Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 47 (2026), No. 5-B, pp. 1989–2001
DOI : 10.47974/JIOS-2290

Predicting IT graduate student's employability using machine learning and deep learning hybrid approaches

Ankita Chopra *
Department of Engineering and Technology
Jagannath University
Jaipur 302022
Rajasthan
India

Madan Lal Saini †
Department of Computer Science & Engineering
Apex Institute of Technology
Chandigarh University
Mohali 140413
Punjab
India

Vivek Kumar Sharma §
Department of Engineering and Technology
Jagannath University
Jaipur 302022
Rajasthan
India

Abstract

The growing demand for skilled professionals highlights the need for accurate prediction of employability and placement success among engineering students. Traditional methods of assessing placement success often rely on static parameters like academic percentage, written test score, interviews, and resume information. These methods fail to capture the dynamic interdependencies between academic performance, technical abilities, interpersonal skills, analytical problem-solving, critical analysis, and decision-making capabilities. This study proposes machine learning and deep learning-based approach to

* *E-mail:* **chopraankita3007@gmail.com (Corresponding Author)**
† *E-mail:* **mlsaini@gmail.com**
§ *E-mail:* **vivek.sharma@jagannathuniversity.org**

address these limitations, leveraging data-driven techniques to enhance prediction accuracy and also helps in identifying the important graduate feature for placement success.

The proposed methodology integrates diverse datasets, including academic records, aptitude marks, soft skills assessment marks, and behavioural attributes. After data pre-processing 32 feature are taken to train and validate the proposed model. Findings reveal that the proposed model attains an 81.7% prediction accuracy, surpassing conventional approaches such as LR, RF, and GBM. Moreover, Feature importance analysis emphasizes the significance of the dominant role of specific parameters, such as technical skills and aptitude competency, in influencing employability outcomes.

This study underscores the potential of machine learning and deep learning techniques in forecasting placement success based on their skills. Education institutes can identify skill gaps in students who are less likely to be placed and they can run specific training programs, workshops, or certifications to boost employability.

Subject Classification: *68T07, 68T05, 62H30.*

Keywords: *Employability prediction, Placement success, Academic scores, Technical skills, Interpersonal skills, Analytical problem-solving, Critical analysis, Decision-making capabilities.*

1. Introduction

The facts generation area has emerged as a prime driving force of financial boom and improvement in India. Advances in computing, connectivity, and software program and facts structures are swiftly remodelling companies and society. This exponential tempo of generation evolution has brought about skyrocketing call for professional IT experts in domain names like software engineering, artificial intelligence, synthetic intelligence, cybersecurity, and cloud computing, finance, medical, banking, manufacturing, marketing, etc. However, current years have visible mounting worries concerning the employability of IT graduates in India. Employability refers back to the cap potential of graduates to benefit preliminary employment of their subject of look at and feature a hit careers. According to Bai and Hira [1], engineering graduates in India face acute talent gaps and a loss of activity readiness, with handiest 20-25% taken into consideration quite simply employable via way of means of the industry.

Employability skills encompass both general skills needed for most jobs and those that enable individuals to excel in their careers Kapil [3]. While job-specific skills will always be sought after by employers, they also place great importance on general skills. The general skills desired by employers include effective communication (both written and oral), teamwork, problem-solving abilities, basic numeracy, leadership qualities,

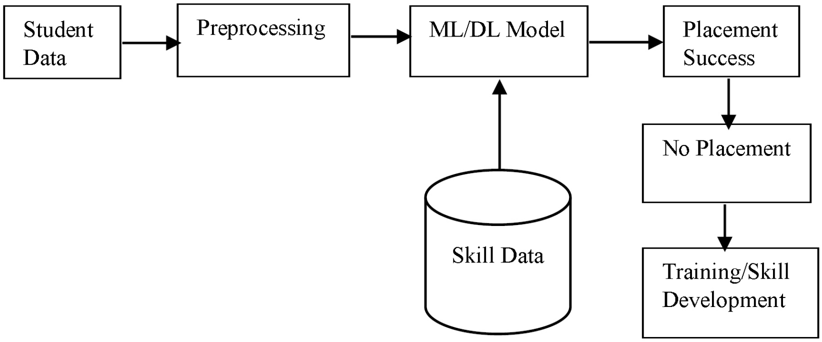

Figure 1

Student's employability prediction system flowchart

adaptability to new situations and technologies, creativity, and proficiency in IT, including familiarity with MS Office and computer hardware. A general model for employability prediction is presented in Figure 1, which is divided into 4 phases. The first phase is data acquisition in which data are collected from various college/universities' placement cells. Data pre-processing is performed in the second phase and important features are extracted. In the third phase model is trained and tested and the fourth phase gives the prediction results for placed and not placed students.

1.1 *Problem Statement and Research Objective*

The growing demand for skilled professionals highlights the need for accurate prediction of employability and placement success among engineering students. Traditional methods of assessing placement success often rely on static parameters of student performance which fail to capture the dynamic interdependencies between academic performance, technical skills, interpersonal abilities, analytical skills, problem solving capabilities, critical reasoning, and decision-making skills. The proposed method uses real data that is collected from various colleges/universities' placement cells. The study proposes a machine learning and deep learning-based hybrid approach to predict IT students' employability. Some students do not have good academic score but they are good at other skills. The dataset contains 41 features which includes 30 feature apart from academic scores and makes the model more accurate.

To develop and evaluate a predictive model that accurately assesses the employability of IT graduates based on academic scores, analytical skills, problem solving skills, technical skills, internships, and soft skill indicators, with the goal of identifying key factors influencing successful job placement.

2. Literature review

Engineering education in India produces over 1.5 million graduates annually, yet employability remains a significant concern, with only 20–25% of graduates deemed readily employable by the industry Mezhoudi et al. [5]. Several research have explored elements affecting employability and predictive fashions for placement achievement. This segment provides an evaluation of applicable literature, which specialize in instructional achievement prediction, elements impacting employability, and upgrades via predictive modeling.

Gupta et al. [6] proposed a method for prediction salary and placement using logistic regressing, SVM, RF and applied ridge and lasso regularization techniques. They analysed the overall employability of engineering graduate and predicted the domain of employment and salary range. Gokuladas [7] identified parameters for employability prediction of undergraduate engineering students in college/university campus placement process. They analysed the technical and non-technical skills of students and concluded that how other parameters plays role in placements apart from academic performance of a student.

Vinutha and Yogisha [8] analysed the issues of not getting placements and proposed a two-step strategy for improving the placements. They used ANN, RF, Naïve Bayes, and Xgboost and got highest accuracy for RF algorithm. Thavasi and Revathi [9] designed a job prediction system on the basis on their academic performance, course outcome and other required skills. The proposed system needs the required skills for job role and identify the suitable candidates for the said job role. They used KNN, Naïve Bayes, and SVM for their prediction system and got an accuracy of 93.33 and 98% precision.

Patacsil and Tablatin [11] used LSTM networks to forecast employment outcomes, demonstrating advanced overall performance over conventional models and diagnosed graduate attributes along with teamwork, leadership, and conversation as vital to employability. Wei et al. [12] proposed a feature selection prediction model for employment on basis of improved bat algorithm and SVM. Their study includes student's leader experience, family situation, and other relevant characteristics that can affect employment outcomes. They analyzed activity commercials to decide important smooth abilities for entry-stage job roles.

Tu et al. [13] developed an adaptive support vector machine framework for predicting college students' entrepreneurial intention. They collected data from planning and development cell of their university personnel training department and incorporated dependent and unstructured statistics the usage of deep getting to know, accomplishing better accuracy in predictive fashions. This aligns with the prevailing study`s goal of leveraging synthetic neural networks (ANN) to decorate employability predictions. Qostal, Moumen, and Lakhrissi [14] examined various deep learning approaches for resume classification and extracting information from it. They used GRU, LSTM and CNN and compared these models for text data. Zayed, Salman, and Hasasneh [15] designed a system that recommends graduate program to students on the basis of student's characteristics and academic interest. Alyahyan and Düştegör [2] reviewed academic success prediction models in higher education. Krishnaiah and Kadegowda [4] proposed hybrid machine learning approaches for employability prediction. Koen, Klehe, and Vianen [10] discussed employability challenges among graduates. Tu et al. [13] developed an adaptive SVM-based prediction framework. Singh, Saini, and Savita [16] analysed activation functions in classification models, while Chopra and Saini [17, 19] explored neural network-based employability prediction approaches

Recent studies published in journals of Taru Publications have also explored machine learning-based predictive modelling in educational analytics and optimization domains. These studies highlight the growing importance of AI-driven approaches [18] in improving prediction accuracy and decision-making in academic and professional settings [21–25].

3. Proposed Methodology

This section offers a comprehensive explanation of the methodology employed in this paper to forecast the likelihood of job placements for engineering students, taking into account their employability skills. The methodology involved studying the university system offering IT courses, collecting important information, data preparation, developing an artificial neural network model, analysing the model results, and evaluating the overall methodology.

3.1 *Source and Collection of Data*

In this phase, we accumulated applicable facts from 3 important sources: graduates, recruiters, and located students. Data is collected from the reputed colleges/universities' placement cell. Each supply of facts affords treasured insights into the elements that affect employability and site success, and together, they shape a complete dataset for our analysis. The data set has 20,000 records and it can be accessed by GitHub repository link provided in the reference [20].

3.2 *Data Pre-processing*

The key steps concerned in pre-processing for the employability prediction dataset are removing Irrelevant Features: Attributes that don't make a contribution to employability prediction (together with recruiter names, Student IDs, and non-relevant features) had been eliminated to lessen noise and enhance version efficiency. The dataset consists of specific variables together with Placement Outcome (Placed / Not Placed) and recruiter ratings (High, Medium, and Low). The Placement Outcome variable, which has handiest categories, turned into label-encoded as: Placed (1), Not Placed (0). To make sure that numerical capabilities along with aptitude, communication, technical subject marks, and personal rankings make contributions similarly to the model, we have to apply normalization techniques. Before training the model, the dataset was shuffled to eliminate any order-based bias. After pre-processing, the final dataset contains 20000 records and 38 relevant features.

3.3 *Experimental Set Up*

The machine learning model was developed and run on the subsequent system specifications: Processor: Intel i7, 11th Gen, RAM: 16GB DDR4, Storage: 512GB SSD, OS: Windows 10. The implementation was carried out using the following tools: Jupyter Notebook, Programming Language: Python 3.12. The employability prediction system utilizes the following Python libraries: pandas – for handling tabular data, numpy – for numerical computations, scikit-learn – for dataset splitting, pre-processing, machine learning algorithm, tensorflow / keras – for building and training the ANN model, scikit-learn – for label encoding and dataset splitting, matplotlib – for plotting performance metrics, seaborn – for visualizing feature correlations, joblib – for model saving/loading

3.4 *Development of ANN Model*

The input layer of the ANN accepts skills data as input features, consisting of four neurons, each representing one of the employability skills: Aptitude, Communication, Technical, and Personality. The hidden layers are essential for feature extraction and nonlinear transformation of the input data, with this model featuring two hidden layers. With the pre-processed data, an ANN model was created to estimate the likelihood of placement based on skills data. The feature set has 38 features and these feature are selected after data processing (correlation and heat-map). The ANN model architecture and its parameters are presented in Table 1.

Table 1

ANN Model Architecture and its Parameters

Parameter	Value	Parameter	Value
Input_size	(16000, 38)	Output_size	(16000, 1)
Hidden_layers	[64, 32]	Activation	'relu'
Output_activation	sigmoid	Loss_function	binary_crossentropy
Optimizer	'adam'	Learning_rate	0.001
Batch_size	32	Epochs	100
Dropout_rate	0.2	Weight_initialization	'glorot_uniform'
Bias_initializer	'zeros'	Regularization	None
Validation_split	0.1	Shuffle	True

3.5 *Development of Logistic Regression, Random Forest, and GBM*

Logistic Regression, which is a baseline model for binary classification tasks. After cleaning the data, we encoded the categorical features using one-hot encoding and scaled the numerical variables then split the dataset into training and testing sets. The model was trained and evaluated using metrics like accuracy, precision, recall, and ROC-AUC. Random Forest Classifier being an ensemble of decision trees, is more powerful in handling complex patterns in the data. Since it's not very sensitive to scaling so we didn't performed scaling but encoded the categorical variables. Using Random Forest Classifier from scikit-learn, we trained the model and observed significantly better performance than logistic regression, particularly in handling nonlinear relationships. We explored the feature importance scores to identify which student attributes were most predictive of employment. Finally, we implemented Gradient Boosting Machine (GBM), which is known for its high predictive power. Gradient Boosting Classifier is available in scikit-learn, we tuned parameters like learning rate, number of estimators to avoid overfitting. GBM builds trees sequentially with each new tree trying to correct the errors of the previous ones. However this is more computationally expensive and requires careful tuning. This model gave good accuracy on the test set.

4. Results & Discussions

The Artificial Neural Network (ANN) model was developed in Python utilizing the Keras library. The model was trained using resilient backpropagation with weight backtracking as the optimization approach. This technique updates the weights and biases of the network based on the gradient of the error, similar to conventional backpropagation. However, it focuses solely on the sign of the gradient to determine the direction of weight changes, ignoring its magnitude. Additionally, a backtracking mechanism is included to revert to a previous state if there is a substantial rise in error, which improves the convergence speed during training.

The training dataset comprised placement records for 20,000 engineering students, which were divided into training, validation, and testing sets in a 60:20:20 split. Each record contained scores for the employability skills 38 attributes along with the placement outcome (1 indicating placement and 0 indicating non-placement). The model underwent training for 100 epochs, utilizing a learning rate of 0.001 and a momentum term of 0.9. The momentum term aids in achieving faster convergence by incorporating a portion of the previous weight update into the current adjustment. These hyperparameter values were determined through preliminary experimentation and an analysis of the performance on the validation set.

4.1 *Model Evaluation*

The trained artificial neural network (ANN) model underwent evaluation using a test set comprising 16000 previously unseen records. The performance was assessed using the following metrics: Accuracy, Precision, Recall, and F1-score. Accuracy measures the overall correctness of predictions, precision assesses the trustworthiness of positive predictions, and recall evaluates the model's capacity to recognize positive instances. The results of the model assessment are presented in Table 2.

The accuracy of 82.4% suggests that the model can make overall correct predictions in a majority of cases. The precision of 80.6% indicates that when the model predicts placement, there is an 80.6% chance that the student is actually placed. The recall of 84.7% means the model correctly identifies 84.7% of students who got placed.

Table 2

Performance metrics for ANN model on test set

Metric	Value
Accuracy	82.4%
Precision	80.6%
Recall	84.7%
F1-Score	82.6%

4.2 *Cross-Validation*

To evaluate the robustness of the model's performance, 5-fold cross-validation was performed. The training set was divided into 5 folds, and during each iteration, 4 folds were utilized for training and 1 fold for validation. The average performance across the 5 validation folds is presented in Table 3.

Table 3

Cross-validation results for ANN model

Metric	5-Fold Average
Accuracy	81.2%
Precision	79.4%
Recall	83.1%
F1-Score	81.2%

4.3 *Comparative Analysis*

While the ANN model demonstrates strong predictive performance on the campus placement dataset, it is informative to compare it against alternate modelling techniques as a benchmark. Three alternate models Logistic Regression (LR), Random Forest (RF) and Gradient Boosting Machine (GBM) are trained on the dataset for comparison: The comparative evaluation was conducted using repeated 5-fold cross-validation to minimize variability. The average test metrics across the 5 folds for each model are presented in Table 4.

Table 4

Benchmark model comparison using cross-validation

Model	Accuracy	Precision	Recall	F1-Score	AUC
ANN	81.7%	80.1%	83.5%	81.8%	0.88
LR	79.2%	77.3%	80.4%	78.8%	0.82
RF	80.1%	78.6%	81.7%	80.1%	0.85
GBM	80.8%	79.2%	82.4%	80.8%	0.86

The ANN model achieves the best performance across all evaluation metrics compared to the other benchmark models. This indicates that the ANN's ability to learn complex nonlinear relationships results in higher predictive accuracy for this problem. Logistic regression performance is the lowest, likely because of its linearity assumptions. Random forest and gradient boosting machine achieve competitive performance, benefiting from their ensemble modelling approaches.

5. Conclusion & Future Work

This study introduced an AI-powered approach using ANN to enhance skill-based prediction of placement outcomes. Results presented in Table 4 show that the ANN model outperformed other machine learning models, achieving the highest accuracy (81.7%) in predicting employability. ANN model outperformed other machine learning models, achieving the highest accuracy (81.7%) in predicting employability. Furthermore, feature attribution analysis revealed that technical skills had the most significant impact on placement decisions, followed by aptitude, communication, and personality traits. The effectiveness of the ANN model was validated through a range of evaluation metrics, including F1-score, recall, precision, and AUC. The model demonstrated strong predictive abilities, highlighting the potential of AI in recruitment to enhance data-driven, transparent, and fair decision-making processes. Additionally, comparisons with traditional models such as Logistic Regression, Gradient Boosting Machine, and Random Forest, emphasized the benefits of deep learning techniques for complex employment predictions. Future research can be concentrated more on the following points: Utilizing a broader range of datasets from various institutions and sectors to enhance the model's resilience and minimize potential biases. Utilizing supplementary data sources, including resume text analysis, interview performance metrics, and social media engagement, to enhance predictive accuracy.

References

[1] A. Bai and S. Hira, "An intelligent hybrid deep belief network model for predicting students employability," *Soft Comput.*, vol. 25, no. 14, pp. 9241–9254 (2021), doi: 10.1007/s00500-021-05850-x.

[2] E. Alyahyan and D. Düştegör, "Predicting academic success in higher education: literature review and best practices," *Int. J. Educ. Technol. High. Educ.*, vol. 17, no. 1 (2020), doi: 10.1186/s41239-020-0177-7.

[3] P. Kapil, "Bridging the Industry- Academia skill gap A conceptual investigation with special emphasis on the management education in India," *IOSR J. Bus. Manag.*, vol. 16, no. 3, pp. 08–13 (2014), doi: 10.9790/487x-16330813.

[4] V. Krishnaiah and Y. H. Kadegowda, "Undergraduate engineering students employment prediction using hybrid approach in machine learning," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 3, pp. 2783–2791 (2022), doi: 10.11591/ijece.v12i3.pp2783-2791.

[5] N. Mezhoudi, R. Alghamdi, R. Aljunaid, G. Krichna, and D. Düştegör, "Employability prediction: a survey of current approaches, research challenges and applications," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 3, pp. 1489–1505 (2023), doi: 10.1007/s12652-021-03276-9.

[6] T. Gupta, A. Singh, A. Kaushik, and B. T. Student, "Placement Predict: A Review of Engineering Graduate Placement Statistics in India," *Int. J. Eng. Sci. Comput.* (2018), [Online]. Available: http://ijesc.org/

[7] V. K. Gokuladas, "Predictors of employability of engineering graduates in campus recruitment drives of Indian software services companies," *Int. J. Sel. Assess.*, vol. 19, no. 3, pp. 313–319 (2011), doi: 10.1111/j.1468-2389.2011.00560.x.

[8] K. Vinutha and H. K. Yogisha, "Prediction of employability of engineering graduates using machine learning techniques," *Proc. 2021 8th Int. Conf. Comput. Sustain. Glob. Dev. INDIACom 2021*, pp. 742–745 (2021), doi: 10.1109/INDIACom51348.2021.00132.

[9] S. Thavasi and T. Revathi, "A personalized machine learning–based system to evaluate students' skillset and analyze the gap between academia and industry for engineering students," *Kybernetes* (2024), doi: 10.1108/K-01-2024-0053.

[10] J. Koen, U. C. Klehe, and A. E. M. Van Vianen, "Employability among the long-term unemployed: A futile quest or worth the

effort?," *J. Vocat. Behav.*, vol. 82, no. 1, pp. 37–48 (2013), doi: 10.1016/j.jvb.2012.11.001.

[11] F. F. Patacsil and C. L. S. Tablatin, "Exploring the importance of soft and hard skills as perceived by it internship students and industry: A gap analysis," *J. Technol. Sci. Educ.*, vol. 7, no. 3, pp. 347–368 (2017), doi: 10.3926/jotse.271.

[12] Y. Wei, X. Rao, Y. Fu, L. Song, H. Chen, and J. Li, "Machine learning prediction model based on enhanced bat algorithm and support vector machine for slow employment prediction," *PLoS One*, vol. 18, no. 11 (2023), doi: 10.1371/journal.pone.0294114.

[13] J. Tu, A. Lin, H. Chen, Y. Li, and C. Li, "Predict the Entrepreneurial Intention of Fresh Graduate Students Based on an Adaptive Support Vector Machine Framework," *Math. Probl. Eng.*, vol. 2019 (2019), doi: 10.1155/2019/2039872.

[14] A. Qostal, A. Moumen, and Y. Lakhrissi, "CVs Classification Using Neural Network Approaches Combined with BERT and Gensim: CVs of Moroccan Engineering Students," *Data*, vol. 9, no. 6 (2024), doi: 10.3390/data9060074.

[15] Y. Zayed, Y. Salman, and A. Hasasneh, "A Recommendation System for Selecting the Appropriate Undergraduate Program at Higher Education Institutions Using Graduate Student Data," *Appl. Sci.*, vol. 12, no. 24 (2022), doi: 10.3390/app122412525.

[16] Y. Singh, M. Saini, and Savita, "Impact and Performance Analysis of Various Activation Functions for Classification Problems," *Proc. IEEE InC4 2023 - 2023 IEEE Int. Conf. Contemp. Comput. Commun.* (2023), doi: 10.1109/InC457730.2023.10263129.

[17] A. Chopra and M. Lal Saini, "Comparison Study of Different Neural Network Models for Assessing Employability Skills of IT Graduates," *Int. Conf. Sustain. Commun. Networks Appl. ICSCNA 2023 - Proc.*, pp. 189–194 (2023), doi: 10.1109/ICSCNA58489.2023.10368605.

[18] M. L. Saini, R. S. Telikicharla, Mahadev, and D. C. Sati, "Handwritten English Script Recognition System Using CNN and LSTM," *Proc. InC4 2024 - 2024 IEEE Int. Conf. Contemp. Comput. Commun.* (2024), doi: 10.1109/InC460750.2024.10649099.

[19] A. Chopra and M. L. Saini, "Neural Network Based Employability Index for IT Graduates," *Commun. Comput. Inf. Sci.*, vol. 2268 CCIS, pp. 118–127 (2025), doi: 10.1007/978-3-031-75167-7_9.

[20] A. Chopra, "Employability Prediction Dataset," GitHub (2024). Available: https://github.com/ankit737257/employability-dataset

[21] A. Sharma and R. Gupta, "Machine learning techniques for student performance prediction," *Journal of Statistics and Management Systems*, vol. 24, no. 5, pp. 1123–1135 (2021).

[22] P. Verma and S. K. Singh, "Data-driven approaches in educational analytics," *Journal of Information and Optimization Sciences*, vol. 43, no. 2, pp. 567–580 (2022).

[23] R. Mehta and M. Singh, "Artificial intelligence in higher education systems," *Journal of Statistics and Management Systems*, vol. 25, no. 3, pp. 789–802 (2022).

[24] S. Agarwal and M. Jain, "Predictive modelling using neural networks," *Journal of Information and Optimization Sciences*, vol. 42, no. 6, pp. 1345–1358 (2021).

[25] K. Singh and A. K. Singh, "Optimization techniques in machine learning applications," *Journal of Statistics and Management Systems*, vol. 23, no. 4, pp. 945–960 (2020).

Received April, 2025