Deep-Feature Stacking Ensemble for Predictive Modeling of Student Academic Performance Using Socio-Behavioral and Institutional Factors

N. Rajesh[1], Mukkala Rachana[2], Konduru Rakesh[2], Thipparapu Anjali[2], Gantala Chandu[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (Data science)

[1,2]Vaagdevi Engineering College, Bollikunta, Warangal, 506005, Telangana, India

To Cite this Article
N. Rajesh, Mukkala Rachana, Konduru Rakesh, Thipparapu Anjali, Gantala Chandu, "Deep-Feature Stacking Ensemble for Predictive Modeling of Student Academic Performance Using Socio-Behavioral and Institutional Factors", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 04, April 2026,pp: 410-418, DOI: http://doi.org/10.64771/jsetms.2026.v03.i04.pp410-418
Submitted: 28-02-2026 Accepted: 01-04-2026 Published: 09-04-2026

Abstract

This study addresses the growing need for intelligent, data-driven systems in the education sector to analyze student performance and behavioral patterns. Educational datasets encompass critical factors such as study habits, attendance, and engagement, which can be utilized to predict academic outcomes and stress levels. However, traditional machine learning models including Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), Decision Tree (DT), Logistic Regression (LR), and Linear Regression (LinR) often struggle to capture complex non-linear relationships and sequential dependencies within such data. Additionally, these models face challenges such as overfitting, limited generalization, and reduced interpretability, particularly when handling both classification and regression tasks within a unified framework. To address these limitations, this research proposes a hybrid model, Bidirectional LSTM with Attention and Rule-based Model (BARM). The model integrates Bidirectional Long Short-Term Memory (BiLSTM) with an attention mechanism to effectively capture temporal patterns and feature importance, alongside rule-based approaches including Optimal Decision Rule List Classifier (ODRLC) and Optimal Decision Rule List Regressor (ODRLR) for enhanced interpretability. The system performs classification tasks such as Final Grade and Stress Level prediction, as well as regression tasks like Exam Score estimation. A dynamic model selection mechanism is employed to evaluate and choose the most suitable model based on performance metrics. The proposed system is implemented as a Flask-based web application, enabling real-time predictions, batch processing, and visualization through Exploratory Data Analysis, By improving scalability, usability, and decision-making in academic environments.

Keywords: Educational Data Mining, Machine Learning, Deep Learning, Hybrid Model, Bidirectional LSTM (BiLSTM), Attention Mechanism, Rule-Based Learning

1. Introduction

Over the past few years, institutions of higher learning have increasingly adopted information and communication technology (ICT)-based learning approaches, resulting in the generation of vast amounts of educational data. These data, sourced from Learning Management Systems (LMS), Student Information Systems (SIS), video-assisted courses, and other digital learning platforms, offer valuable insights into student behavior and academic performance patterns [1]. However, this influx of data brings new challenges, including data complexity, diversity, and high dimensionality. To make sense of this vast array of information, Educational Data Mining (EDM) has gained recognition as a significant tool, enabling researchers and educators to analyze data and predict student academic success [2,3]. EDM is an increasingly expanding domain that applies data-driven approaches to analyze and improve various aspects of education, as shown in figure 1. With the increasing availability of data from digital learning platforms, student management systems, and online learning behaviors, EDM provides opportunities to uncover patterns and insights that can inform personalized learning, early intervention strategies, and policy decisions [3]. By harnessing advanced approaches including machine learning and data representation, EDM has proven effective in addressing key educational challenges, including dropout prediction, performance forecasting, and engagement analysis.

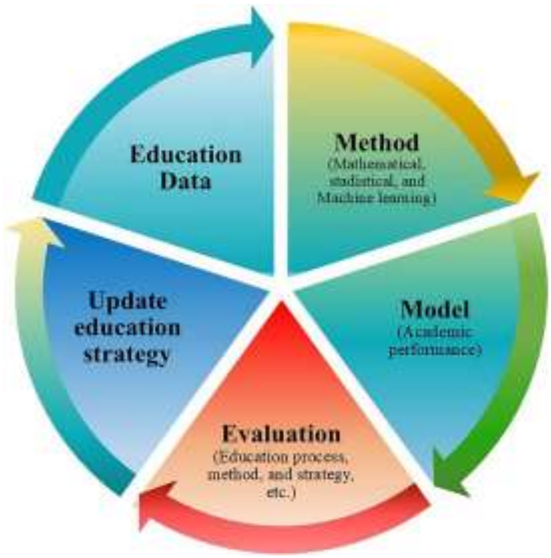

Figure. 1: Student academic performance

The task of predicting student academic success has become a critical research area, as accurate predictions can guide interventions and support strategies, particularly for students at risk of poor performance or dropout [4]. Traditional predictive models have primarily relied on structured data, such as test scores, attendance records, and demographic information. While these features provide valuable insights, they often oversimplify the multifaceted nature of student performance by overlooking behavioral data, which can reveal deeper patterns of engagement, effort, and learning dynamics [5]. Behavioral data, such as online learning interactions, video engagement, and assignment submission patterns, are inherently unstructured and multidimensional, making them challenging to analyze using conventional methods. As a result, valuable information embedded in these data remains untapped in many predictive frameworks. This limitation underscores the need for innovative approaches that can effectively extract and model behavioral data to enhance the precision and robustness of academic success predictions[6,7]. Data-driven methods have greatly transformed this field, as machine learning and deep learning techniques now enable the analysis of various data types, including engagement metrics, grades, and participation patterns [8]. However, a persistent challenge in predicting student success is the complexity and high dimensionality of behavioral data collected from online learning platforms. Traditional predictive methods typically rely on one-dimensional numerical or text-based data, which can overlook latent patterns within the data that may significantly influence academic outcomes [9].

2. Literature Survey

Tang et al. [10] proposed method employed an optimization strategy to concurrently configure and train the deep neural networks within their ensemble system. Furthermore, the proposed ensemble model used weighted voting among its learning components for more accurate prediction. Put simply, the suggested approach enhanced the accuracy of academic performance predictions for students not only by employing weighted ensemble techniques, but also by optimizing the parameters of deep learning models. Adefemi et al. [11] aimed to bridge this gap by proposing a deep learning model to predict student academic performance with greater accuracy. The approach combined a convolutional neural network (CNN) and a bidirectional gated recurrent unit (BiGRU) network to enhance predictive capabilities. To improve the model's performance, they addressed key data preprocessing challenges, including handling missing data, addressing class imbalance, and selecting relevant features.

Yu et al. [12] employed five-fold cross-validation to assess the model's performance. In comparison with the four single models, the two fusion models based on the four single models both showed significantly better performance. The prediction accuracies of the bagging fusion model and stacking fusion model were 83% and 84%, respectively. Ruiz-Camacho et al. [13] comprised 1014 Spanish university students (64.5% women, 35.5% men; M = 20.56, SD = 3.50). Participants completed the Academic Stressors Scale (E-CEA) and the Stress Responses Scale (R-CEA). Hierarchical regression analyses were conducted in two blocks: sociodemographic variables were entered in the first block, followed by academic stressors in the second. Results showed that academic stressors accounted for substantial variance in all five stress response dimensions: negative thoughts (47.8%), physical exhaustion (39.5%), physical agitation (32.9%), irritability (29.7%), and sleep disturbances (26.8%). The most recurrent predictors were beliefs about performance, exams, and academic overload.

Hahn et al. [14] aimed to explore whether the correlation structures could also be confirmed under different conditions, particularly in the context of the COVID-19 pandemic. To answer these questions, data collected under pandemic study conditions (NLA1 = 510) and post-pandemic study conditions (NLA2 = 433) were used and analysed by SEM. The results showed that the Study Demands–Resources model was applicable in the two different contexts based on its validation in both study contexts. Ismiyana Putri et al. [15] aimed to identify student stress and academic performance using machine learning. A Systematic Literature Review (SLR) was conducted following the PRISMA protocol, focusing on articles published between 2019 and 2025 from Scopus and other academic databases. Twenty relevant studies were reviewed, covering algorithm types, data preprocessing, feature types, and model performance. Random Forest, Support Vector Machine (SVM), and Neural Networks emerged as the most frequently used algorithms. Predictive features included psychological, behavioral, and physiological data collected from standardized instruments (e.g., DASS-42, PHQ-9), wearable biosensors, and facial expression analysis. Common preprocessing steps included text cleaning, tokenization, normalization, and dimensionality reduction.

3. Proposed System

The system integrates data ingestion, preprocessing, model training, evaluation, and deployment into a unified pipeline for student performance prediction. It begins with user authentication and dataset loading, followed by structured preprocessing to prepare data for modeling. Figure. 2 illustrates how both existing ML models RF, SVM, GB, DT, LR and the proposed BARM model operate within the same framework. The architecture supports both classification tasks FinalGrade, StressLevel and regression tasks ExamScore. The BARM model combines BiLSTM, Attention, ODRLC, and ODRLR to enhance predictive performance. A model comparison and selection module evaluates all models using performance metrics.

Data Acquisition and Ingestion: The system begins by collecting student-related data from structured sources such as CSV files, which include academic, behavioral, and institutional attributes. This data is loaded into a processing environment where it is organized into a tabular format. Proper data ingestion ensures consistency and prepares the dataset for further transformation and analysis.

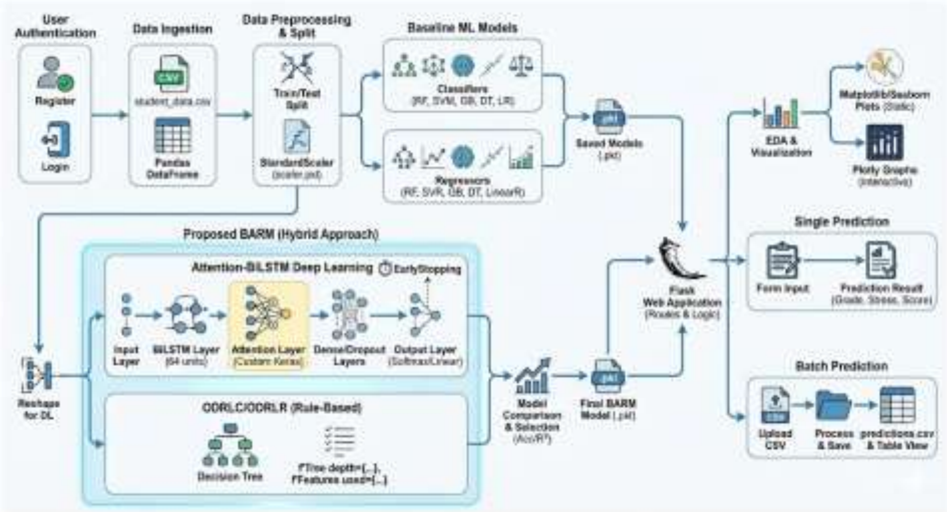

Figure. 2: Proposed System Architecture

Data Preprocessing and Transformation: In this stage, the collected data undergoes cleaning, normalization, and scaling to remove inconsistencies and improve quality. The dataset is then divided into training and testing subsets to enable proper evaluation of model performance. This step ensures that the data is standardized and suitable for both traditional and advanced analytical methods.

Baseline Model Evaluation: Multiple baseline models are applied to the processed data to establish initial performance metrics for 2CA1RT. These models help in understanding the dataset characteristics and provide a reference for further improvements. The outputs generated here are stored and later used for comparative analysis.

Hybrid Model Processing and Feature Learning: The system utilizes a hybrid framework that processes input data through deep learning layers to capture complex patterns and sequential dependencies. Simultaneously, rule-based components analyze structured decision paths to enhance interpretability. This combined approach enables the system to handle both complexity and transparency in predictions.

Model Comparison and Selection: All trained models are evaluated using performance metrics, and a comparison mechanism identifies the best-performing model. This dynamic selection process ensures that the most accurate and reliable model is chosen for deployment. It also helps in improving overall system robustness and adaptability.

Deployment and Prediction Interface: The final selected model is integrated into a web-based application that allows users to perform both single and batch predictions. The system provides outputs such as grades, stress levels, and scores along with visualizations for better understanding. This stage ensures practical usability and real-time interaction with the predictive system.

4. Results discussion

The results and discussion section presents the outcomes obtained after executing the system and evaluating its performance. It focuses on analyzing how effectively the system performs based on defined evaluation metrics. The results include numerical values, graphical representations, and comparisons that help in understanding system behavior. This section interprets the outputs to identify patterns, trends, and performance differences. It also highlights the strengths and weaknesses observed during execution. Visual tools such as graphs and charts are used to support the analysis. The discussion provides meaningful insights derived from the results.

Confusion Matrix (CM)

	Pred 0	Pred 1	Pred 2	Pred 3
True 0	1398	22	18	7
True 1	15	1218	13	12
True 2	12	16	1431	15
True 3	16	18	9	1220

Classification Report (CR)

	precision	recall	f1-score	support
Average	0.97	0.97	0.97	1474
Excellent	0.97	0.97	0.97	1445
Good	0.96	0.97	0.96	1258
Poor	0.97	0.97	0.97	1263
accuracy			0.97	5440
macro avg	0.97	0.97	0.97	5440
weighted avg	0.97	0.97	0.97	5440

Figure. 3: Confusion Matrices and Classification Report obtained for Final Grade Classification BARM

Figure 3 illustrates the confusion matrix and classification report for the proposed BARM model. It depicts that the model achieves the highest accuracy of approximately 0.97, indicating superior classification performance. The confusion matrix shows very high diagonal values such as 1398, 1218, 1431, and 1220 with minimal misclassification. The classification report confirms that precision, recall, and F1-scores are consistently around 0.97 across all classes (0–3). This indicates excellent prediction capability and balanced performance. The model effectively distinguishes between all categories with high reliability.

Confusion Matrix (CM)

	Pred 0	Pred 1	Pred 2
True 0	1077	5	22
True 1	13	1544	37
True 2	28	36	2678

Classification Report (CR)

	precision	recall	f1-score	support
High	0.98	0.98	0.98	2742
Low	0.96	0.98	0.97	1104
Medium	0.97	0.97	0.97	1594
accuracy			0.97	5440
macro avg	0.97	0.97	0.97	5440
weighted avg	0.97	0.97	0.97	5440

Figure. 4: Confusion Matrices and Classification Report obtained for Stress Level Classification BARM

Figure. 5: Scatter Plots obtained for Exam Score Regression BARM

Figure 4 illustrates the confusion matrix and classification report for the proposed BARM model. It depicts that the model achieves the highest accuracy of approximately 0.97, indicating superior performance. The confusion matrix shows very high diagonal values such as 1077, 1544, and 2678 with minimal misclassification. The classification report confirms that precision, recall, and F1-scores are consistently around 0.97–0.98 across all classes (0–2). This indicates highly accurate and balanced predictions. The model effectively distinguishes between all stress levels with minimal error.

Figure 5 illustrates the scatter plot for the proposed BARM model in ExamScore regression. It depicts a strong alignment of predicted values with the diagonal reference line across the full range (40–100). The distribution of points closely follows the ideal prediction line, indicating high correlation. Minimal deviation from the reference line suggests accurate predictions. The model effectively captures variations in input data. The spread of points reflects balanced and consistent performance.

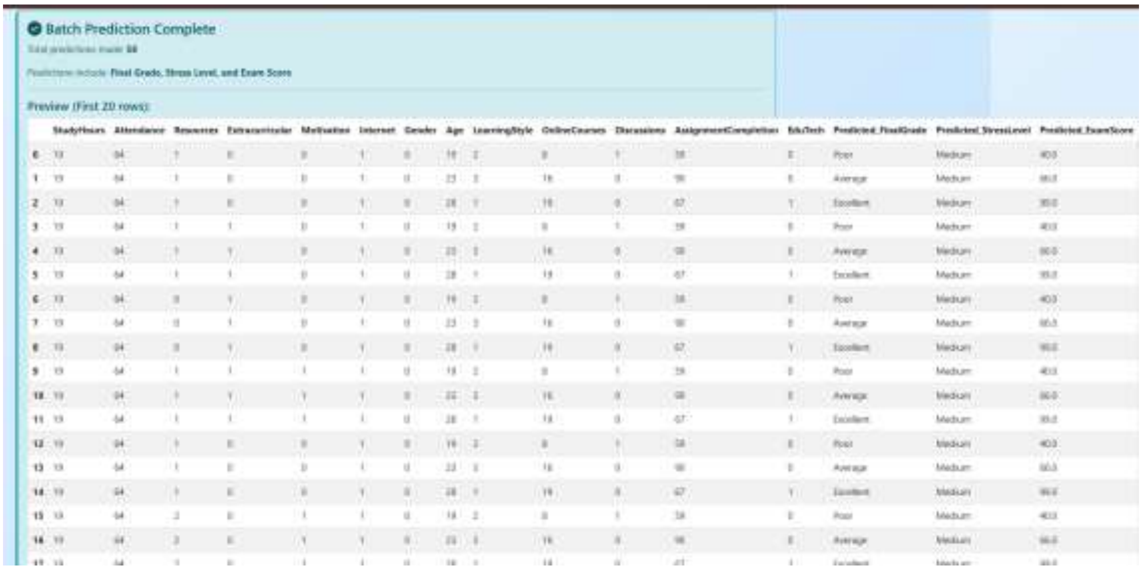

Figure. 6: Batch Prediction Interface

Figure 6 depicts the batch prediction interface of the system, which processes multiple input records simultaneously. It illustrates the structured tabular output containing predicted values for Final Grade (0–3), Stress Level (0–2), and Exam Score (40–100). The figure shows that predictions are generated for multiple samples, enabling large-scale analysis. The interface supports efficient handling of batch datasets. It highlights the integration of prediction results with input features. The system ensures consistent output formatting for all records.

Table 1: Final Grade – Model Performance

Model	Accuracy	Precision	Recall	F1 Score
RF	0.8509	0.8514	0.8509	0.8509
SVM	0.4553	0.4605	0.4553	0.4498
GB	0.5026	0.5069	0.5026	0.5003
DT	0.2864	0.1534	0.2864	0.1984
LinR	0.2860	0.2804	0.2860	0.2394
Proposed BARM	0.9682	0.9682	0.9682	0.9682

Table 1 presents the performance comparison of various classification models for predicting final grades. The Proposed BARM model significantly outperforms all baseline models, achieving the highest accuracy, precision, recall, and F1 score (0.9682). Among traditional methods, RF shows strong performance with balanced metrics around 0.85. In contrast, SVM and GB demonstrate moderate results with accuracies below 0.55. DT and LR perform poorly, indicating limited capability in capturing complex patterns. The consistent metrics of BARM suggest robust and reliable classification performance.

Table 2: Stress Level – Model Performance

Model	Accuracy	Precision	Recall	F1 Score
RF	0.8985	0.8988	0.8985	0.8986
SVM	0.5344	0.6422	0.5344	0.4119
GB	0.5526	0.6036	0.5526	0.4565
DT	0.5040	0.2541	0.5040	0.3378
LinR	0.5040	0.2541	0.5040	0.3378
Proposed BARM	0.9741	0.9741	0.9741	0.9741

Table 3: Exam Score – Model Performance

Model	MAE	MSE	RMSE	R^2 Score
RF	1.7103	49.5114	7.0364	0.8409
SVM	14.5936	298.3599	17.2731	0.0412
GB	14.3744	278.4714	16.6875	0.1051
DT	15.1999	308.8406	17.5739	0.0075
LinR	15.2205	310.1464	17.6110	0.0033
Proposed BARM	0.8270	25.5347	5.0532	0.9179

Table 2 compares different models for stress level classification. The Proposed BARM model again achieves the best performance, with all evaluation metrics reaching 0.9741. Random Forest is the second-best model, delivering high accuracy and balanced precision and recall values close to 0.90. SVM, GB show moderate effectiveness but with noticeable drops in recall and F1 score. DT and LR yield relatively weak results, indicating poor generalization. The higher precision of SVM suggests some class bias despite lower overall accuracy.

Table 3 presents regression model performance for predicting exam scores using error metrics and R^2 score. The Proposed BARM model achieves the lowest MAE (0.8270), MSE (25.5347), and RMSE (5.0532), along with the highest R^2 score (0.9179), indicating excellent prediction accuracy. RF performs well among baseline models with relatively low error values and a strong R^2 of 0.8409. Other models, including SVM and GB, show higher errors and significantly lower R^2 scores. DT and LinR on exhibit the weakest performance, with minimal explanatory power. The results demonstrate that BARM effectively captures underlying patterns in exam score prediction.

5. Conclusion

The research successfully developed an intelligent prediction system for analyzing student performance and stress levels using both ML models and the proposed BARM model. The system integrates preprocessing, EDA, and multiple models including RF, SVM, GB, DT, LR, and LinR within a unified framework. The proposed BARM model, which combines BiLSTM, Attention, ODRLC, and ODRLR, demonstrated superior performance compared to all baseline models. In Final Grade classification, BARM achieved an accuracy of approximately 0.97, outperforming RF (~0.85) and significantly surpassing SVM, GB, DT, and LR models. Similarly, in Stress Level classification, BARM maintained high accuracy around 0.97–0.98, while other models showed moderate to low performance. In regression tasks, BARM achieved a high R^2 score (0.91) with lower error values, whereas LinR and other models showed limited predictive capability. The hybrid architecture effectively captured both complex patterns and interpretable decision rules, leading to improved generalization and reliability. The integration with a Flask-based interface enabled real-time prediction, batch processing, and visualization, making the system practical and scalable.

References

1. Santthosh Saai Reddy Purmani. (2026). Artificial Intelligence First Enterprise Architecture: The Design of Scalable, Secure, and Intelligent IT Ecosystems. American Journal of AI Cyber Computing Management, 6(1(2)), 1–8. https://doi.org/10.64751/ajaccm.2026.v6.n1(2).pp1-8
2. Patel, S., & Patyrykin, K. (2025). Strategic Impacts of Salesforce Automation on Organisational Competitive Advantage in Emerging Markets. Journal of Posthumanism, 5(12), 357–372. https://doi.org/10.63332/joph.v5i12.3782
3. Vasagam, M., Kumar, A., & Garg, A. (2026). Learning Execution Plan Embeddings for Multi-Dimensional Query Resource Prediction. IEEE Access.
4. Kalae, U. K. (2021). Enhancing data analytics and reporting efficiency using Power BI and SQL in cloud computing environments. Journal of Computational Analysis and Applications, 29(6), 2021. https://doi.org/10.48047/jocaaa.2021.29.06.48

5. Poojari, R. Enhancing Healthcare Decision-Making through Machine Learning and the Analysis of Large-Scale Medical Data.
6. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
7. Prodduturi, S. M. K. To Secure Your Paper as Per UGC Guidelines We Are Providing A ElectronicBar code.
8. Gaddam, S. From Fixed Specifications to Self-Adapting Systems: A Machine Learning Perspective on Software Engineering.
9. Asif, R.; Merceron, A.; Ali, S.A.; Haider, N.G. Analyzing undergraduate students' performance using educational data mining. Comput. Educ. 2017, 113, 177–194. [Google Scholar] [CrossRef]
10. Explainable AI Framework for Policy-Compliant Anomaly Detection in Data Pipelines. (2025). Inaternational Journal of Communication Networks and Information Security, 16(4). https://doi.org/10.48047/ijcnis.16.4.2111
11. Adefemi, K.O.; Mutanga, M.B.; Jugoo, V. Hybrid Deep Learning Models for Predicting Student Academic Performance. Math. Comput. Appl. 2025, 30, 59. https://doi.org/10.3390/mca30030059
12. Yu, F.; Liu, X. Research on Student Performance Prediction Based on Stacking Fusion Model. Electronics 2022, 11, 3166. https://doi.org/10.3390/electronics11193166
13. Ruiz-Camacho, C.; Gozalo, M. Predicting University Students' Stress Responses: The Role of Academic Stressors and Sociodemographic Variables. Eur. J. Investig. Health Psychol. Educ. 2025, 15, 163. https://doi.org/10.3390/ejihpe15080163
14. Hahn, E.; Kuhlee, D.; Zimmermann, J.; Serrano-Sánchez, J. The Mediating Role of Perceived Stress and Student Engagement for Student Teachers' Intention to Drop Out of University in Germany: An Analysis Using the Study Demands–Resources Model Under Pandemic and Post-Pandemic Conditions. Educ. Sci. 2025, 15, 719. https://doi.org/10.3390/educsci15060719
15. Ismiyana Putri, Dwi & Putra, Mardi Yudhi & Nurul Alfian, Ari & Arifin, Rita & Shadiq, Jafar & Indaryani, Putri. (2025). Machine Learning Models for Student Stress and Academic Performance Prediction: A Systematic Literature Review. 1-5. 10.1109/ICIC68054.2025.11309547.