive contributions rather than causal effects.

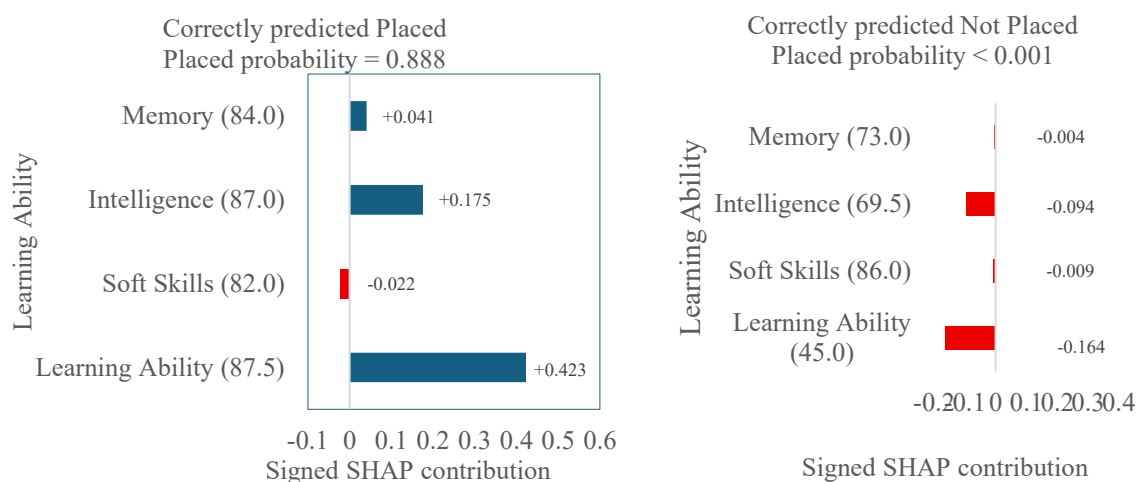


Figure 4(b). Local SHAP explanations for representative correctly predicted placed and not placed student's positive values increase and negative values decrease the predicted probability of the placed class

Cross-Validation Stability

To evaluate performance stability, stratified five-fold cross-validation was conducted separately for the unfiltered and readiness-filtered scenarios. Accuracy, precision, recall and F1-score were calculated for each fold. Table 9 reports the means, standard deviations and 95% confidence intervals calculated using Student's t-distribution.

Table 9. Five-fold cross-validation results for the ANN

Scenario	Metric	Mean (%)	SD (%)	95% CI (%)
Unfiltered	Accuracy	78.33	0.89	77.23–79.43
Unfiltered	Precision	76.42	1.33	74.76–78.07
Unfiltered	Recall	69.98	1.43	68.20–71.76
Unfiltered	F1-score	73.05	1.12	71.66–74.44
Readiness-filtered	Accuracy	93.80	0.81	92.80–94.81
Readiness-filtered	Precision	83.64	2.28	80.81–86.48
Readiness-filtered	Recall	94.75	0.66	93.93–95.57
Readiness-filtered	F1-score	88.84	1.30	87.22–90.46

Figure 5 presents the fold-wise accuracy and F1-score of the readiness-filtered ANN. Mean accuracy was 93.80% (95% CI: 92.80–94.81%), while mean F1-score was 88.84% (95% CI: 87.22–90.46%). The limited fold-to-fold variation indicates stable performance.

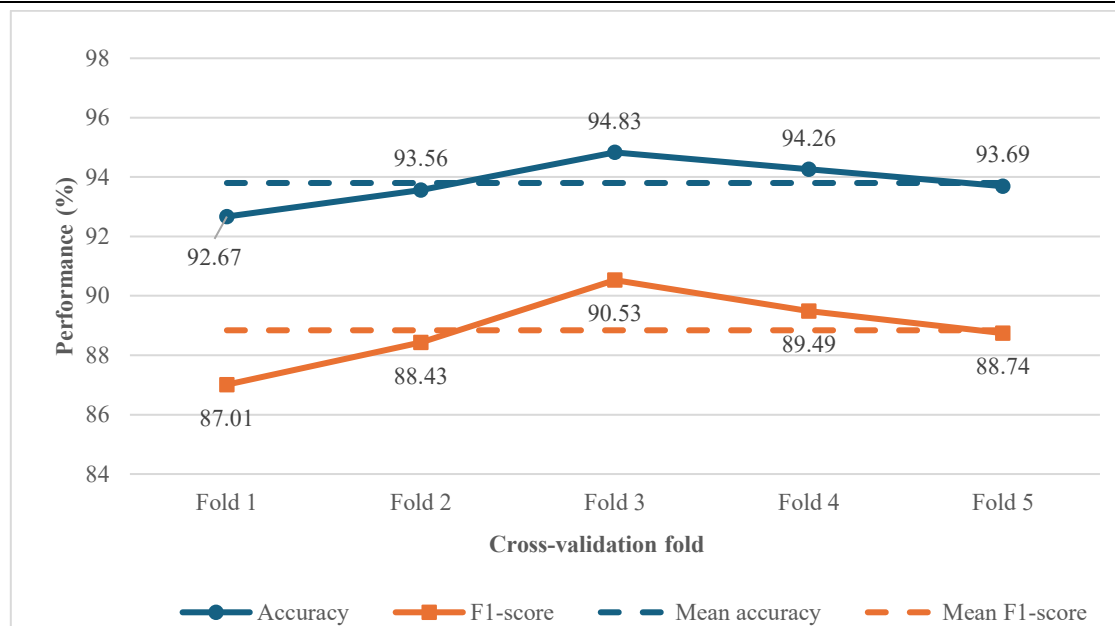


Figure 5. Five-fold cross-validation accuracy and F1-score of the readiness-filtered ANN. Dashed lines indicate the corresponding mean performance

Statistical Significance Testing

Pairwise exact McNemar tests were used to compare ANN predictions with three conventional classifiers using identical test records within each scenario. Holm correction was applied separately within each scenario to account for multiple comparisons. As shown in table 10, none of the differences remained statistically significant after correction ($p > 0.05$). Thus, the ANN demonstrated competitive performance, but the results do not establish its statistical superiority over the alternative classifiers.

Table 10. Pairwise McNemar test results for the ANN and three selected conventional classifiers

Scenario	Comparison	Raw p -value	Holm-adjusted p -value	Significant
Unfiltered	ANN vs. Logistic Regression	0.8854	1.0000	No
Unfiltered	ANN vs. SVM	1.0000	1.0000	No
Unfiltered	ANN vs. Random Forest	1.0000	1.0000	No
Readiness-filtered	ANN vs. Logistic Regression	0.0489	0.1466	No
Readiness-filtered	ANN vs. SVM	0.2669	0.5337	No
Readiness-filtered	ANN vs. Random Forest	0.8238	0.8238	No

Ablation Study

To evaluate the contribution of each individual design component within the proposed framework, an ablation study was conducted by systematically removing key modules. Table 11 summarizes the impact on classification performance and model interpretability.

Table 11. Ablation study of framework components

Configuration / Removed Component	Accuracy (%)	F1-Score (%)	ROC-AUC	SHAP Explanations Available
Full Framework (Proposed)	94.58	90.13	0.981	Yes
<i>w/o Cognitive Factor Derivation (Raw Attributes Input)</i>	93.85	88.70	0.965	Partial (Low-level features)
<i>w/o Cognitive Readiness Filter (Unfiltered Dataset)</i>	78.00	72.57	0.863	Yes
<i>w/o ANN Nonlinearity (Linear Logistic Regression Baseline)</i>	93.63	87.86	0.961	Yes (Model-agnostic SHAP)
<i>w/o SHAP Analysis</i>	94.58	90.13	0.981	No (Black-box model)

DISCUSSION

The readiness-filtered scenario increased ANN accuracy from 78.00% to 94.58% and F1-score from 72.57% to 90.13%. Similar gains across the comparison models suggest that the operational readiness criteria produced a dataset with clearer class separation and that the observed improvement was not attributable solely to the ANN. Five-fold cross-validation demonstrated stable readiness-filtered performance, with a mean accuracy of 93.80% (95% CI: 92.80–94.81%) and a mean F1-score of 88.84% (95% CI: 87.22–90.46%). However, no pairwise ANN comparison remained statistically significant after Holm correction. The ANN should therefore be regarded as competitive rather than universally superior to the alternative classifiers. Learning Ability produced the largest mean absolute SHAP value (0.2257), consistent with evidence linking internships and practical learning activities to employability [14][15]. Intelligence ranked second (0.0705), supporting the relevance of aptitude and academic performance reported in placement research [2][3]. Its greater contribution than Memory (0.0113) suggests that combining aptitude with academic attainment provided more predictive information than CGPA alone. Soft Skills made a comparatively smaller contribution (0.0203), despite the importance of communication and interpersonal competence reported previously [11][13]. This difference may reflect limited variation in the available rating or interactions with the other cognitive factors. Local SHAP explanations further demonstrated how individual factor values contributed to Placed and Not Placed predictions, extending educational SHAP applications to placement readiness [19][22]. These contributions represent predictive associations and should not be interpreted as causal effects.

LIMITATIONS, IMPLICATIONS AND FUTURE SCOPE

The use of a single public dataset limits the external validity of the findings across institutions, academic disciplines and recruitment contexts. The cognitive factors were operationally constructed from the available attributes and may not capture the full breadth of the corresponding cognitive or psychometric concepts. The absence of protected demographic variables also prevented formal group-fairness assessment. Furthermore, the readiness filter changes the dataset composition; consequently, comparisons between the two scenarios do not represent performance on an unchanged student population.

The framework may assist institutions in identifying academic, aptitude, interpersonal and experiential areas requiring student development. Global and local SHAP explanations provide interpretable evidence that can support targeted placement preparation. Nevertheless, the framework should be used only as a decision-support tool and not as an independent mechanism for determining placement eligibility.

Future research should evaluate the framework using multi-institutional and longitudinal datasets and externally validate the operational cognitive-factor formulation. Data-driven selection of factor coefficients and readiness thresholds should also be investigated. The inclusion of protected demographic attributes would enable formal fairness assessment, while model calibration and counterfactual explanations could provide more reliable and actionable student-level guidance.

CONCLUSION

This study developed an interpretable cognitive-readiness-based framework for predicting campus-placement outcomes by transforming seven observable student attributes into four operational factors: Memory, Intelligence, Soft Skills and Learning Ability. These factors were used as inputs to an artificial neural network and evaluated under unfiltered and readiness-filtered experimental scenarios. On the unfiltered test set, the ANN achieved 78.00% accuracy, an F1-score of 72.57% and a ROC–AUC of approximately 0.86. After applying the predefined readiness rule, the ANN achieved 94.58% accuracy, an F1-score of 90.13% and a ROC–AUC of 0.981. The removal of 2,159 historically placed records that failed at least one operational readiness criterion produced a more clearly separated prediction setting; however, it also changed the dataset composition and should not be interpreted as correcting erroneous placement labels. The $\pm 10\%$ coefficient-sensitivity analysis indicated that the main findings remained stable under moderate variations in the predetermined scale-conversion coefficients. Five-fold cross-validation further demonstrated consistent readiness-filtered performance, with a mean accuracy of 93.80% (95% CI: 92.80–94.81%) and a mean F1-score of 88.84% (95% CI: 87.22–90.46%). Performance improvements across the alternative classifiers indicate that the readiness criteria, rather than the ANN alone, contributed to the increased class separability. Importantly, although the ANN yielded the highest nominal post-filter accuracy, pairwise McNemar tests with Holm correction established no statistically significant superiority over classical baselines such as Logistic Regression, RBF-SVM, or Random Forest, underscoring that the neural network operates as a statistically competitive alternative rather than a definitively superior one. Kernel SHAP provided global and student-level explanations of the ANN predictions. Learning Ability and Intelligence produced the largest predictive contributions, followed by Soft Skills and Memory. The framework therefore provides an understandable decision-support approach for identifying placement-related strengths and areas requiring focused preparation. Nevertheless, the findings are limited by the use of a single public dataset, operationally constructed factors, altered cohort composition and the absence of demographic attributes for fairness evaluation. Future research should examine multi-institutional and longitudinal data, data-driven coefficient and threshold selection, formal fairness assessment, probability calibration, counterfactual explanations and external generalisability.

REFERENCES

- [1] Dake DK. Information technology graduates employability prediction model in a low-income country using tree-based machine learning classifiers. Discover Global Society. 2025 Nov 26;3(1):158. <https://doi.org/10.1007/s44282-025-00304-3>
- [2] Molla MJ, Obaidullah SM, Sen S, Weber GW, Jana C. Developing a predictive model for engineering graduates placement using a data-driven machine learning approach. Journal of applied research on industrial engineering. 2024 Dec 1;11(4):536-59. <https://doi.org/10.22105/jarie.2024.468839.1641>
- [3] Maurya LS, Hussain MS, Singh S. Developing classifiers through machine learning algorithms for student placement prediction based on academic performance. Applied Artificial Intelligence. 2021 May 12;35(6):403-20. <https://doi.org/10.1080/08839514.2021.1901032>
- [4] Chen LS, Huynh-Cam TT, Nguyen VC, Lu TC, Le-Huynh DK. Predicting early employability of Vietnamese graduates: Insights from data-driven analysis through machine learning methods. Big Data and Cognitive Computing. 2025 May 19;9(5):134. <https://doi.org/10.3390/bdcc9050134>
- [5] Ruparel M, Swaminarayan P. Enhanced student placement prediction using machine learning: A comparative evaluation of algorithms. International Journal of Engineering Trends and Technology. 2025 Jan;73(1):225-36. <https://doi.org/10.14445/22315381/IJETT-V73I1P119>
- [6] ElSharkawy G, Helmy Y, Yehia E. Employability prediction of information technology graduates using machine learning algorithms. International Journal of Advanced Computer Science and Applications. 2022;13(10):359–367. <https://doi.org/10.14569/IJACSA.2022.0131043>
- [7] Kumar M, Walia N, Bansal S, Kumar G, Cengiz K. Predicting college students' placements based on academic performance using machine learning approaches. International Journal of Modern Education and Computer Science. 2023;15(6):1–13. <https://doi.org/10.5815/ijmecs.2023.06.01>
- [8] Bayardalai U, Lee JH. Predicting employment status and types of university graduates in South Korea using machine-learning techniques. Industrial Engineering and Management Systems. 2025;24(2):265-81. <https://doi.org/10.7232/iems.2025.24.2.265>

- [9] Mezhoudi N, Alghamdi R, Aljunaid R, Krichna G, Düşteğör D. Employability prediction: a survey of current approaches, research challenges and applications. *Journal of Ambient Intelligence and Humanized Computing*. 2023 Mar;14(3):1489-505. <https://doi.org/10.1007/s12652-021-03276-9>
- [10] Nouib H, Qadech H, Benatiya Andaloussi M, Chowdhury SJ, Moumen A. Predicting graduate employability using hybrid AHP-TOPSIS and machine learning: a Moroccan case study. *Technologies*. 2025 Sep 1;13(9):385. <https://doi.org/10.3390/technologies13090385>
- [11] Amarathunga B, Khatibi A, Talib ZM, Azam SF, Tham J. Graduate employability skills, trending avenues and research gaps: a systematic literature review and bibliometric analysis. *Asian Education and Development Studies*. 2024 Aug 23;13(4):320-39. <https://doi.org/10.1108/AEDS-04-2024-0085>
- [12] Eden S. Out of the comfort zone: enhancing work-based learning about employability through student reflection on work placements. *Journal of Geography in Higher Education*. 2014;38(2):266–276. <https://doi.org/10.1080/03098265.2014.911826>
- [13] Cseh Papp I, Molnar C, Juhász T. Soft skills of business students in relation to higher education internships. *Problems and Perspectives in Management*. 2023;21(4):113-26. [https://doi.org/10.21511/ppm.21\(4\).2023.09](https://doi.org/10.21511/ppm.21(4).2023.09)
- [14] Baert S, Neyt B, Siedler T, Tobback I, Verhaest D. Student internships and employment opportunities after graduation: A field experiment. *Economics of Education Review*. 2021 Aug 1;83:102141. <https://doi.org/10.1016/j.econedurev.2021.102141>
- [15] Pianda D, Hilmiana H, Widiyanto S, Sartika D. The impact of internship experience on the employability of vocational students: a bibliometric and systematic review. *Cogent Business & Management*. 2024 Dec 31;11(1):2386465. <https://doi.org/10.1080/23311975.2024.2386465>
- [16] Di Pietro G. International internships and skill development: A systematic review. *Review of Education*. 2022 Aug;10(2):e3363. <https://doi.org/10.1002/rev3.3363>
- [17] Ndamase M, Lukman Y. The Impact of the Internship Programme on Students in a Selected Public Higher Institution in the Eastern Cape, South Africa. *Research in Social Sciences and Technology*. 2024;9(2):246-61. <https://doi.org/10.46303/ressat.2024.34>
- [18] Anjum S. Impact of internship programs on professional and personal development of business students: a case study from Pakistan. *Future business journal*. 2020 Jan 10;6(1):2. <https://doi.org/10.1186/s43093-019-0007-3>
- [19] Nnadi LC, Watanobe Y, Rahman MM, John-Otumu AM. Prediction of students' adaptability using explainable AI in educational machine learning models. *Applied Sciences*. 2024 Jun 13;14(12):5141. <https://doi.org/10.3390/app14125141>
- [20] Johora FT, Hasan MN, Rajbongshi A, Ashrafuzzaman M, Akter F. An explainable AI-based approach for predicting undergraduate students academic performance. *Array*. 2025 Jul 1;26:100384. <https://doi.org/10.1016/j.array.2025.100384>
- [21] Bettahi A, Belouadha FZ, Harroud H. A modular and explainable machine learning pipeline for student dropout prediction in higher education. *Algorithms*. 2025 Oct 18;18(10):662. <https://doi.org/10.3390/a18100662>
- [22] Islam MM, Sojib FH, Mihad MF, Hasan M, Rahman M. The integration of explainable AI in Educational Data Mining for student academic performance prediction and support system. *Telematics and Informatics Reports*. 2025 Jun 1;18:100203. <https://doi.org/10.1016/j.teler.2025.100203>
- [23] Mai J, Wei F, He W, Huang H, Zhu H. An explainable student performance prediction method based on dual-level progressive classification belief rule base. *Electronics*. 2024 Nov 6;13(22):4358. <https://doi.org/10.3390/electronics13224358>
- [24] Rainio O, Teuho J, Klén R. Evaluation metrics and statistical tests for machine learning. *Scientific reports*. 2024 Mar 13;14(1):6086. <https://doi.org/10.1038/s41598-024-56706-x>
- [25] Ziwei M, Han LL. Scientometric review of sustainable land use and management research. *Aquatic Ecosystems and Environmental Frontiers*. 2023 Nov 30;1(1):21-4. <https://doi.org/10.70102/AEEF/V1I1/5>
- [26] Ghorab M. The Socio-Economic Effects of Globalization on Labor Markets in Developing Countries. *IAJSS*. 2025 Sep. 29;12(3):1-7. <https://doi.org/10.71086/IAJSS/V12I3/IAJSS1209>
- [27] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI. From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence*. 2020 Jan;2(1):56-67. <https://doi.org/10.1038/s42256-019-0138-9>
- [28] García A, Hernández V. Sustainable finance models for addressing climate change impacts on emerging economies. *International Academic Journal of Economics*. 2024;11(1):11–19.