International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

Unleashing Deep Learning for Academic Success: A Survey of Predictive Models in Educational Data Mining

V. Loganayaki[1], N. Balakumar[2]

[1]Research Scholar, Department of Computer Science, United College of Arts and Science, Coimbatore, Tamil Nadu, India
Email: *logabalu415phd[at]gmail.com*

[2]Associate Professor and Head, Department of Computer Science, United College of Arts and Science, Coimbatore, Tamil Nadu, India
Email: *msg2balakumar[at]gmail.com*

Abstract: ***Educational Data Mining (EDM) is a fast-emerging area of influence in Educational Informatics which aims to enhance educational outcomes by applying foundations and techniques of data mining to educational data. The prediction of students' academic performance is one of the many applications of EDM, which is of great significance in providing timely academic support to at-risk students once detected in their early stages. This review examines the application of Deep Learning (DL) techniques for student performance prediction within EDM. The increasing availability of educational data and the limitations of traditional statistical and machine learning approaches have encouraged the adoption of deep learning models capable of automatically learning complex feature representations. The study reviews major deep learning architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), attention-based models, graph neural networks (GNNs), and hybrid frameworks. A comparative analysis of datasets, evaluation metrics, strengths, limitations, and predictive performance is presented. The findings indicate that deep learning models generally achieve high predictive accuracy, particularly when temporal and behavioural learning patterns are incorporated. The review also highlights challenges related to interpretability, computational complexity, and data dependency. Finally, future research directions are identified to support the development of reliable, scalable, and practical student performance prediction systems.***

Keywords: Educational Data Mining, Student Performance Prediction, Deep Learning, Learning Analytics, Convolutional Neural Networks, Long Short-Term Memory Networks, Educational Artificial Intelligence.

1. Introduction

Applying data mining techniques to educational data in order to evaluate and comprehend learning processes and educational systems is the goal of Educational Data Mining (EDM), an interdisciplinary field of study [1]. Data analysis in educational settings aims to improve our understanding of student learning, spot trends in student behaviour, and back up data-driven decision-making in the field. Extracting valuable, actionable, and undiscovered information from massive datasets is known as Data Mining (DM). In the educational sector, these techniques are adapted to handle the unique characteristics of educational data, which are often complex, hierarchical, temporal, and context dependent [2]. Figure 1 shows the process EDM.

A great deal of educational data has been generated as a result of the fast expansion of educational technology. EDM draws contributions from different educational settings and research domains, which can be broadly categorized as follows:

- **Offline (Traditional) Education**: It focuses on face-to-face teaching and learning environments where knowledge and skills are transmitted through direct interaction between instructors and students. Data collected in such settings include student performance records, assessment results, attendance, and classroom behaviour.
- **Learning Management Systems (LMS) and Online Education:** The delivery of educational information using the Internet is referred to as online education, e-learning, or web-based learning. Learning Management Systems (LMS) are tools for communication, collaboration, teaching and learning administrator, and assessment. These systems can generate huge amounts of data over time, including log files and database records, that record the interactions students make, how they move around, and how engaged they are with learning materials.
- **Adaptive Educational Hypermedia Systems (AEHS) and Intelligent Tutoring Systems (ITS):** These are designed to tailor instruction and feedback to each student's requirements and, consequently, have as an objective to adapt the content and the presentation. They gather information like user models, interactions and learning progress monitors [3].

Applications of EDM: EDM encompasses a wide variety of tasks aimed at analysing and improving educational processes. Several researchers have proposed frameworks to categorize these tasks based on their objectives and applied methodologies. According to Baker [4], the key application areas of EDM include:

- Enhancement of both student and domain models,
- A review of the instructional resources offered by learning management systems, including
- Extensive studies on the many aspects of learning and how people learn.

Similarly, according to Castro [5],

- Evaluation of students' academic progress,
- Instructional strategies that are both flexible and tailored to each student's needs,
- Assessing course materials as well as virtual education,

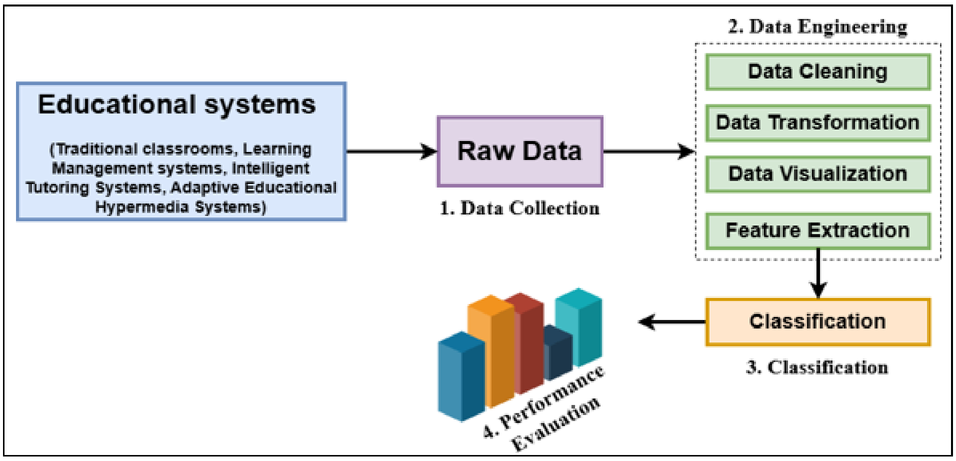

Figure 1: Educational Data Mining

- The ability for both teachers and students to provide and receive feedback, and
- Finding unusual or out-of-the-ordinary patterns of cognitive development.

Among the various educational tasks addressed through EDM, student performance prediction has emerged as one of the most prominent and widely studied areas. This task focuses on forecasting academic outcomes such as grades, pass/fail status, or risk of dropout by analysing historical and behavioural data. Academic achievement among students is the gold standard by which every nation's educational system is measured. To rephrase, student performance is defined as the degree to which learners accomplish both short-term and long-term goals [6]. One of the most important criteria for determining a university's excellence according to rankings is its academic record. Institutions with stellar reputations and records of success tend to rise in the rankings. From the standpoint of the student, keeping up their exceptional academic performance improves their chances of getting a job, as companies primarily look at this as an indicator of potential success [7]. Educators can help students who are struggling to learn by anticipating their performance and implementing interventions at an early stage. Research across different educational settings, particularly e-learning platforms, LMS, and ITS, shows that predictive tasks commonly rely on data mining techniques such as regression [8], classification [9], clustering [10], and association rule mining [11]. A dependent variable and its connection to one or more independent variables can be discovered using regression analysis. The term "classification" refers to a process whereby a training set of previously labelled objects and quantitative data on one or more item properties are used to arrange individual things into categories. Association rule mining extracts valuable insights from massive datasets by discovering intriguing correlations between variables and presenting them as robust rules ranked by potential relevance. The term "clustering" refers to the process of dividing a dataset into smaller groupings of observations with shared characteristics.

But the conventional statistical approaches used to forecast students' performance have a number of drawbacks. These models often have strong assumptions made about distributions and independence of data, linear correlations between variables, etc. The educational data, however, generally suffers from some drawbacks such as non-linearity, noise and high dimensionality, and as a consequence, statistical methods do not permit to capture a complex education pattern [12, 13]. Furthermore, statistical models are usually not capable of processing a large amount of data and dynamic behavioural data from the modern e-learning environment. Admission to complex teaching situations therefore could result in a drop in their prediction skills.

This is because of the limitations in statistical approaches, and ML techniques have been applied widely in student performance prediction. Use of ML models such as Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbour (KNN), Random Forest (RF) or Neural Networks (NN) to learn complex patterns from data without assuming what underlying distribution is required [14]. These models have been found to greatly improve the accuracy of predictions, demographic data and behavioural data when utilizing various information, especially academic data, in an efficient way. Still, a lot of ML models need a lot of feature engineering, and effective parametrization, which is tricky. In addition, a lack of interpretability prevents teachers from getting an understanding of what aspects of a forecast influence the results of that forecast. This can also be prone to overfitting, particularly if there are not enough or unbalanced training data.

In recent years, DL methods have captivated scholars' attention due to their capability to self-organize multi-level feature representations from huge amounts of data. Much research interest has been paid to the prediction of student performance from recent years, and the ability of DL methods to self-organize multi-level feature representations from massive amounts of data has attracted research interest. Temporal and sequential patterns in educational data were successfully captured using CNN, RNN and LSTM based models [15]. DL models are particularly effective in analysing log data from LMS platforms and sequential learning behaviors over time. In the context of EDM, the study is an in-depth analysis of student success models prediction.

In EDM, the study provides a detailed assessment of student success models prediction. The survey looks into different DL approaches, its model architectures, feature representations, training strategies, pros and cons of these approaches. Furthermore, a comparative evaluation of the existing approaches underlines their properties of predictability, interpretability and applicability in various educational contexts and shows up open challenges and open gaps of research and theory, which require further study.

1.1 Research Methodology

Relevant studies were retrieved from major scientific databases using predefined search terms related to deep learning, educational data mining, student performance prediction, dropout prediction, and learning analytics. The search covered publications published between 2022 and 2025. The article selection process involved identification, screening, eligibility assessment, and final inclusion of studies. Inclusion criteria comprised peer-reviewed journal articles published in English that focused on student performance prediction. Studies unrelated to student performance prediction, review articles, and papers lacking experimental evaluation were excluded. Bibliographic management and duplicate removal were performed using Mendeley. After applying the inclusion and exclusion criteria, 15 relevant studies were selected for detailed analysis and synthesis.

Section II examine some DL methods which try to predict student's performance in class. In Section III, we analyse these DL models in detail based on various metrics, datasets and techniques. This is followed by a discussion of the effectiveness of the existing methods in Section IV. Finally, the article concludes and some further study vehicles suggested in Section V.

2. Survey on Deep Learning Models Based on Student Performance Prediction

Currently, research in education has focused on using DL for the prediction of student outcomes as it has excellent representation and analysis of complex data structures.
Begum & Padmannavar [16] proposed a hybrid model of CNN and Binary Particle Swarm Optimization (BPSO) for Student performance prediction. To remove redundant and most informative attribute of student before classification, BPSO is used to make an optimal subset of student attribute. The selected features were then fed into a CNN to model complex nonlinear relationships in educational data. The approach was designed to support both binary and multi-class student performance prediction tasks.

Poudyal et al. [17] devised academic achievement prediction using a 2D-CNN hybrid model. First, data were pre-processed and merged into a single dataset containing demographic information, assessment results, VLE interactions, and final student outcomes. Second, the pre-processed one-dimensional feature vectors were transformed into two-dimensional inputs suitable for CNN processing. Third, a hybrid CNN architecture was constructed by combining two parallel 2D CNN models with different depths. A single feature vector was created by merging the two branches' results. The last step in making the forecast was to run the combined characteristics through fully linked layers.

Liu et al. [18] presented an MCAG model that utilised DL, which incorporated CNN, Attention Mechanism (AM), Gate Recurrent Unit (GRU), and Maximum Information Coefficient (MIC) to predict students' achievements. Initially, the relationships between student achievement and multiple influencing factors were analyzed using the MIC module to identify highly correlated attributes and reduce the dimensionality of the input features. Then, the selected features were processed by two parallel sub-models: a CNN-AM network to extract spatial feature representations and a GRU network to capture the temporal characteristics of student learning behaviors. A fully linked layer was employed to fuse the representations and get the final forecast of student success.

Wen & Juan [19] developed an DNN-based early student performance prediction approach that represents learners' online activity sequences as a three-dimensional tensor capturing students, time phases, and activity types. To address the high dimensionality of these representations, a pre-trained deep autoencoder was used to extract low-dimensional latent features from partial or complete learning sequences. In order to provide end-to-end predictions about students' performance, the characteristics that were retrieved were input into a feedforward neural network.

Huang & Chen [20] developed an Academic Performance Prediction model based on Temporal Graph Neural Networks (APP-TGN) to capture temporal learning behaviors and interaction patterns in MOOCs. Initially, online learning activity logs were collected, cleaned, and transformed, and dynamic temporal graphs were constructed to represent students' learning behaviours and their temporal interactions. Then, a low–high filtering temporal graph network was applied to the dynamic graphs to learn local representations of students' academic performance, which was then sent to global sampling module which reduced data bias and mitigate false correlations. Finally, these representations were fused using a MHSA and fed into an MLP classifier to predict students' academic performance.

Leelaluk et al. [21] suggested an Attention-Based Artificial Neural Network (Attn-ANN) model to improve student performance prediction by jointly modeling temporal learning behaviors with activity features. Initially, students' weekly learning behaviour sequences were processed using a recurrent neural network to extract time-dependent hidden representations for each lecture. The next step was to determine which lectures were more important by using a time-dimension attention mechanism, and for each lecture, the most impactful learning activities were identified using a feature-dimension attention mechanism. The last step in using a fully connected neural network to determine whether kids were at danger or not was to combine these representations into a context vector.

Qin et al. [22] devised an MFO-Attention-LSTM model for forecasting in-class performance based on data from the learning log of the course. At first, Unit learning records were used to extract the fine-grained behavioural traits and pre-processed through normalization and data reshaping to construct the model inputs. Next, an attention-enhanced LSTM network was employed to capture temporal learning patterns while assigning different importance weights to behavioural features. Finally, MFO was applied to optimize the attention layer parameters, resulting in improved prediction accuracy.

Shou et al. [23] suggested a Multidimensional Time-Series Analysis–Based Student Performance Prediction (MTAPSP) model to support early identification of at-risk learners in online education. In the first stage, learning behaviour logs, assessment scores, and demographic data were pre-processed, temporally segmented, encoded, and concatenated to form multidimensional time-series inputs. Then, a multi-layer LSTM with a multi-head self-attention mechanism (MHSA) was employed to extract temporal behavioural features and evaluate the relative importance of the pre-processed data. Finally, ANN integrated these features to predict students' performance in a multiclass setting, achieving superior accuracy and early-prediction capability.

Zhou and Yu [24] presented an Multi-Graph Spatial–Temporal Synchronous Network (MGSTSN) to improve the accuracy of the student performance prediction. The framework consists of the input module which uses spatial-temporal embeddings to further improve hidden feature representations, several stacked MGSTSL layers to learn complex spatial–temporal correlations and the output module which concatenates the representations learnt from the previous layers and outputs the final prediction results.

Adefemi & Mutanga [25] created a model for predicting students' academic achievement using a combination of Convolutional Neural Networks and Long Short-Term Memory (CNN-LSTM). First, the model pre-processes educational data in order to deal with missing values, encode categorical features, normalize feature input, address class imbalance, and select relevant features. LSTM networks were then used to capture the temporal dependence embedded in the sequences of the students' learning behaviours; and CNNs were employed to obtain discriminative feature representations from the sequence of information. Finally, the learned representations were integrated through fully connected layers to produce the final performance predictions. Bin Nuweeji & Alzubi [26] devised an Activation Ensemble Deep Neural Network (AcEnDNN) for accurate student performance prediction. Initially, raw student data were preprocessed through data cleaning, encoding, and imbalance handling, followed by statistical feature extraction to characterize students' academic and demographic patterns. Then, the extracted statistical features were integrated with the original dataset and used as inputs to a deep neural network composed of multiple hidden layers. The last step was to use an activation ensemble mechanism that used swish, sigmoid, ReLU, and tanh activation functions to understand complicated nonlinear correlations in the student data and to accurately forecast the students' performance.

Gu [27] devised a hybrid framework for predicting students' academic performance and supporting individualized academic planning. Initially, raw educational data were pre-processed through feature transformation, missing-value imputation using K-Nearest Neighbour (KNN), and normalization. After that, we utilised an approach to feature selection based on fuzzy logic. Here, we ranked features using Mutual Information (MI) and ANOVA, a fuzzy inference system combined these rankings, and backward elimination was employed to select the most relevant attributes. Next, a combination of CNN, LSTM, and MLP, among others, was used to mimic the students' performance. Lastly, in order to get reliable final results, a stacked MLP meta-model combined the predictions of the base learners.

Junejo et al. [28] suggested an architecture for early multiclass prediction of online student success utilising learning analytics data based on neural networks. Gathering demographic, evaluation, and clickstream data, dealing with missing values, encoding category variables, and scaling features were all part of the pre-processing. These features were then refined by feature engineering module which reduces the data dimensionality. A One-Dimensional CNN (1D-CNN) was trained to extract temporal and behavioural patterns from the engineered features while addressing class imbalance through weighted training. Finally, the model generated multiclass predictions using the SoftMax function that creates probability scores over different categories.

Li [29] developed a Genetic Algorithm–based Graph CNN (GA-GGCNN) for student grade prediction. Initially, student academic and demographic data were pre-processed and transformed into a graph structure representing student–course relationships. Then, stacked graph convolutional layers were applied to extract relational feature representations by aggregating information from neighbouring nodes. Finally, a genetic algorithm was employed to optimize the network parameters through selection, crossover, and mutation, transforming the learned representations into accurate student grade predictions.

Yang et al. [30] introduced a student academic performance prediction model that integrates multi-source behavioural hypergraphs with TabNet to capture higher-order student associations and enhance interpretability. Multi-source campus behaviour data were pre-processed, normalized, and transformed into statistical behavioural features, and informative features were selected using information gain. Hypergraph convolution was used to aggregate neighbourhood information and learn student embedding representations after KNN was used to generate multi-source behavioural hypergraphs. TabNet model was used to adaptively select important features and predict students' academic performance. Table 1 summarizes these models.

3. Comparative Analysis

Table 1 shows the results of this section's comparison of the aforementioned literature, which focuses on the datasets, metrics for performance, merits and demerits.

Table 1: Summary of various DL models for student's academic performance predictions

Ref No.	Techniques Used	Merits	Demerits	Datasets	Performance Metrics
[16]	CNN, BPSO	It effectively reduces redundant features and improves prediction efficiency using feature selection.	Performance depends on BPSO parameter tuning and may increase computational cost.	UCI Machine Learning repository	UCI Maths data: Accuracy = 93.33%. UCI Portuguese data: Accuracy = 96.6%
[17]	Hybrid 2D-CNN	This model effectively averted the issues of overfitting	Data transformation to 2D may introduce artificial structure and limit interpretability.	OULAD dataset	Accuracy = 88%
[18]	MCAG	The model captures both spatial and temporal patterns while highlighting important features.	Model complexity is high, making training and deployment more challenging.	A poll of Nanjing University's sophomore, junior, and senior classes	Accuracy = 94.22%, Precision = 92.85%, Recall = 95.53%, F1-score = 94.89%
[19]	Feedforward Neural Network, Autoencoder	Effectively reduces high-dimensional activity sequences for early performance prediction.	Temporal dependencies may not be fully captured after dimensionality reduction.	Open University Learning Analytics (OULA) dataset	Accuracy = 0.84, Precision = 0.64, Recall = 0.57, F1-score = 0.59
[20]	Temporal Graph Neural Network, MHSA	The model is particularly effective in mining the dynamic relationship between learning behaviour data and accurately predicting at-risk students	Graph construction and training require significant computational resources.	OULA dataset	Accuracy (Pass/ fail) = 83.22%, Accuracy (Pass/ Withdrawn) = 77.06%
[21]	Attn-ANN	It provides interpretability by identifying important lectures and learning activities.	Performance depends heavily on the quality of sequential learning logs.	Moodle and BookRoll (M2B) system data	For PT course (Week 7): F1-score = 0.607, Accuracy = 0.643, AUC= 0.937. For DSP course (Week 7): F1-score = 0.834, Accuracy = 0.895, AUC= 0.928.
[22]	MA-LSTM	MA-LSTM has the ability to detect and prioritise important characteristics automatically, while downplaying the significance of less significant ones.	Optimization process increases training time and model complexity.	Seven undergraduate courses' worth of data on learning behaviours from 2019 and 2020	F1-score (Enthusiasm for Class) = 70.08%, (Classroom participation) = 76.44%, (Knowledge mastery) = 80.16
[23]	MTAPSP, Multilayer LSTM, MHSA, ANN	The model poses strong effective handling of multidimensional time-series data.	It requires careful preprocessing and large datasets for stable performance.	OULAD dataset	Precision= 0.692, Recall = 0.656, F1-score = 0.668, Accuracy = 0.744
[24]	MGSTSN	Captures complex spatial–temporal correlations using multiple graphs.	Model design is complex and less interpretable for educators.	A questionnaire survey of senior students at a university in Nantong, China	MAE = 9.03, MAPE = 11.25, RMSE = 12.27.
[25]	CNN–LSTM	This model minimized the training time thereby improving computational efficiency	Performance may degrade with noisy or incomplete time-series data.	OULAD and Western Ontario University (WOU) Datasets	Accuracy = 98.93%, Precision = 98.93%, Recall = 98.93%, F1-score = 98.93%
[26]	AcEnDNN	Learns complex nonlinear patterns efficiently using multiple activation functions.	Ensemble activations increase computational overhead and tuning complexity.	Datasets include student-mat.csv, student-por.csv, and real-time	MAE = 1.28, MAPE = 2.36, MSE = 4.55, and RMSE = 2.13 with Student mat. csv dataset (best result)
[27]	CNN, LSTM, MLP, MI, ANOVA	Improves robustness by combining multiple deep models and fuzzy logic.	System complexity is high and difficult to deploy in real-time settings.	Questionnaires data from students at universities in Nanjing, China	RMSE = 1.4908, MAPE = 0.0895, R squared = 0.7890, PLCC = 0.8883, SROCC = 0.8900, CCC = 0.8835
[28]	1D-CNN	Efficiently captures temporal patterns from clickstream and behavioural data.	The model's performance may change in different online learning platforms because of their unique characteristics.	OULAD dataset	For 100% course duration: Accuracy = 98%
[29]	GA-GGCNN	This model effectively captured relational information between students and courses	Genetic optimization can be computationally expensive and slow to converge.	Student Performance Kaggle dataset	For 100 epoch and 1000 students: F1-score = 0.96,

					Precision = 0.96, Recall = 0.97
[30]	TabNet	Captures higher-order student associations with improved interpretability.	Hypergraph construction and training add additional computational complexity.	Information on the on-campus learning and living habits of 14,565 students	F1-score = 0.931

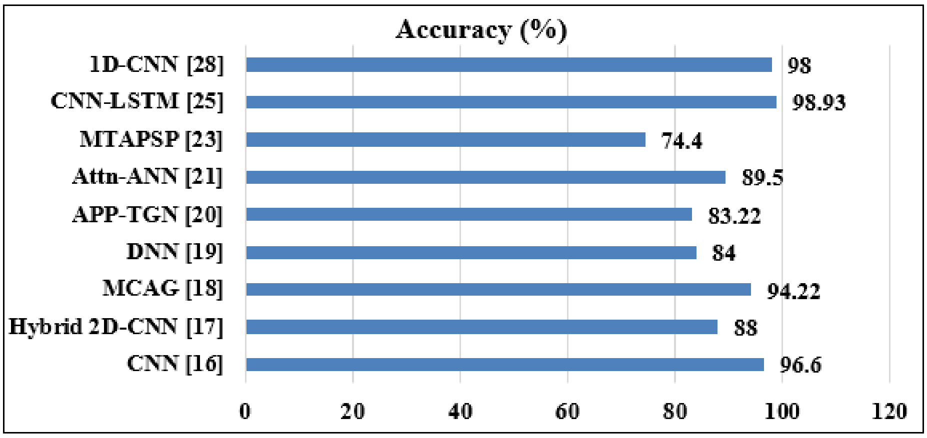

Figure 2: Comparison of various DL models for students' performance prediction on diverse datasets

4. Performance Evaluation

The accuracy reported in Table 1 after the student performance prediction models using the DL framework is summarized, indicating the reliability of the models in learning environments and various educational aspects. The research analyzed for the evaluation has a wide range of data types ranging from institutional LMS data to questionnaire-based surveys to benchmark repositories such as UCI and OULAD to real-time learning behaviour recordings. Here we are comparing the accuracy reported by the various existing approaches based on DL which are used for predicting students' academic performance.

Based on the results shown in Figure 2, CNN-LSTM model was determined to have the highest accuracy of 98.93%, having high potential in terms of temporal and behavioural learning patterns. Likewise, in the case of full course duration, the 1D-CNN model [28] achieved a 98% accuracy on the OULAD dataset, which suggested that using convolution operation to learn the temporal feature is effective for predicting student performance. CNN [16] also obtained excellent accuracy of 96.6%, highlighting the need for selectively reducing the features for further predictive performance enhancement. The above discussions suggest that the DL models that fuse temporal learning patterns and behaviours tend to outperform less complex DL architectures. Often greater accuracy is obtained at a higher cost in terms of computation and information required. The findings from these observations reveal a tension between prediction and efficiency in models used for student performance prediction in DL and the importance of two things: being able to deploy an efficient student performance prediction model in an educational setting while being clear about the models' contents and explainable.

5. Conclusion

Deep learning has emerged as a powerful approach for student performance prediction within Educational Data Mining, offering substantial improvements in modelling complex educational data and learning behaviours. The reviewed studies demonstrate that architectures such as CNNs, LSTMs, attention-based networks, graph neural networks, and hybrid models can achieve strong predictive performance across diverse educational datasets. Despite these advances, challenges related to interpretability, computational requirements, data quality, and model generalizability remain important research concerns. Future studies should focus on explainable and efficient deep learning frameworks, multimodal educational data integration, and real-world deployment in educational environments. These developments can contribute to more effective early intervention strategies and improved student success.

References

[1] L. Calvet Liñán, and Á. A. Juan Pérez, "Educational Data Mining and Learning Analytics: differences, similarities, and time evolution," International Journal of Educational Technology in Higher Education, vol. 12, no. 3, pp. 98-112, 2015.

[2] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI magazine, vol. 17, no. 3, pp. 37-37, 1996.

[3] C. Romero, S. Ventura, and P. D. Bra, "Knowledge discovery with genetic programming for providing feedback to courseware authors," User Modeling and User-Adapted Interaction, vol. 14, no. 5, pp. 425-464, 2004.

[4] R. S. Baker, and K. Yacef, "The state of educational data mining in 2009: A review and future visions," Journal of educational data mining, vol. 1, no. 1, pp. 3-17, 2009.

[5] F. Castro, A. Vellido, A. Nebot, and F. Mugica, "Applying data mining techniques to e-learning

problems," In Evolution of teaching and learning paradigms in intelligent environment, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 183-221, 2007.

[6] Y. Li, K. M. Sheldon, J. N. Rouder, D. A. Bergin, and D. C. Geary, "Long-term prospects and college students' academic performance," Journal of Psychoeducational Assessment, vol. 37, no. 3, pp. 358-371, 2019.

[7] D. R. Sadler, "Grade integrity and the representation of academic achievement," Studies in Higher Education, vol. 34, no. 7, pp. 807-826, 2009.

[8] R. S. J. D. Baker, "Data mining for education," International encyclopedia of education, vol. 7, no. 3, pp. 112-118, 2010.

[9] C. Jalota, and R. Agrawal, "Analysis of educational data mining using classification," In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing, COMITCon, IEEE, pp. 243-247, 2019.

[10] A. Dutt, S. Aghabozrgi, M. A. B. Ismail, and H. Mahroeian, "Clustering algorithms applied in educational data mining," International Journal of Information and Electronics Engineering, vol. 5, no. 2, p. 112, 2015.

[11] E. García, C. Romero, S. Ventura, and C. De Castro, "A collaborative educational association rule mining tool," The Internet and Higher Education, vol. 14, no. 2, pp. 77-88, 2011.

[12] E. Alshehri, H. Alhakami, A. Baz, and T. Alsubait, "A comparison of EDM tools and techniques," International Journal of Advanced Computer Science and Applications, vol. 11, no. 12, pp. 824-831, 2020.

[13] A. Hernández-Blanco, B. Herrera-Flores, D. Tomás, and B. Navarro-Colorado, "A systematic review of deep learning approaches to educational data mining," Complexity, vol. 2019, no. 1, p. 1306039, 2019.

[14] S. Hilbert, S. Coors, E. Kraus, B. Bischl, A. Lindl, M. Frei, and C. Stachl, "Machine learning for the educational sciences," Review of Education, vol. 9, no. 3, p. e3310, 2021.

[15] Z. Xuan, "DRN-LSTM: a deep residual network based on long short-term memory network for students behaviour recognition in education," Journal of Applied Science and Engineering, vol. 26, no. 2, pp. 245-252, 2022.

[16] S. Begum, and S. S. Padmannavar, "Student performance prediction with bpso feature selection and cnn classifier," International Journal of Advanced and Applied Sciences, vol. 9, no. 11, pp. 84-92, 2022.

[17] S. Poudyal, M. J. Mohammadi-Aragh, and J. E. Ball, "Prediction of student academic performance using a hybrid 2D CNN model," Electronics, vol. 11, no. 7, p. 1005, 2022.

[18] Y. Liu, Y. Hui, D. Hou, and X. Liu, "A novel student achievement prediction method based on deep learning and attention mechanism," Ieee Access, vol. 11, pp. 87245-87255, 2023.

[19] X. Wen, and H. Juan, "Early prediction of students' performance using a deep neural network based on online learning activity sequence," Applied Sciences, vol. 13, no. 15, p. 8933, 2023.

[20] Q. Huang, and J. Chen, "Enhancing academic performance prediction with temporal graph networks for massive open online courses," Journal of Big Data, vol. 11, no. 1, p. 52, 2024.

[21] S. Leelaluk, C. Tang, T. Minematsu, Y. Taniguchi, F. Okubo, T. Yamashita, and A. Shimada, "Attention-Based Artificial Neural Network for Student Performance Prediction Based on Learning Activities," IEEE Access, 2024.

[22] X. Qin, C. Wang, Y. Yuan, and R. Qi, "Prediction of in-class performance based on MFO-ATTENTION-LSTM," International Journal of Computational Intelligence Systems, vol. 17, no. 1, p. 13, 2024.

[23] Z. Shou, M. Xie, J. Mo, and H. Zhang, "Predicting student performance in online learning: a multidimensional time-series data analysis approach," Applied Sciences, vol. 14, no. 6, p. 2522, 2024.

[24] Y. Zhou, and X. Yu, "Multi-Graph Spatial-Temporal Synchronous Network for Student Performance Prediction," IEEE Access, 2024.

[25] K. O. Adefemi, and M. B. Mutanga, "A Robust Hybrid CNN–LSTM Model for Predicting Student Academic Performance," Digital, vol. 5, no. 2, p. 16, 2025.

[26] H. Bin Nuweeji, and A. B. Alzubi, "Early Prediction of Student Performance Using an Activation Ensemble Deep Neural Network Model," Applied Sciences, vol. 15, no. 21, p. 11411, 2025.

[27] J. Gu, "Predicting student academic achievement using stacked ensemble learning with deep neural networks and fuzzy-based feature selection," Scientific Reports, vol. 15, no. 1, p. 37195, 2025.

[28] N. U. R. Junejo, M. W. Nawaz, Q. Huang, X. Dong, C. Wang, and G. Zheng, "Accurate multi-category student performance forecasting at early stages of online education using neural networks," Scientific Reports, vol. 15, no. 1, p. 16251, 2025.

[29] T. Li, "Prediction and Optimization of Student Grades Based on Genetic Algorithm and Graph Convolutional Neural Networks," International Journal of Computational Intelligence Systems, vol. 18, no. 1, p. 59, 2025.

[30] Y. Yang, S. Zhao, S. An, X. Li, and Y. Zhang, "Student academic performance prediction via hypergraph and TabNet," Journal of Big Data, vol. 12, no. 1, p. 119, 2025.

Paper ID: SR26612152545 DOI: https://dx.doi.org/10.21275/SR26612152545