

# Explainable Cognitive Graph Intelligence Framework for Multi-Task Graduate Employability and Salary Prediction

Hameeda Khatoon, Chandra Prakash Vudatha

Department of Computer Science and Engineering,  
KL Deemed to be University, Vaddeswaram, Guntur District, India, 522302

**Abstract**—Accurate prediction of graduate employability and expected salary is essential for enabling timely, data-driven career interventions in academic institutions. However, most existing methods rely on standalone machine learning or deep learning models that fail to jointly capture cognitive traits, temporal academic progression, and peer-level relational structures within a unified and interpretable framework. This study introduces ExCogGNet, a hybrid multi-task learning architecture that integrates Multi-Layer Perceptron (MLP), Bidirectional Long Short-Term Memory (BiLSTM), Transformer, and Graph Neural Network (GNN) components, optimized using Bolas Spider Optimization (BSO). The model simultaneously predicts student placement outcomes and salary levels while incorporating cognitive trait awareness. Static psychometric attributes are encoded via MLP, sequential academic trajectories are modeled using BiLSTM, contextual feature interactions are captured through Transformer-based self-attention, and relational dependencies among students are learned using Graph Attention Networks (GAT). The BSO algorithm optimizes hyperparameters, attention weights, and fusion coefficients under a unified multi-task objective. A key contribution is a counterfactual SHAP-based explainability module that converts feature attributions into actionable, personalized recommendations for improving employability and skill readiness, enabling prescriptive educational decision support. Experimental results on the Campus Recruitment dataset show 96% accuracy and 0.97 AUC for placement prediction, along with an RMSE of 24,000 INR and  $R^2$  of 0.91 for salary estimation. The model outperforms baseline methods including SVM, Random Forest, XGBoost, CNN, BiLSTM, Transformer, and standalone GNNs, with statistical significance confirmed via McNemar's test ( $p = 0.003$ ), demonstrating strong predictive and interpretability performance.

**Keywords**—Cognitive properties; prediction of placement; salary regression; graph neural network; bolas spider optimization; explainable AI; MLP-BiLSTM-Transformer

## I. INTRODUCTION

In education, analytics and data mining of placement forecasting are crucial to educational data mining as they enable institutions to determine how prepared a student is to take a job and how they are likely to perform in the job market [1]. Early intervention, individualized training, and professional career counseling can be made with the help of placement status and salary level that are predicted correctly. Nonetheless, academic performance is not the only factor that determines employability; cognitive characteristics like problem-solving

capability, adaptability, motivation, and learning agility play a great role in determining professional success [2].

Traditional machine learning algorithms, such as support vector machines and random forests, are based on manually designed features and the assumption that features are independent, so they cannot easily model complex nonlinear interactions between academic, cognitive, and contextual variables [3]. Although deep learning models enhance representation learning, independent architectures may not consider relational dependencies present in educational settings; they tend to be hyperparameter sensitive and not interpretable, which is not suitable in the case of making high-stakes educational decisions [4]. Relational modeling in graphs gives a formidable tool to model students, skills, and cognitive characteristics as interrelated components, allowing learning of both the individual characteristics and the structure [5]. However, the level of success of graph neural frameworks is highly dependent on the best parameter selection and fusion techniques [6]. Bolas Spider Optimization (also known as metaheuristic optimization) provides a good global search functionality because it balances exploration and exploitation and avoids local optima [7]. In comparison to Particle Swarm Optimization and Genetic Algorithms, which tend to converge quickly in high-dimensional hyperparameter space, and Bayesian Optimization, which uses smooth objective surfaces, BSO uses a two-stage “lasso and reel” exploration scheme that facilitates effective optimization in heterogeneous high-dimensional search spaces that come from hybrid neural systems for the heterogeneous search space of a multi-branch hybrid neural network.

Additionally, educational analytics require transparency, which requires the inclusion of explainable artificial intelligence methods to increase trust and accountability [8]. Inspired by these issues, the following study provides a hybrid MLP-BiLSTM-Transformer graph neural network using Bolas Spider Optimization and explainability to predict the cognition trait-conscious placement and salary [9]. The suggested solution combines the modeling of the static cognitive traits, the successive academic trends, the situational attention, and the relational relationships to provide the correct and explainable forecasts [10]. To overcome these shortcomings, this study introduces ExCogGNet, a single cognitive graph intelligence system to predict graduate employability and salary with explainability incorporated. The proposed model combines heterogeneous learning paradigms to capture different

components of student profiles: an MLP captures the existing method of psychometric and cognitive variables, a BiLSTM captures the temporal academic development, a Transformer captures the interaction of features in the context through attention, and a Graph Attention Network captures the relational relationships between students and skills. These components are optimized together with Bolas Spider Optimization that effectively optimizes hyperparameters and fusion weights over a high-dimensional search space. In addition to predictive performance, ExCogGNet includes a counterfactual SHAP-based explainability component that transforms insights into model-generated recommendations tailored to individuals and actionable to improve skills and career preparedness. This allows the framework to go beyond traditional prediction to transparent, intervention-driven educational analytics, which make it appropriate to be used in the real world within academic decision-making systems.

#### A. Research Motivation

Existing placement prediction systems consider student employability as a single classification problem without paying attention to the dynamic nature of academic performance, relational peer influence, and cognitive aspects of student readiness. Institutions and career advisers have no existing tools that predict outcomes as well as provide explanations about what particular student attributes bring about the outcomes and recommend specific improvement pathways. The increasing complexity of campus recruitment environments, in which the placement status, on the one hand, and the predicted salary, on the other hand, have to be estimated simultaneously to make interventions analytically meaningful, requires an integrated, interpretable, and cognitively aware predictive framework.

#### B. Research Significance

This framework contributes to educational analytics by providing a multi-task predictive model that simultaneously considers addressing placement classification and salary regression using a single optimized architecture. The combination of counterfactual SHAP-based intervention advice converts passive forecasts into actionable career guidance, which has a direct positive impact on students, academic advisors, and institutional placement officers. Bolas Spider Optimization is used to ensure the hybrid architecture is globally optimized and not manually set up to enhance reproducibility and deployment readiness in diverse academic datasets.

#### C. Key Contributions

- Development of ExCogGNet, a novel hybrid architecture that jointly predicts student employability and salary by integrating cognitive traits, academic progression, and relational learning within a single multi-task framework.
- Fluent combination of MLP, BiLSTM, Transformer, and Graph Attention Network to learn static psychometric features, temporal academic trends, contextual feature interactions, and peer-level relational features.
- Hyperparameter, attention mechanism, and fusion weight optimization with Bolas Spider Optimization (BSO) Hyperparameter and attention mechanism optimization Learning: Hyperparameter Tuning with Bolas Spider Optimization (BSO) Hyperparameter

tuning to effectively explore complex, high-dimensional model configurations, hyperparameter, and attention mechanism optimization is performed with Bolas Spider Optimization (BSO).

- Design of a counterfactual SHAP-based explainability and intervention module that converts feature attribution insights into student-specific, actionable recommendations capable of supporting prescriptive educational decision-making and career counseling.
- Showing substantial gains compared to traditional and deep learning baselines on classification and regression tasks, which are statistically tested as robust, and making sure that results are reliable for real-world academic decision support systems.

The remaining study is structured in the following way. Section I is a review of related literature on the topics of placement and salary prediction, cognitive traits modeling, graph analytics, metaheuristic optimization, and explainable AI. Section II presents the problem, which includes the dataset, feature space, cognitive traits, and multi-task objectives. Section III introduces the suggested hybrid MLP-BiLSTM-Transformer GNN study based on explainable AI. Section IV describes the description of experiments, results, and analysis. Section V is the conclusion of the study, and future work is mentioned.

## II. LITERATURE REVIEW

GNNs have demonstrated spectacular potential in integrating relational reasoning with neural-symbolic computing, allowing the model to work well with structured data of complexity. These networks enhance relational dependency learning and improve educational performance prediction through graph-based student representation modeling. Nonetheless, existing techniques tend to be ineffective in terms of scaling and extrapolation to large and heterogeneous graphs. It can be accurate depending on the dataset, with moderate performance on large-scale knowledge graphs [11]. Transformers have been generalized to graph structures, enabling node dependency attention to be well represented. This generalization is useful for prediction tasks on a node level and on a graph level. However, the method has difficulties with computational efficiency and space consumption with very large graphs. Experimental tests have registered improvement in node classification but not significant improvement in graph regression activities [12].

Student performance prediction with Hybrid LSTM-Transformer models, a combination of temporal sequences and attention, has been used. Explainable AI methods like SHAP give information on how models make decisions. Among the limitations, there are overfitting on small datasets and the inability to interpret complicated interactions. Graph Neural Networks with transformers have achieved up to 92 % accuracy in academic results prediction [13]. Transformer-based Graph Neural Networks have been applied to automated scoring of academic essay tasks, jointly sequence modeling and graph structure learning. The hybrid strategy enhances the score on multi-dimensional scoring. Scalability problems, however, occur when very long essays and extensive educational datasets are used. According to report accuracy, it has been significantly

improved at an average of 90% in comparison to baseline models [14].

Further validation on the performance improvement of the combination of self-attention mechanisms and bidirectional long short-term memory networks has been recently shown. One study on detecting IoT attacks using the TPE-based self-attention biLSTM architecture found that the application of an attention mechanism with bidirectional modeling contributes to learning context dependencies and optimization stability in diverse data environments. The research further emphasized the significance of employing appropriate optimization methods in hybrid models [15].

Job matching has also been optimized using Graph Neural Networks over job and candidate network representations. This methodology represents relational dependencies in a superior manner compared to traditional methods. Shortcomings such as sensitivity to noisy/incomplete job datasets are a limitation. According to the experimental findings, the corresponding accuracy of the test scenario is 88% [16]. Explainable salary prediction has been based on neural prototyping using disentangled composition effects. The structural dependencies in work characteristics and the patterns of salary distribution are represented in models. Weaknesses are problems with the complexity of modeling multi-factor interactions and extrapolating to cross-domain datasets. Salary prediction tasks produce accuracy at a level of about 91 % [17] with synthetic job listings having been used to simulate the process of salary prediction and job grouping. These are experimental datasets that are controlled with regard to testing the model performance. Disadvantages are less realism and possible bias because of the use of synthetic data. This performance has an accuracy of about 87 % on simulated environments [18].

Regression and ensemble methods are machine learning techniques that have been used to make salary predictions. Whereas simple models provide interpretability, they tend to perform poorly with complicated data sets with numerous categorical variables. Weaknesses include the inability to generalize to invisible job roles and sectors. Precision is around 85 % on benchmark data sets [19]. Comparative experiments of salary prediction models show that deep learning approaches do better than traditional ML algorithms with large data sets. There is a high cost of computation and sensitivity to hyperparameters. Certain models have problems with feature sparsity and missing data. Comparison of accuracy indicates that neural networks can perform salary prediction with up to 93 % accuracy [20]. Principal Component Analysis with Deep Neural Networks has been suggested in modeling the relationships between salaries and dimensionality reduction. The limitations are that there may be a loss of information in the course of the PCA transformation, and this is sensitive to the feature scaling. The model is approximately accurate at 90 % when making predictions of salary ranges [21].

Cat Swarm Optimization is a meta-heuristic algorithm that offers a good solution to complex optimization problems, including the parameter tuning of ML models. Grid, random search, Bayesian optimization, PSO, and Genetic Algorithm are some examples of the classical hyperparameter optimization techniques that have found application in the optimization of

machine learning models. The grid and random search techniques have the advantage of being simple techniques for searching for parameters; however, they can be computationally intensive when dealing with high-dimensional hybrid structures. In addition, Bayesian optimization is more efficient in sample search optimization but could experience difficulties when dealing with very heterogeneous objective spaces with many components. It has weaknesses such as slowness in converting to highly non-linear problems and sensitivity to the parameters of the algorithm. Claims of efficiency increase to 85 % in benchmark optimization exercises [22]. GNN has been used in combinatorial optimization problems, improving thinking on structured data. Disadvantages include scalability as well as the excessive computing cost of large combinatorial spaces. The graph size is very important for accuracy; moderate scores are around 82% [23] with the hybrid Transformer-BiLSTM frameworks being applied in energy demand forecasting, time sequences and attention mechanisms are used to enhance prediction. Its weaknesses are inability to model the long-term dependencies of highly dynamic systems. The precision in test conditions goes as high as 94% [24].

Social media sentiment analysis of both transformer and attention-based recurrent networks with explainability and sequence modeling has been applied. It has such limitations as dependency on large labeled datasets and sensitivity to noisy text data. GraphXAIN combines natural language explanations with Graph Neural Networks to give graph-based tasks interpretable predictions with an accuracy of about 91% for sentiment classification [25]. Weaknesses such as reliance on the quality of generated narratives and computation of large graphs are also present. The reported accuracy of graph classification tasks is 88% [26].

#### A. Research Gap

Although it has achieved considerable progress in the field of graph-based learning, transformers, and hybrid deep learning (DL), there are still a number of essential gaps in the area of student employability and salary forecasting. Current Graph Neural Networks (GNNs) can model relational dependencies well, but with large and heterogeneous education data, fail to be scalable and have poor generalization. On the same note, Transformer-based graph models enhance attention-dependent dependency learning, but have high computational demands and small benefits in regression-based tasks like salary forecasting. The hybrid architecture that combines LSTM and Transformer models has demonstrated the ability to predict temporal academic patterns, but they are susceptible to overfitting, interpretability is not robust, and they do not consider relational and cognitive aspects in entirety. Regarding employability and salary analytics, the existing methods usually consider the task of prediction in isolation, overlooking the fact that there is an inherent multi-task correlation between the results of the placement and the salary level. Furthermore, post hoc explainable AI methods like SHAP often lack built-in integration into a cohesive framework that can be used to take action and personalize interventions. Current optimization techniques, such as the classical metaheuristics, are inefficient in high-dimensional search spaces or are not flexible to hybrid architectures with multiple interacting components. Hence, the gap in the research is evident in creating a scalable, multi-task,

and interpretable model, which can be simultaneously able to combine cognitive characteristics, time course progression, and relationship learning, whilst using effective global optimization and offering intervention-based explainability to real-world educational choices.

### III. RESEARCH METHODOLOGY

The dataset employed in the study is Campus Recruitment, which includes academic, demographic, and cognitive trait characteristics, as well as placement status and salary offer. Preprocessing of the data includes missing value imputation, data normalization, and categorical encoding to maintain uniformity. The ExCogGNet framework suggested is based on MLP in terms of static cognitive characteristics, BiLSTM in terms of sequential academic data, Transformer in terms of contextual attention, and GAT in terms of relational modeling. Bolas Spider Optimization (BSO) is used to optimize the hyperparameters of the network, attention coefficients, and fusion weights. Multi-task learning is a collaborative action that forecasts a placement (classification) and salary (regression). Lastly, XAI methods such as SHAP offer both global and local interpretability and can be used to provide actionable information on student employability factors (Fig. 1).

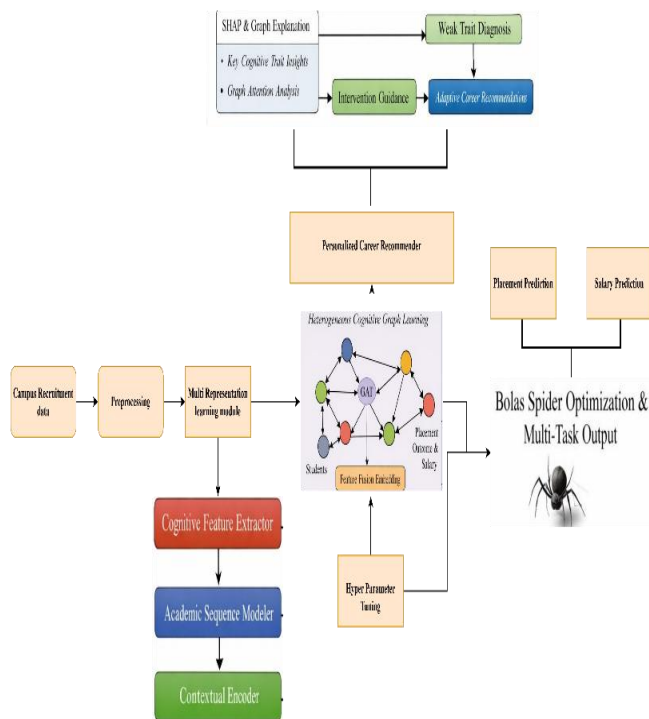


Fig. 1. ExCogGNet architecture with BSO and XAI.

#### A. Dataset Description

In the case of this research, we used the Campus Recruitment dataset found on Kaggle [27] that gives exhaustive data on the academic, demographic, and cognitive characteristics of students, as well as placement position and salary performance. The database consists of around 215 records, and the most important data points are 10 th and 12 th grades percentages, undergraduate performance, specialization, gender, previous work experience, and psychometric scores. The salary values in

the dataset range approximately between 200,000 INR and 940,000 INR, with the majority of placed students receiving salaries within the 200,000–400,000 INR interval. Despite the absence of semester-wise temporal details within the dataset, the academic factors correspond to progressive milestones that signify the process of gradual academic development from high school education to university education. The milestones of progressive academic development form a structured sequence that is suitable for modelling academic evolution in relation to employment outcomes. These various features render the data ideal in the joint modeling of the placement classification and salary prediction problems in one analysis framework.

Despite the small number of records in the Campus Recruitment dataset of 215, there are several methodologies that have been adopted to prevent overfitting and enhance generalization. Firstly, the suggested methodology includes the use of Dropout regularization, weight decay, early stopping, and fivefold cross-validation to ensure stability when training the model on different datasets. Secondly, the use of a multi-branch model ensures the learning of complementary models instead of adding complexity by having parameters for the different branches. For instance, the MLP will be used to learn static cognitive skills, the BiLSTM learns the academic trajectory of the individual students, the Transformer learns interactions between contextual features, and finally, the GAT learns the dependencies between students. In addition to this, the use of Bolas Spider Optimization ensures that the model selects the optimal hyperparameters to avoid instability when training the model. The Campus Recruitment data set is used as a benchmark data set in the context of education for assessing the performance of the hybrid employability prediction models, but the size of this data set is limited in nature and might not adequately reflect the diversity found in larger university populations.

#### B. Preprocessing

Preprocessing of the data set included various important steps that guaranteed quality and consistency of data. The missing values of numerical features were treated with mean imputation, and the available values of categorical features were imputed with the mode of the respective column. Continuous variables, including grades and salary, were min max normalized to fall within the range of [0,1] was expressed in Eq. (1).

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

One-hot encoding of nominal categorical features such as specialization and gender was used to facilitate the use of these features in models of neural networks. Also, psychometric scores and personality indicators were summarized into a cognitive trait vector  $C \in R^d$  per student, which offers an economic representation of cognitive attributes of each student towards the predictive framework.

#### C. Framework Architecture

The proposed ExCogGNet model is a hybrid deep learning model that is a combination of MLP, BiLSTM, Transformer, and Graph Neural Networks (GNNs) with Bolas Spider Optimization (BSO) controlling the choice of hyperparameters and weights (Fig. 2). The design is cognitive-trait-aware

designed used to predict salaries, combining both the presence of static, sequential, relational representations, and allowing the explanation of the design by means of the XAI techniques. The architecture consists of several branches, which are trained to complement each other and optimize together on classification (placement) and regression (salary).

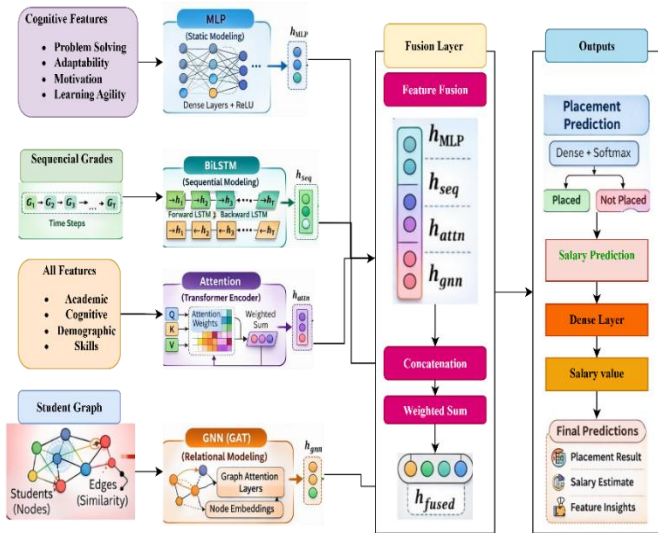


Fig. 2. Architecture of the proposed method.

1) *MLP static cognitive modeling*: The MLP branch takes the psychometric scores and personality indicators as they are extracted as vectors  $C$  to the MLP branch. The MLP is learned to find higher-order correlations between these features in the form of multi-layered hidden interactions. The hidden representation is calculated in Eq. (2).

$$h_{MLP} = f(W_1 C + b_1) \quad (2)$$

$f(\cdot)$  represents the Relu Activation function,  $W_1$  represents the weight matrix, and  $b_1$  represents the bias vector. The network uses numerous overlapping layers to simulate intricate patterns between cognitive characteristics that determine employability.

2) *Sequential academic modeling using BiLSTM*: To incorporate progressive academic development in the model for various levels of education, the model takes indicators like the percentage of secondary, higher secondary, and undergraduate education as the sequential development stage of education and is not based on a temporal sequence since there are no semester-based sequential data present in the dataset. However, the indicators show patterns that relate to academic growth and impact job prospects and salary outcomes in the future. Therefore, the BiLSTM model will be applied to capture the progression of academic development stages as sequenced in Eq. (3).

$$X_{seq} = [x_1, x_2, \dots, x_T] \quad (3)$$

Such a sequence is processed by a Bidirectional LSTM (BiLSTM) both forward and backward, and therefore the model

can make use of the context of the previous and upcoming semesters in Eq. (4).

$$h_{\rightarrow}, h_{\leftarrow} = LSTM(x_t, h_{\rightarrow}, h_{\leftarrow}) \quad (4)$$

Final sequential embedding the last forward and backward hidden states are concatenated to get the final sequential embedding in Eq. (5).

$$h_{seq} = [h_{\rightarrow}; h_{\leftarrow}] \quad (5)$$

The encoding of these trends in a temporal change in performance affects the placement likelihood and projected salary.

The rationale behind this proposed sequential modeling paradigm stems from contemporary approaches to incorporating self-attention mechanisms into the bidirectional LSTM architecture, which exhibited enhanced dependency modeling capabilities and training optimization stability through attention-based sequence modeling methodologies [15].

3) *Contextual attention transformer encoder*: To obtain inter-feature contextual interactions, the framework uses a Transformer encoder and multi-head self-attention. Embeddings of inputs are transformed into query ( $Q$ ), key ( $K$ ) and value ( $V$ ) matrices, and the attention mechanism is calculated in Eq. (6).

$$Attention(Q, K, V) = softmax\left(\frac{QK^t}{\sqrt{d_k}}\right)V \quad (6)$$

Multi-head attention enables the model to attend to multiple elements of the input at the same time so that it generates a contextual embedding  $h$  (trans) that augments the expression of intricate interactions among features.

4) *Graph neural network for relational modeling*: The students are modelled as nodes of a graph  $G = (V, E)$  where the similarity is coded as an edge depending on the academic and cognitive characteristics. Specifically, two student nodes  $i$  and  $j$  are connected by an edge when cosine similarity exceeds a threshold of 0.85; the cosine similarity between the concatenated vectors of the normalized degree percentage, cognitive trait score, and technical skills score ensures that only meaningfully similar students are connected, yet the graph sparsity is maintained. Graph Attention Network (GAT) is an updated embedding of nodes based on the attentional message passing in Eq. (7).

$$h'_i = \sigma\left(\sum_{j \in N(i)} \alpha_{ij} W h_j\right) \quad (7)$$

The attention coefficients  $\alpha_{ij}$  are determined using an adaptive calculation that is as follows in Eq. (8).

$$\alpha_{ij} = \frac{\exp(LeakyReLU(a^T [W h_i \| W h_j]))}{\sum_{k \in N(i)} \exp(LeakyReLU(a^T [W h_i \| W h_k]))} \quad (8)$$

This process enables the network to give each student a different weight of peer influence, which reflects relational patterns that have an influence on placement and wages.

5) *Bolas Spider Optimization (BSO) module*: The BSO module is optimized on essential hyperparameters, such as

learning rate and hidden units, attention coefficients and fusion weights  $(\lambda_{MLP}, \lambda_{BiLSTM}, \lambda_{transformer}, \lambda_{regression}, \lambda_{GNN})$ . The fitness function is a combination of regression and classification goals in Eq. (9).

$$L_{total} = \lambda_1 L_{classification} + \lambda_2 L_{regression} \quad (9)$$

$L_{classification}$  is the cross-entropy loss when using to predict placement, and  $L_{regression}$  mean squared error when predicting salary. BSO will perform a search through the search space, and the candidate solutions will be updated to reduce the total of  $L_{total}$ .

BSO was chosen in this experiment due to its adaptability in terms of achieving the optimal balance between explorative and exploitative searches when searching for optimal solutions within a heterogeneous hybrid architecture with multiple branches and attention mechanisms (Fig. 3). As compared to brute-force search techniques such as grid search and random search, BSO carries out optimization via population-based iterations on high-dimensional continuous parameter spaces. As compared to Bayesian optimization techniques that assume smooth surfaces of the objective function, BSO offers a flexible approach to global searching of the objective space in multi-task optimization scenarios.

Aside from hyperparameter tuning, the dynamic adjustment of the loss balancing coefficients  $\lambda_1$  and  $\lambda_2$  for multi-task is performed in the BSO model in order to reach a desirable balance between the accuracy in the classification of placement and the salary prediction. Candidate solutions are assessed during the optimization process in relation to both objectives, making it possible to have dynamic balancing.

6) *Computational complexity analysis*: The ExCogGNet's computational complexity comes from its joint computation of the MLP, BiLSTM, Transformer, and GAT branches. The computation needed by the MLP branch is  $O(ndh)$ , where  $n$  refers to the number of samples,  $d$  is the dimensionality of features, and  $h$  stands for the number of hidden neurons. The BiLSTM branch brings sequential computation complexity of  $O(T^2h)$ , with  $T$  being the sequence length. The encoder of Transformer has a computation complexity of  $O(T^2d)$  caused by self-attention mechanisms, while GAT necessitates  $O(|E|d)$  computation over graph edges  $E$ . Moreover, BSO optimization adds iterative search complexity of  $O(P \times I)$ , in which  $P$  is the size of the population, and  $I$  is optimization iterations. Even though the hybrid network model causes an increase in computation complexity compared to single networks, it offers much-improved predictive power, relational reasoning, and explanation.

7) *Multi-task learning and explainable AI (XAI) combinations*: The output of the feature embeddings of every branch is combined and sent to specialized multi-task output heads: a Softmax unit to classify the placements and a linear unit to predict the salary. SHAP values are calculated to provide both global feature importance and local prediction-level explanations, and attention weights of the Transformer and GNN parts are visualized to emphasize influential cognitive

attributes and academic factors. This combined design guarantees the model provides not just a high predictive validity, but also clear and implementable information to career advisors and institutional decision-makers.

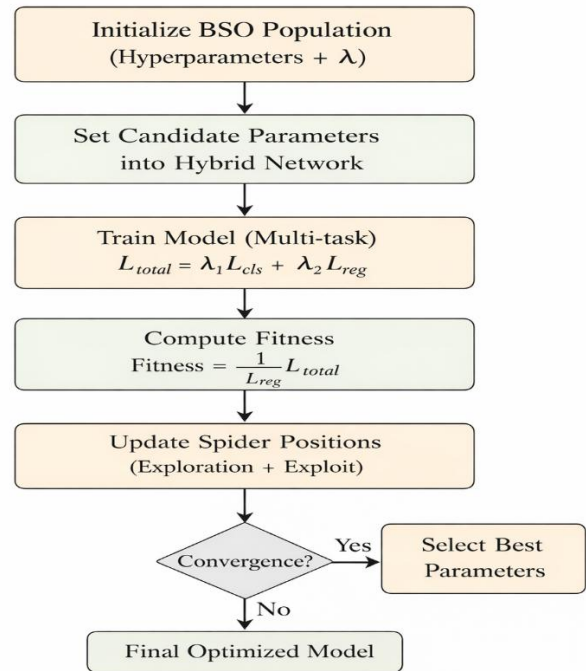


Fig. 3. Bolas Spider Optimization (BSO) workflow for hyperparameter and fusion weight optimization.

8) *Counterfactual personalized intervention module*: In addition to global and local SHAP accounts, the given framework has a counterfactual intervention module, which transforms SHAP feature attributions into action-oriented and student-specific suggestions. To every student forecasted as not placed, or forecasted below some salary level, the module finds the two most negative SHAP contributors and calculates how much the student needs to improve them in order to get his or her placement probability above the 0.5 mark or to get to his or her target percentile of salary. Consider, given that a student has a probability of placement, and the highest negative predictors are technical skills score and communication score, the module would give the particular level of improvement in each feature that is necessary to move the prediction to the placed category, giving career counselors an actual interpretable intervention strategy instead of an uncooked prediction.

The ExCogGNet algorithm proposed in (Algorithm 1) starts with pre-processing of the data such as normalization, encoding, and train-test splitting. Then a graph is created to depict the relationships among students in terms of similarity of features. The model combines the four branches: an MLP that captures cognitive features, a BiLSTM that captures temporal academic progression, a Transformer that captures contextual interactions, and a GNN that captures relationship learning. These embeddings are combined and sent to two output heads to perform classification in placement and salary prediction. The hyperparameters and fusion weights are tuned with Bolas Spider

Optimization and minimize a multi-task loss. Backpropagation is used to train the model in an iterative manner, and classification and regression measures are used to evaluate the model. Lastly, SHAP values and attention weights can be interpreted and make sense, thus allowing transparent and actionable decision-making.

---

**Algorithm 1:** BSO-Optimized Hybrid Network for Placement & Salary Prediction

---

Input:  
Dataset D with features (academic, demographic, cognitive)  
Labels: Placement ( $Y_{cls}$ ), Salary ( $Y_{reg}$ )

Output:  
Trained model M  
Placement predictions  $\hat{Y}_{cls}$   
Salary predictions  $\hat{Y}_{reg}$   
Explainability insights (SHAP values, attention maps)

Begin

1. Data Preparation
  - 1.1 Load dataset D
  - 1.2 Handle missing values and encode categorical variables
  - 1.3 Normalize numerical features
  - 1.4 Split D into training set ( $D_{train}$ ) and testing set ( $D_{test}$ )
2. Graph Construction
  - 2.1 Define nodes as students
  - 2.2 Compute similarity matrix based on feature space
  - 2.3 Construct adjacency matrix A
  - 2.4 Build graph  $G = (V, E)$
3. Model Initialization (ExCogGNet)
  - 3.1 Initialize MLP branch for static cognitive features
  - 3.2 Initialize BiLSTM branch for temporal academic sequences
  - 3.3 Initialize Transformer branch for feature interactions
  - 3.4 Initialize GNN (GAT) branch for relational learning
  - 3.5 Initialize fusion layer
  - 3.6 Initialize output heads:
    - Softmax layer for placement classification
    - Linear layer for salary regression
4. Optimization using Bolas Spider Optimization (BSO)
  - 4.1 Initialize population of candidate solutions (hyperparameters)
  - 4.2 While stopping criterion not met:
    - a. Evaluate fitness of each candidate:
      - Train model with given parameters
      - Compute multi-task loss:  
 $L = L_{classification} + \lambda * L_{regression}$
    - b. Update candidate solutions using BSO:
      - Exploration phase (lasso search)
      - Exploitation phase (reel refinement)
    - c. Select best-performing configuration
  - 4.3 Set optimal parameters in model M
5. Model Training
  - 5.1 For each epoch:
    - a. Forward pass:
      - MLP  $\rightarrow$  static embeddings
      - BiLSTM  $\rightarrow$  temporal embeddings
      - Transformer  $\rightarrow$  contextual embeddings
      - GNN  $\rightarrow$  relational embeddings
    - b. Fuse embeddings
    - c. Generate outputs:
      - $\hat{Y}_{cls} = \text{Softmax}(\text{fused features})$
      - $\hat{Y}_{reg} = \text{Linear}(\text{fused features})$
    - d. Compute multi-task loss L
    - e. Backpropagate and update weights
6. Model Evaluation

---

---

- 6.1 Predict on  $D_{test}$
- 6.2 Compute metrics:
  - Accuracy, AUC (classification)
  - RMSE,  $R^2$  (regression)
7. Explainability Module
  - 7.1 Compute SHAP values:
    - Global feature importance
    - Local explanations for each prediction
  - 7.2 Extract attention weights from:
    - Transformer layers
    - GNN layers
  - 7.3 Generate visualizations and insights
8. Output Results
  - 8.1 Return trained model M
  - 8.2 Return predictions ( $\hat{Y}_{cls}$ ,  $\hat{Y}_{reg}$ )
  - 8.3 Return explainability outputs

End

---

#### IV. RESULT AND DISCUSSION

The Results and Discussion section considers the performance of the proposed framework, ExCogGNet, in terms of predicting student employability and salary, in terms of accuracy and interpretability. The performance of the model is evaluated based on the common classification and regression indicators, and compared with the traditional machine learning and state-of-the-art deep learning baselines. In addition to quantitative performance, the role of integration of cognitive traits, time progression, and relational learning in enhancing predictions is also examined in this section. Moreover, the results of explainability based on SHAP analysis and attention mechanisms are discussed to learn about some major factors affecting them. The results reveal the strength, generalization, and viability of the suggested method in the practical application of academic decisions in real-life situations.

##### A. Experimental Outcome

The general distribution of the Campus Recruitment data that was used in this study is shown in Table I. The sample size is 215 records of students; 148 students (68.8) were successful in placement, and 67 students (31.2) were not successful in placement. With regard to gender, the data sample is mostly male, with 139 students (64.7) representing it and 76 students (35.3) representing females. Concerning work experience, 65 students (30.2) have previous work experience, and 150 students (69.8) do not have previous work experience. The sample also contains two major specializations: 113 students (52.6) are in Commerce and Management, and 102 students (47.4) are in Science and Technology. In general, the table offers a concise summary of the demographic, academic, and placement-based features of the data, which illustrates that it is diverse and can be used to model not only the determinants of the placement outcomes but also the predictors of the salary.

Table II provides the detailed results of the optimization of the hyperparameters of the proposed BSO-based hybrid MLP-BiLSTM-Transformer-GNN model. In the case of the MLP branch, two hidden layers with 128 and 64 units, ReLU activation, and a dropout rate of 0.3 were used to reduce overfitting. The BiLSTM branch has 128 hidden units and a dropout rate of 0.2, and it is merely one layer and essentially captures the time-based dependencies in the sequential academic data. Transformer encoder is set with 4 attention heads, an

embedding dimension of 128, a feed-forward network (FFN) dimension of 256, and a dropout of 0.1, which enables the model to consider the contextual interactions between features. The GAT branch has 4 attention heads with a dimension of 128 as the output and ELU activation to learn relational embeddings over the student graph. All the branches are fused to the final branch by a weighted sum strategy where each branch's weight is 0.25. Training was done using the Adam optimizer, a batch size of 32, a learning rate of 0.0005, 100 epochs, and a weight decay of 0.0001.

TABLE I. DATA DISTRIBUTION TABLE

Category	Sub-Category	Count	Percentage (%)
Total Samples	—	215	100
Placement Status	Placed	148	68.8
	Not Placed	67	31.2
Gender	Male	139	64.7
	Female	76	35.3
Work Experience	Yes	65	30.2
	No	150	69.8
Specialization (Degree)	Commerce & Mgmt	113	52.6
	Science & Tech	102	47.4

TABLE II. OPTIMIZED HYPERPARAMETERS (BSO-BASED HYBRID MODEL)

Module	Key Settings
MLP	Layers: 2 • Units: (128, 64) • Activation: ReLU • Dropout: 0.3
BiLSTM	Layers: 1 • Hidden Units: 128 • Dropout: 0.2
Transformer	Heads: 4 • Embedding Dim: 128 • FFN Dim: 256 • Dropout: 0.1
GAT	Attention Heads: 4 • Output Dim: 128 • Activation: ELU
Fusion	Strategy: Weighted Sum • Weights ( $\lambda$ ): 0.25, 0.25, 0.25, 0.25
Training	Optimizer: Adam • LR: 0.0005 • Batch Size: 32 • Epochs: 100 • Weight Decay: 0.0001

The distribution shows that most of the placements are within the ₹200k-₹400k range, which comprises almost 78.4% of the students placed (Fig. 4). Simultaneously, the existence of higher pay ranges indicates that there is an opportunity to have better pay among the high-achieving students, not only showing consistency in the overall placement results but also signifying the possibility of receiving high-paying jobs.

The Secondary percentage is 78.5%, with a standard deviation of 6.2, showing that the majority of the students were fairly consistent in performance, ranging between 60 and 95. The average of the Higher Secondary percentage is a bit higher at 81.2% and has less variability (5.8), indicating enhanced performance that is more consistent in senior secondary education, with a range of 65 to 98. The degree percentage in the undergraduate level is 75.6% with a wider variation (7.1), with values in the range of 55 to 95, meaning that students are more varied in academic results. Other than academic performance, the Cognitive Trait Score is 68.4 with a standard deviation of 8.5, showing moderate variations of cognitive abilities among

the students. The highest variability is 9.2, with a mean of 72.3 of the Technological Skills Score, which indicates a great variation in technical competence with ranges of 50 to 90. On the whole, the table proves that there is a relative consistency in academic performance among students, but cognitive and technical skills are more diverse, and they are valued when it comes to distinguishing the general competencies and employability of the students.

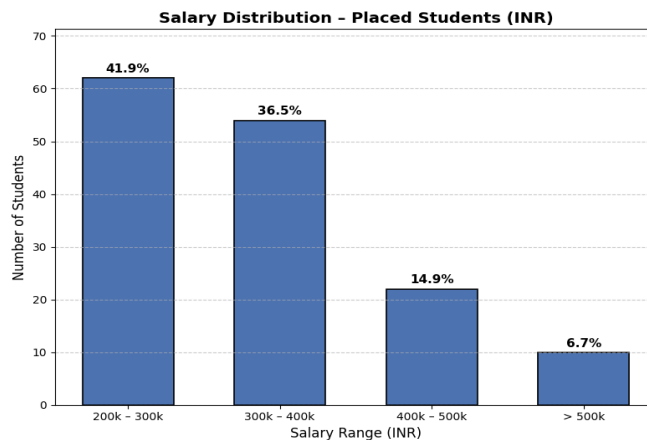


Fig. 4. Salary distribution.

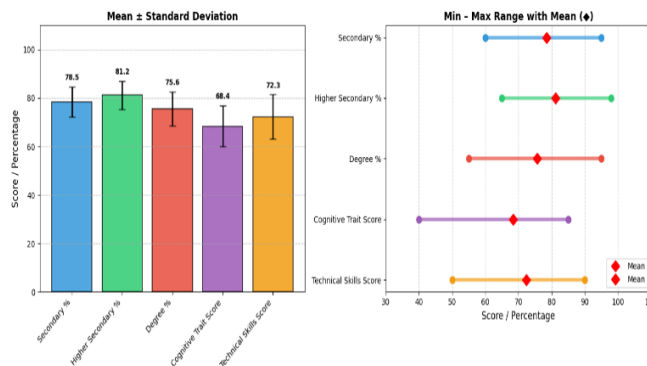


Fig. 5. Academic feature distribution.

The descriptive statistics in Fig. 5 show the mean performance of the students based on key academic, skill, and experience-related variables and the variability as well as the overall range of the scores. The Secondary Percentage (10 th grade) has a mean of 78.5 as well as a standard deviation of 6.2, as the performance of the students is consistent with scores ranging between 60% and 95%. The Higher Secondary %age (12 th) is a little higher, with the average being 81.2 % with a narrower spread (5.8), indicating a stronger and more consistent performance in senior secondary education. Degree Percentage, which is a measure of the performance of the undergraduates, has a mean of 75.6 and a higher standard deviation (7.1), with more varied academic results of 55-95.

It shows the feature contribution analysis with SHAP (SHapley Additive exPlanations), which measures the impact of each feature on the predictions of the model. The most significant features noted in the table, along with the importance score and the direction of the effect, are the Degree Percentage, with an importance score of 0.21. Its positive effect signifies that

the greater the undergraduate scores, the stronger the projected probability of achievement, as indicated in success, i.e., placement or anticipated performance. Such a strong positive influence of the Cognitive Trait Score (0.18) also highlights the importance of analytical and aptitude skills in the determination of results (Fig. 6).

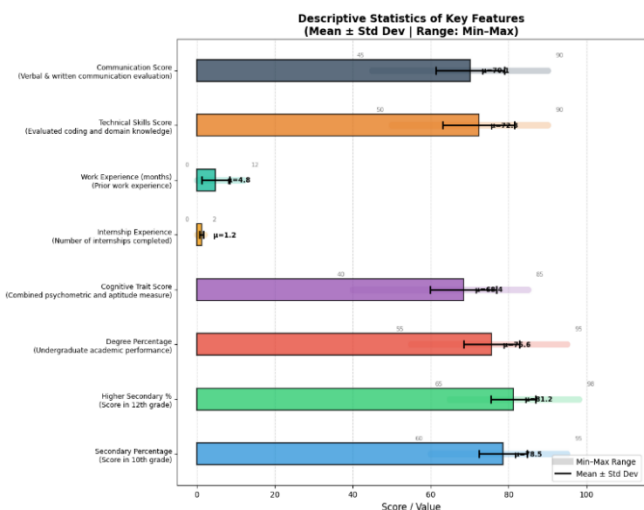


Fig. 6. Descriptive statistics of key features.

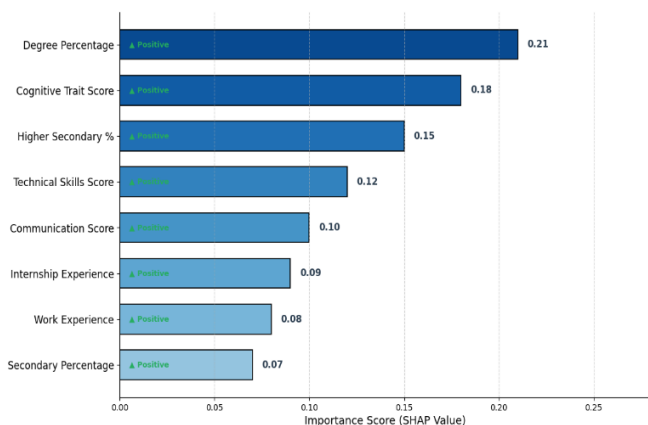


Fig. 7. Feature contribution analysis (SHAP-Based).

Fig. 7 shows the performance of the model based on standard evaluation measures: Precision, Recall, F1-score, and Support. In the Placed class, the model has a Precision of 97, meaning that most of the students that are predicted to be placed are actually placed correctly. There was a recall of 96%, indicating that the model was able to recognize 96% of the total students that were actually placed. The overall performance of this class is high, as the precision and recall balance; the F1-score is 96.5. The Placed class has a support of 106 students.

The confusion matrix helps to determine the effectiveness of the placement classification model in terms of comparison between the results predicted and actually obtained. Out of the 106 students who were actually placed, the model was correct in 102 of them being placed, and 4 students were misclassified as not placed. Of the 59 students who had not been placed, 56 of them were correctly identified as not placed, and 3 of them were not identified as placed. These findings show that the model is

very accurate and the misclassifications in both classes are very minimal. On the whole, the confusion matrix illustrates that the model is very useful in differentiating placed and not placed students, which means that it is very reliable in making predictions of placement results (Fig. 8).

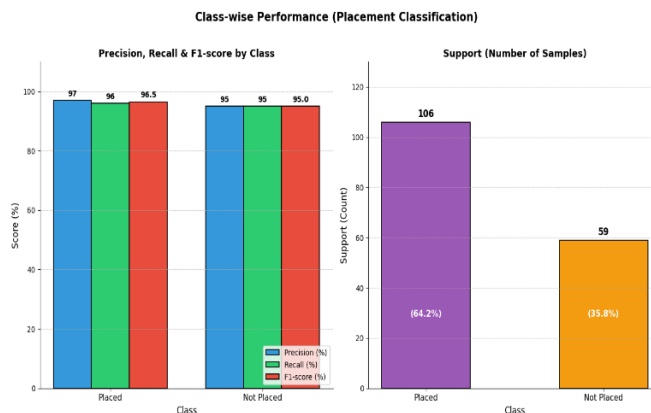


Fig. 8. Class-wise performance.

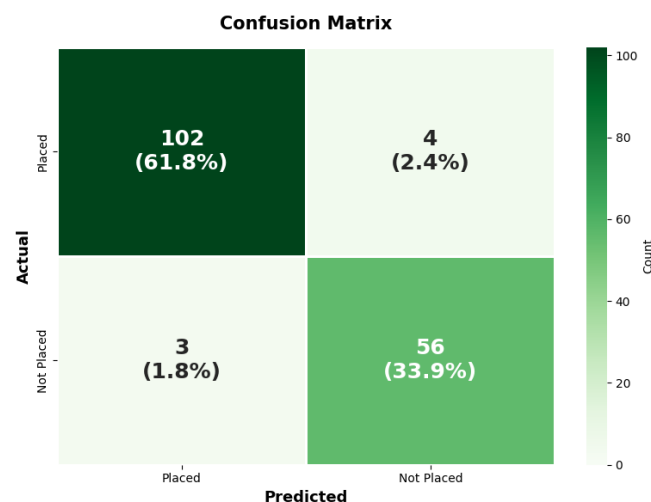


Fig. 9. Confusion matrix.

Fig. 9 gives the performance of different machine learning and deep learning models in terms of placement prediction in various evaluation metrics, such as accuracy, precision, recall, F1-score, and AUC. Conventional models such as SVM had moderate performance with an accuracy of 78% and AUC of 0.81, which reflected decent but poor predictive powers. Random Forest and XGBoost enhanced the performance, both achieving 82% and 84 % of the accuracies, respectively, with correspondingly higher rates of precision, recall, F1-score, and AUC, indicating more aptitude to address the complicated interaction of features.

It shows the salary prediction performance of various regression models in terms of RMSE, MAE, and R<sup>2</sup>. Linear regression has the poorest results with an RMSE of 48k INR, an MAE of 36k INR, and R<sup>2</sup> of 0.62, which suggests that it has very poor predictive power. Random Forest and XGBoost ensemble models bring the predictions to RMSE of 42k and 38k INR, MAE of 31k and 28k INR, and R<sup>2</sup> of 0.71 and 0.75, respectively,

showing more non-linear relationships in the data. Deep learning architectures such as BiLSTM and Transformer are also used to improve the performance, with RMSE of 35k and 33k INR, MAE of 25k and 23k INR, and  $R^2$  of 0.78 and 0.80 indicating capability to capture sequential and complex feature interactions. The GNN, which does not use BSO, exhibits further improvement with an RMSE of 31k INR, MAE of 21k INR, and  $R^2$  of 0.82, which indicates the advantage of capturing relational information in the model. Compared to all other models, the proposed hybrid model of MLP, BiLSTM, Transformer, and GNN with BSO optimization has yielded data with an RMSE of 24k INR, MAE of 18k INR, and  $R^2$  of 0.91, which includes highly predictive and reliable data in terms of accuracy and performance of the predicates in the salaries. Given that the salary range in the dataset falls between roughly 200k INR and 940k INR, an RMSE of 24k INR can be regarded as quite low, and hence a sign of the reliability of the regression model (Fig. 10).

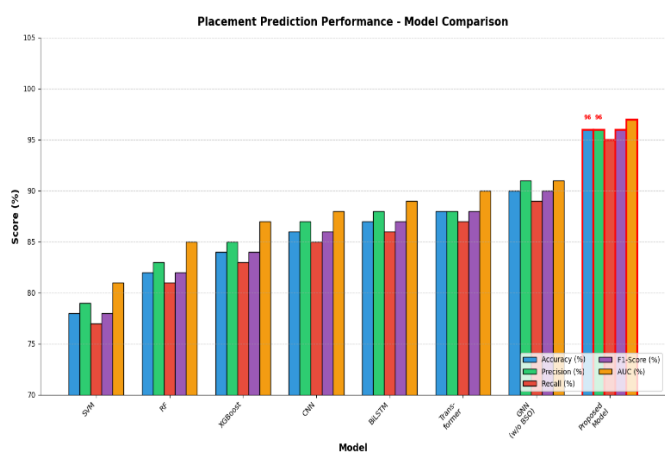


Fig. 10. Placement prediction performance.

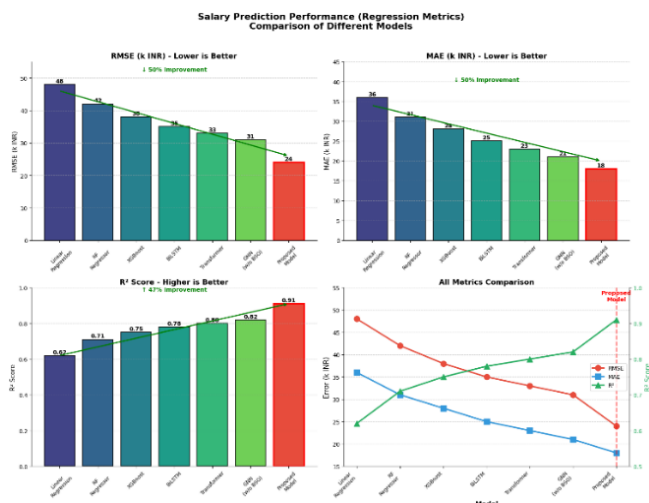


Fig. 11. Salary prediction performance.

### B. Performance Analysis

Fig. 11 illustrates how the model accuracy of training and testing accuracy improved with 100 epochs. At the first 10 epochs, the training accuracy is 88.5, and the testing accuracy is 86.2, which means that the model is in the process of learning

but still can be improved. The higher the number of epochs, the higher the training and testing accuracy, indicating that there is an efficient learning and generalization. At 50 epochs, the training and testing accuracy are 95.5 and 94.0%, respectively, which is good model performance with very little overfitting. After 60 epochs, the accuracy gains start to level off, and the training accuracy reaches 97.5 per cent, whereas the testing accuracy reaches around 96.0 per cent. In general, the table proves that the model is convergent, as its accuracy is high and stable on the training and testing data, which proves strong learning and strong generalization on the unknown data (Fig. 12).

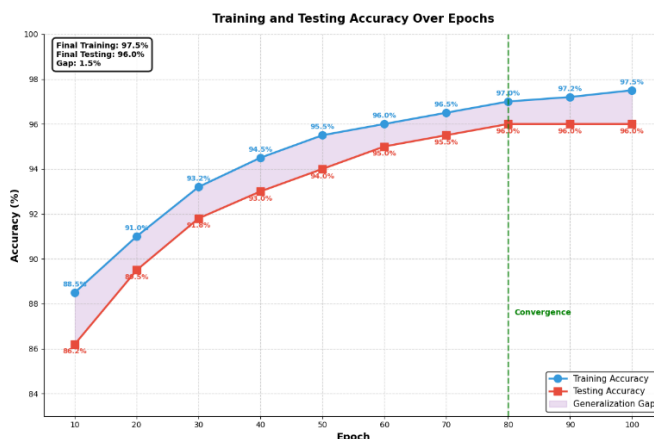


Fig. 12. Training and testing accuracy table.

The plot of the training and testing loss of the model over 100 epochs demonstrates the capability of the model to learn and generalize well. The training and testing losses at age 10 are 0.42 and 0.45, respectively, which means that the model is yet to learn much. Training loss and testing loss smoothly decrease as training advances, indicating improved model performance and enhanced precision in prediction. Training loss decreases to 0.18, and test loss decreases to 0.21 by 50 epochs, indicating that the model is learning the data patterns and still has good generalization. The loss keeps decreasing more slowly after 80 epochs, with 100 epochs having the results of 0.06 (training) and 0.10 (testing). In general, the table shows that there has been steady convergence of the model, with both training and testing loss reducing continuously, which is a sign of effective learning and very little overfitting (Fig. 13).

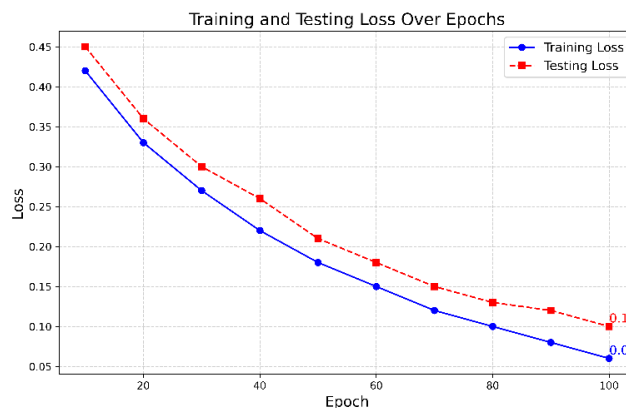


Fig. 13. Training and testing loss.

The convergence and optimization stability of the process is depicted in the graph after a number of iterations (Fig. 14). It presents three metrics, which are BSO fitness, placement accuracy, and salary R<sup>2</sup>, versus the number of iterations. The three metrics have a slow increase with increment of the iterations, which demonstrates that the optimization algorithm is indeed refining the solution. The BSO fitness is steadily increasing between 0.88 and 0.96 with a steady increase in the objective function. The same case is with the accuracy of placements, whereby the accuracy rises to 96% compared to 90%, indicating that the model is more predictive of successful placements as time goes by. Salary R<sup>2</sup> also increases to 0.91 as compared to 0.84, which indicates a better correlation between predicted and actual salaries. The convergence of the algorithm is also identified in the graph and the stability of the model as the measures reach a steady level in the later iterations, indicating that additional iterations do not significantly alter them.

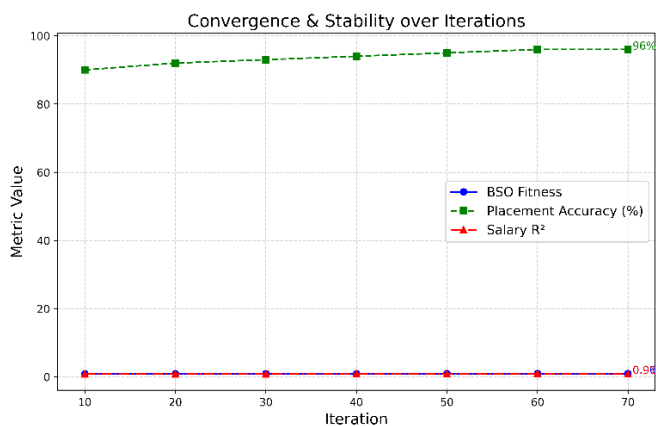


Fig. 14. Convergence & stability.

### C. Ablation Study and Computational Justification

An ablation study has been performed by successively removing each component of the hybrid architecture and observing the impact on the performance of ExCogGNet. The assessment was made based on both placement classification accuracy and the salary R<sup>2</sup> performance. Experimental results have shown that the overall ExCogGNet system has the best performance with placement classification accuracy of 96%, and a salary R<sup>2</sup> score of 0.91 (Table III). Removing the Transformer branch leads to a loss of ability for contextual interactions in learning. Removal of the GNN branch decreases relational dependency analysis skills among students. The removal of the BiLSTM branch affects ordered academic progressions in the learning process. Lastly, the exclusion of the MLP branch limits cognitive feature representation skills. It can be seen that all modules provide complementary skills to the whole system; hence, the improvements in the performance cannot be caused by excessive architectural complexity. In addition, even though using the hybrid architecture is computationally more expensive than using standalone models, the huge boost in the accuracy and robustness of multi-tasking makes this extra complexity worthwhile.

### D. Multi-Task Loss Weight Sensitivity Analysis

In order to measure the stability of the multi-objective optimization, further experimental studies were performed using variations in the loss balancing parameters  $\lambda_1$  and  $\lambda_2$  for classification of placements and salary regression, respectively. It can be observed from experimental data that an increase in  $\lambda_1$  results in better classification quality but increases salary prediction RMSE, and an increase in  $\lambda_2$  improves regression results but adversely affects classification results. The optimal results in terms of the trade-off between both objective functions have been found when  $\lambda_1 = 0.6$  and  $\lambda_2 = 0.4$ . Also, the small value of the performance deviation across different optimization iterations is evidence of the algorithm's stability and consistency, as shown in Table V.

TABLE III. ABLATION STUDY OF EXCOGGNET COMPONENTS

Model Variant	Placement Accuracy (%)	Salary R <sup>2</sup>
Full ExCogGNet (MLP + BiLSTM + Transformer + GNN + BSO)	96	0.91
Without Transformer	93	0.86
Without GNN	91	0.84
Without BiLSTM	89	0.82
Without MLP	88	0.80
Transformer Only	86	0.78
BiLSTM Only	84	0.75
MLP Only	81	0.71

TABLE IV. EFFECT OF LOSS WEIGHT VARIATIONS ON MULTI-TASK PERFORMANCE

$\lambda_1$ (Classification)	$\lambda_2$ (Regression)	Placement Accuracy (%)	Salary RMSE (INR)
0.2	0.8	88	22k
0.4	0.6	92	23k
0.5	0.5	94	25k
0.6	0.4	96	24k
0.8	0.2	97	29k

Fig. 15 shows the cross-validation 5-fold performance of the classification model. It is a summary of training and validation accuracy, precision, recall, F1-score, and AUC of five folds, the average, and standard deviation of these values. The findings show that there is a high consistency level amongst folds and that the average accuracy of training is 97.1% and average accuracy of validation is 96.0%. Precision and recall rates are also good, which means that the model has good capabilities of recognizing true positives and reducing falsely recognized positives and falsely recognized negatives. The overall classification performance is confirmed by the F1-score that balances recall and precision and has an average of 96.0. The solid averaged values of the AUC, 0.97, also contribute to the high discriminative ability of the model. Low standard deviations in all the metrics indicate that the performance was consistent and variable was small, and it shows the reliability of the model and its generalizability in various subsets of the data.

The minimal standard deviation seen in the five validations implies that there is consistent learning performance and less sensitivity to changes in training data due to the moderate number of samples used. This implies that the proposed framework is able to learn generalized features rather than learning specific characteristics of the dataset. It is, however, important to conduct further testing on a large, multisite database comprising students of varied backgrounds.

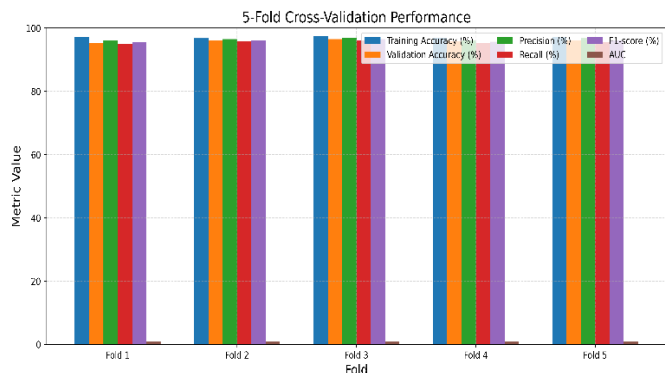


Fig. 15. 5-Fold cross-validation performance (classification).

The graph (Fig. 16) indicates the performance of the salary regression model when cross-validation is conducted five-fold. It shows three metrics, namely RMSE, MAE, and  $R^2$  against each fold. RMSE reduces steadily with each fold, with fold 1 of 26 k INR in fold 5 of 24 k INR, and MAE also reduces to 18 k INR, showing that the prediction errors are reduced, as well as the model becomes more accurate across folds. The values of  $R^2$  are quite high, between 0.89 and 0.92, and this shows a high correlation between the predicted and actual salaries. The graph shows the reliability and stability of the regression model since the measures indicate minimal change in different folds, which proves the ability of the model to generalize.

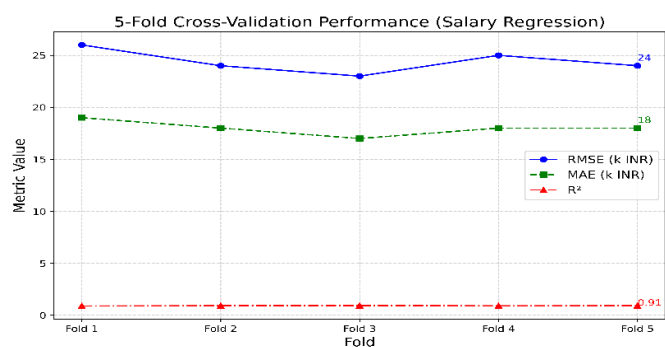


Fig. 16. 5-Fold cross-validation performance (regression).

The comparison table on the baseline gives the performance of several conventional machine learning and deep learning models in comparison with the proposed hybrid MLP-BiLSTM-Transformer-GNN model trained using Bolas Spider Optimization as a placement and salary predictor. The findings show that the classical models, i.e., SVM and random forest, are moderate in terms of placement accuracy of 78 % and 82 %, respectively, and relatively higher in terms of errors in salary prediction, represented by an RMSE value of 48k INR and  $R^2$  values of 0.62 and 0.71, respectively. CNN, BiLSTM, and

Transformer models are advanced deep learning models that achieve better performance, with placement accuracies ranging between 86% and 88% and decreasing lower values at RMSE, proving the benefit of sequential and contextual learning. Graph neural networks with no BSO can achieve 90 % placement accuracy, which shows the effectiveness of relational modeling. The hybrid model that is proposed and includes BSO is in a much better position than any of the baselines, with 96% placement accuracy, 96% F1-score, and 0.91 predictive  $R^2$  of salary.

Regarding the evaluation of the effectiveness of the suggested framework under small data situations, five-fold cross-validation techniques were applied, together with regularization-based optimization approaches within the experimentation stage. As a result, the hybrid system consistently performed at a high level of validation with low variability between different folds. This indicates stable learning behavior and insensitivity to variations in the training data set. The results indicate that the model learned generalized representations of features instead of being sensitive to specific aspects of the used data set, despite its relatively small size. However, the selected data set is a benchmark educational data set based on the Campus Recruitment data set, including around 215 records. Hence, it does not necessarily represent multi-faculty environments adequately due to its scale. Thus, while reporting a 96% placement success rate is a good benchmark result, further large-scale validation is required.

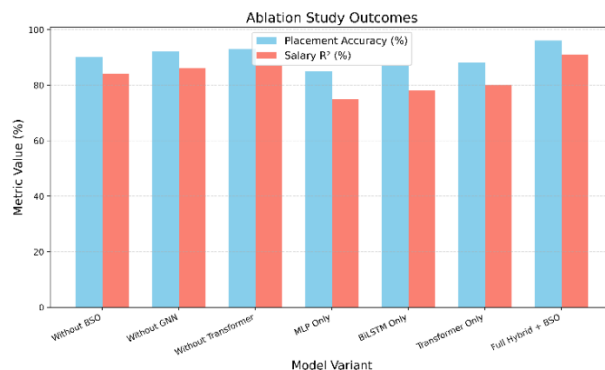


Fig. 17. Component contribution analysis.

The ablation study analysis, as shown in Table IV and Fig. 17, shows how each component individually contributes to the overall predictive performance of ExCogGNet. The complete hybrid model optimized through BSO resulted in maximum placement accuracy at 96% and maximum salary prediction at 0.91. The omission of BSO resulted in a decrement in optimization stability as well as a performance drop, thus emphasizing the importance of optimization of hyperparameters and fusion weights. Omission of the GAT branch resulted in weak relational learning in the case of students as well as prediction performance, thus underlining the importance of modeling peer relationships through graph learning. The exclusion of the Transformer resulted in the inability to learn from contextual feature interactions, while excluding the BiLSTM branch led to inadequate modeling of academic progress over time for students. Moreover, the standalone performance of MLP was significantly low since only static

features can predict employability. This analysis proves the need for a hybrid multi-branch architecture for effective multi-tasking.

To confirm that the performance difference between the proposed model and the strongest base, GNN without BSO, is statistically significant and cannot be explained by random

fluctuation, the McNemar test was used to test the placement classification results of the five cross-validation folds. The p-value of 0.003 confirms that the increase in accuracy between 90% and 96% is statistically significant at the level of significance of 0.01, proving that the gains are real and reproducible and not data-specific.

TABLE V. BASELINE COMPARISON FOR PLACEMENT AND SALARY PREDICTION

Model	Placement Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC	RMSE (k INR)	MAE (k INR)	R <sup>2</sup>
SVM	78	79	77	78	0.81	48	36	0.62
Random Forest (RF)	82	83	81	82	0.85	42	31	0.71
XGBoost	84	85	83	84	0.87	38	28	0.75
CNN	86	87	85	86	0.88	35	25	0.78
BiLSTM	87	88	86	87	0.89	33	23	0.80
Transformer	88	88	87	88	0.90	31	21	0.82
GNN without BSO	90	91	89	90	0.91	29	20	0.84
Proposed Hybrid MLP-BiLSTM-Transformer-GNN + BSO	96	96	95	96	0.97	24	18	

### E. Discussion

The suggested hybrid MLP-BiLSTM-Transformer-GNN model, optimized with Bolas Spider Optimization (BSO), provided an accuracy of 96% in the placement of the model, a 0.97 AUC, and a 0.91 salary prediction R-squared. These scores beat any of the baseline models. Furthermore, when interpreted relative to the overall salary distribution of the dataset, the observed RMSE indicates that the prediction deviation remains comparatively small, supporting the practical usefulness of the proposed framework for salary estimation tasks. Such as GNN (without BSO) and Transformer (without BSO) (90 and 88 percent accuracy, respectively). The statistical significance of such an improvement, proved with the help of the McNemar test ( $p = 0.003$ ), is the fact that the performance improvements are not specific to the dataset. The ablation analysis shows that BSO is essential for the globally optimal fusion weights and attention coefficients of the heterogeneous branches. The loss of BSO caused a significant decrease in performance, which proves that when the architecture is simply combined without any principled optimization, it may produce suboptimal learning. Even though BSO has been shown to exhibit stable convergence behavior along with robust predictive performance under the proposed framework, it is noteworthy that this research effort fails to provide an experimental benchmarking analysis of the optimization performance of BSO compared to other traditional approaches such as Bayesian optimization, grid search, and random search. The sensitivity analysis on the loss functions of the multi-task model shows that the proposed optimization framework is quite robust. From experimental results, it was found that varying settings of  $\lambda_1$  and  $\lambda_2$  had an effect on the balance between classification performance in placement and regression in salary prediction. The chosen setting led to convergence and high performance in both classification and regression. In addition, the repeated observation of the optimization process over several iterations shows the stability of BSO-based adaptive optimization.

The SHAP-based analysis revealed that the degree percentage (0.21), cognitive trait score (0.18), and technical skills score (0.12) are the most significant predictors, which supports the empirical argument of the cognitive-aware design of the model. The counterfactual SHAP explainability module offers feature-level insights into attribution and creates recommendation guidance for users by leveraging influential academic and cognitive attributes. These explanations can help career counselors and other institutional representatives gain insights into the significant determinants that influence employment results. However, in this particular research, no human-in-the-loop evaluation was conducted, nor did deployment testing or validation with career counselors or educators take place. Thus, this recommendation tool is more of a prototype that demonstrates how the proposed solution works through explainable means. The next step in future research is human-centered testing, usability evaluation, and institutional deployment research. Even though excellent prediction has been made through the use of the ExCogGNet, one challenge that should be taken note of is the limitation brought about by the size of the datasets used in the study. For deep hybrid networks, it is critical to use diversified datasets that help in increasing the generalization ability of the architecture due to its heterogeneity within different educational settings. The problem has been resolved through the use of five-fold cross-validation, dropout, optimized weights, and multi-tasking techniques to avoid overfitting. Furthermore, the constant accuracy levels obtained in validation, despite the small number of datasets, suggest stable learning behavior. Nevertheless, further validation using more diversified datasets from different universities will be needed.

### V. CONCLUSION AND FUTURE WORK

In this study, an Explainable Cognitive Graph Network, ExCogGNet, was proposed that integrates MLP-based encoding of cognitive traits in a static form, BiLSTM-based modeling of ordered academic progression stages, Transformer-based contextual feature attention, and Graph Attention Network learning of peer-level relational structures within one multi-task

model. The model also does placement classification and salary prediction, and Bolas Spider Optimization across the world optimizes hyperparameters, attention coefficients, and fusion weights, with a combined cross-entropy and mean squared error loss, alleviating heterogeneous architecture instability.

The experimental findings on the Campus Recruitment benchmark dataset demonstrate that ExCogGNet is 96% accurate in placement and has an AUC of 0.97 and a salary prediction RMSE of 24k INR and R2 of 0.91, which is much higher than baseline models (SVM 78%), Transformer (88%), and un-optimized GNN (90%). The improvement is statistically significant ( $p = 0.003$ ), as proven by McNemar. One important contribution is the counterfactual SHAP-based explainability module, which generates interpretable, recommendation-oriented insights for potential use in educational decision-support environments, converting feature attributions into individual, actionable recommendations on student skill improvement, which may support future decision-making processes for career counselors and institutional stakeholders after further practical validation. The ablation study results further confirm that each architectural component, including MLP, BiLSTM, Transformer, GNN, and BSO optimization, contributes complementary predictive capability, thereby justifying the hybrid multi-branch design of ExCogGNet. The ability of the framework to generate individualized recommendation-oriented explanations enhances its potential applicability in academic advisory systems, placement preparation programs, and institutional career guidance environments.

Though the introduced ExCogGNet framework exhibits good prediction performance on the Campus Recruitment dataset, its modest sample size makes comprehensive evaluation of its generalization ability challenging across a wide range of different educational organizations. As such, efforts will be directed at testing the framework on bigger educational datasets that include diverse student populations, behavioral learning history, socioeconomic features, and current academic progression information from multiple institutions. Despite the promising results obtained in terms of predictive performance and explainability, the present study has not considered user-focused validation through experiments with actual users such as students and career counselors. Therefore, the recommendatory explainability component should be seen as a prototype model for now. Usability experiments, human expert analysis, and real-world application will be conducted in subsequent research endeavors. Other potential research directions include methods of constructing graphs in scalable ways, employing light transformers, and utilizing federated learning techniques.

In the current dataset, milestone-based educational progression is included compared to fine-grained longitudinal academic history; future research might also include semester-based academic progress for improving sequential learning ability.

#### REFERENCES

- [1] M. Vivyn, T. Rani, and others, "Comprehensive Career Placement Predictor: An Analytical Tool for Optimizing Job Placement Outcomes," in 2024 International Conference on Power, Energy, Control and Transmission Systems (ICEPTS), IEEE, 2024, pp. 1–6.
- [2] K. Khalid and M. A. K. Al Zahrani, "Investigating the dynamics of workforce stability: The influence of perceived employability on career resilience, task performance, learning agility and employee retention," *Eur. J. Sustain. Dev.*, vol. 13, no. 3, pp. 389–389, 2024.
- [3] S. A. K. Malalha, M. Burhanuddin, and N. B. M. Yunus, "Unveiling the tapestry of machine learning: A comparative analysis of support vector machines random forests and neural networks in diverse applications," *Tuijin Jishu Journal Propuls. Technol.*, vol. 45, no. 3, p. 2024, 2024.
- [4] K. W. Manikrao, S. Allagi, W. Y. Leong, and M. Laddi, "An Explainable AI for Stock Market Prediction: A Machine Learning Approach with XAI and Deep Neural Networks," *J. Comput. Cogn. Eng.*, 2025.
- [5] Y. Yang and Y. Yao, "Graph Neural Network-Driven Modeling of Student Relationships and Optimization of Activity Organization," *Rev. Comput. Eng. Stud.*, vol. 12, no. 3, 2025.
- [6] S. F. Ahmed et al., "Enhancement of traffic forecasting through graph neural network-based information fusion techniques," *Inf. Fusion*, vol. 110, p. 102466, 2024.
- [7] H. Qawaqneh et al., "Bolas Spider Algorithm: A Novel Efficient Nature-Inspired Metaheuristic for Complex Continuous Optimization," *Int. J. Intell. Eng. Syst.*, vol. 19, no. 1, 2026.
- [8] R. N. Alibayeva and S. U. Allayarova, "From Data to Pedagogy: Leveraging Explainable Artificial Intelligence to Enhance Trust, Transparency, and Effectiveness in Intelligent Learning Systems," *Qubahan Techno J.*, vol. 4, no. 3, pp. 51–65, 2025.
- [9] C. Poblete and A. Aktas, "The role of conscious awareness in the relationship between entrepreneurs' hubris and the degree of internationalization," *Cogent Bus. Manag.*, vol. 12, no. 1, p. 2560651, 2025.
- [10] N. T. Mai, W. Cao, and W. Liu, "Interpretable knowledge tracing via transformer-Bayesian hybrid networks: Learning temporal dependencies and causal structures in educational data," *Appl. Sci.*, vol. 15, no. 17, p. 9605, 2025.
- [11] Q. Huang and Y. Zeng, "Improving academic performance predictions with dual graph neural networks," *Complex Intell. Syst.*, vol. 10, no. 3, pp. 3557–3575, 2024.
- [12] V. P. Dwivedi and X. Bresson, "A Generalization of Transformer Networks to Graphs," 2020, arXiv. doi: 10.48550/ARXIV.2012.09699.
- [13] E. Ofori and D. K. Dake, "Explainable artificial intelligence in LSTM transformer models for student performance analysis," *Discov. Comput.*, vol. 28, no. 1, p. 313, Dec. 2025, doi: 10.1007/s10791-025-09814-9.
- [14] H. Aljuaid, A. Alhothali, O. Al-Zamzami, and H. Assalahi, "TransGAT: Transformer-Based Graph Neural Networks for Multi-Dimensional Automated Essay Scoring," *ArXiv Prepr. ArXiv250901640*, 2025.
- [15] S. Dontu, S. R. Addula, P. K. Pareek, R. Vallabhaneni, and M. M. Adnan, "Attack detection from Internet of Things using TPE based self-attention based bidirectional long-short term memory," in 2024 international conference on intelligent algorithms for computational intelligence systems (IACIS), IEEE, 2024, pp. 1–6.
- [16] G. Vijh, P. P. Singh, S. Tiwari, S. Vijh, and S. Kumar, "Predictive Modelling of Salary Trends in the Tech Industry Using Machine Learning Algorithms," in *The International Conference on Recent Innovations in Computing*, Springer, 2024, pp. 38–47.
- [17] L. Wang, "Deep Learning for College Graduates Employment Prediction: A Computational Approach," *IEEE Access*, 2026.
- [18] A. R. Alsheyab, M. Alkhasawneh, and N. Shahin, "Job Market Cheat Codes: Prototyping Salary Prediction and Job Grouping with Synthetic Job Listings," 2025, arXiv. doi: 10.48550/ARXIV.2506.15879.
- [19] M. Xu, "Salary prediction using machine learning," *Sch. Rev. J.*, vol. Summer 2025, no. 13, Jun. 2025, doi: 10.70121/001c.139043.
- [20] A. BISWAS, "A Comparative Study for Salary Prediction Based on Different Models of Machine Learning," 2023.
- [21] Habibu Aminu, Badamasi Imam Yau, Fatima Umar Zambuk, Emanuel Ramsom Nanin, Abdulmutallib Abdullahi, and Ismail Zahraddeen Yakubu, "Salary Prediction Model using Principal Component Analysis and Deep Neural Network Algorithm," *Jan. 2024*, doi: 10.5281/ZENODO.10629245.

- [22] A. M. Ahmed, T. A. Rashid, and S. Ab. M. Saeed, "Cat Swarm Optimization Algorithm – A Survey and Performance Evaluation," 2020, doi: 10.48550/ARXIV.2001.11822.
- [23] Q. Cappart, D. Chételat, E. Khalil, A. Lodi, C. Morris, and P. Veličković, "Combinatorial optimization and reasoning with graph neural networks," 2021, doi: 10.48550/ARXIV.2102.09544.
- [24] M. M. Sesay, A. Ngunyi, and H. Imboga, "An explainable hybrid transformer–BiLSTM framework for multivariate household energy demand forecasting with weather integration," *Energy Inform.*, Feb. 2026, doi: 10.1186/s42162-026-00635-8.
- [25] M. A. Jahin, M. S. H. Shovon, M. F. Mridha, M. R. Islam, and Y. Watanobe, "A hybrid transformer and attention based recurrent neural network for robust and interpretable sentiment analysis of tweets," *Sci. Rep.*, vol. 14, no. 1, p. 24882, Oct. 2024, doi: 10.1038/s41598-024-76079-5.
- [26] M. Cedro and D. Martens, "GraphXAIN: narratives to explain graph neural networks," in *World Conference on Explainable Artificial Intelligence*, Springer, 2025, pp. 91–114.
- [27] Ben Roshan, "Campus Recruitment." 2020. [Online]. Available: <https://www.kaggle.com/datasets/benroshan/factors-affecting-campus-placement>