International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

EARLY PREDICTION OF STUDENT'S ACADEMIC PERFORMANCE USING AN INTERPRETABLE AND DEEP LEARNING BASED TIME-SERIES PREDICTION FRAMEWORK

Mousoomi Bora[1*], Rupam Baruah[2] and Diganta Baishya[3]

[1*] The Assam Kaziranga University, Assam, India, mousoomi@gmail.com
[2] Jorhat Engineering College, Assam, India, rupam.baruah.jec@gmail.com
[3] Jorhat Engineering College, Assam, India, dbaishya@gmail.com

Abstract

Educational data mining (EDM) is an active research area in present day's higher education system. One of the prevalent objectives of EDM is to predict academic performance of students sin their early academic days. The proposed framework is based on Convolutional Neural Network (CNN) and Bidirectional Long-Short-Term-Memory (LSTM). This study has been projected towards the estimation of student's Cumulative Grade Point Average (CGPA) at the time of graduation based on their previous academic records. A pair of custom activation functions is designed to enhance the performance of the proposed framework. An attention mechanism is implemented to focus on the most dominant features of the prediction process. To interpret the major contributing factors of our study, an interpretable module is used. Looking at the progression of the performance curve, it is observed that student's CGPA can be predicted at their early academic days using the proposed model. Experimental results show that the proposed model with attention mechanism and pair of custom activation functions outperforms the existing models in terms of Mean Square Error (MSE) and Mean Absolute Error (MAE). Based on the evolution of efficient learning pattern, students are recommended pre-emptive counselling towards meeting the attainment gap.

Keywords: Time-series dataset, CNN-BiLSTM, Custom activation function, attention mechanism, interpretability

1. Introduction

Predominant concerns have been accelerated for the quality standard of Higher Education Institutions (HEIs) since ages. To accommodate these uncertainties, India's National Education Policy (NEP 2020) has drawn the vision of Outcome-Based Education (OBE) system. NEP 2020 has launched a self-governing accreditation agency 'National Assessment and Accreditation Council (NAAC)' to assess India's HEIs. To promote OBE, the NEP 2020 has focused on skill development and employability of students pursing their education in HEIs. In order to fulfil the objectives of NEP 2020, India's HEIs have been trying to improve the teaching-learning pedagogy by employing various innovative measures. One of the major milestones of this process is to produce quality graduates from the institutions having strong commands in their respective domains. The purpose of this study is to predict students' final CGPA in their early semesters using their pre-enrollment academic records along with their previous semester grades. The problem statement is formulated as a regression problem and the same is tried to be resolved using a Deep Learning (DL) based time-series forecasting model. The proposed model is deployed on a real-time dataset that contains student's academic records from four cohorts of undergraduate engineering students. Our research is confined to the students who have completed BTech in Computer Science and Engineering (CSE) from a government engineering college from India that includes both regular and lateral entry students.

The students' intake in the regular category is based on the results of Combined Entrance Examination (CEE) and percentage in class twelve board examination. The CEE assesses a student's conceptual

Received: July12,2025

understanding and expertise in three subjects: Chemistry, Math, and Physics (PCM). The marks scored in class twelve board examination especially in four fundamental subjects—Physics, Chemistry, Mathematics, and English are considered as the prerequisites for engineering study.The students who earn their diploma in any engineering discipline are also eligible for admission in BTech program offered by Government engineering colleges. The outcome of this research is aimed to advise both student and teacher community well in advance to take appropriate and productive measures. The teachers can take effective measures for slow learners and proactive steps for the advanced learners.

In addition to the increasing number of youthful and dynamic engineering graduates, this strategy helps the HEIs to achieve a higher ranking from the accrediting body. The suggested approach would serve as a comprehensive road plan for boosting the nation's economy by producing quality graduates from HEIs.

2. Related works

The academic experts and researchers have shown deep interest in holistic development of students. This study reviews the recent research works on students' academic performance and tried to articulate a few interesting ones. Al-Sudani, S. and Palaniappan, R have developed a model based on multi-layered neural network (NN) and the Levenberg-Marquardt learning algorithm to categorize students' educational level [1]. Using the students' browser history, Aydoğdu, S. projected an Artificial Neural Network (ANN) based performance prediction model [2]. An explainable machine learning (EML) framework was created by Guleria, P. and Sood, M. to assist students in deciding their careers after graduation [3]. Iatrellis, O., Savvas, I.K. and Fitsilis, P executed ML approaches for predicting student's outcomes in Higher Education (HE) sector [4]. Costa-Mendes, R., Oliveira, T., Castelli, M. and Cruz Je-sus, F. performed a comparative analysis between multilinear regression, ANN, support vector machine (SVM), RF, and extreme gradient boosting (EGB) as a part of performance prediction [5]. In order to forecast students' total grades, Elmasry, M. created a hybrid model that combines Decision Tree (DT) and Deep Neural Network (DNN) [6]. For the purpose of classifying students' inquiries, Kumar, N., Krishna, H., Shubham, S., and Rout, P. created a text categorization framework. [7]. Envelopment Analysis (DEA) was utilized by Tiancheng, W. to forecast student achievement in online mode [8]. Sekeroglu et al. developed one student's performances prediction model based on Support Vector Regression (SVR), Backpropagation (BP) and Long-Short Term Memory (LSTM) [9]. Çetinkaya, A. and Baykan, O. have used ANN to estimate the analytical, problem solving and programming skills of the high school students [10]. Tadayon, M., Pottie, G. J. established a time-series model using Hidden Markov (HM) model to analyze students' knowledge [11]. Tran et al. proposed a method to predict student performance in next semesters based on the attainment of the previous semesters [12]. Poudyal, S., Mohammadi-Aragh, M., Ball, J. built a hybrid 2D CNN model for academic performance prediction [13]. Jung et al. proposed a load forecasting model using attention based GRU [14]. Aljaloud et al. developed a model based on CNN-LSTM to estimate the learning outcomes of students [15]. Biswas et al. have proposed a set of custom activation functions that outperform existing activation functions [16]. Trottier et al. have mentioned that ReLU activation function suffers from dying cells problem that hinders the performance of the neural network model [17]. The interpretability of the model can be achieved by various tools from explainable AI that help finding the reasons how the black box DL based models provide the prediction [18].

3. Proposed methodology

The proposed framework is designed to predict student's final academic performance based on the performance in semester wise end-term examination in an incremental way. The framework is based

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

on two most popular DL approaches – Convolutional Neural Network (CNN) and Long-Short-Term-Memory (BiLSTM). An interpretable module is implemented to analyse the correlation of different category of subjects with the response variable. The dataset used in this research was collected from the examination department of an anonymous government engineering college from India. Student's pre-enrolment academic records along with their internal and end term examination grades received in a four-year BTech program are considered as independent variables. Likewise, their final "CGPA" is treated as the response variable. **Table 1** shows the summarized information taken from the dataset.

Table 1 Summary of dataset

Type of Column	Description
Sl. No.	Serial Number (of students)
CGPA	Cumulative Grade Point average at the end of 8^{th} semester
BTECH_OVERALL_PERENTAGE	Total marks received by the students in 4 years (in terms of percentage)
1ST_SEM_PERCENTAGE, 2ND_SEM_PERCENTAGE, 3RD_SEM_PERCENTAGE, 4TH_SEM_PERCENTAGE, 5TH_SEM_PERCENTAGE, 6TH_SEM_PERCENTAGE, 7TH_SEM_PERCENTAGE, 8TH_SEM_PERCENTAGE,	End semester percentage starting from 1^{st} to 8^{th} semester
Regular_lateral, Gender, CEE, PCM, ENGLISH, HS_Result	Pre-enrollment records (Regular/ Lateral students, Male/Female, Marks obtained in CEE, % obtained in Physics, Chemistry, Mathematics (PCM) and English in Class twelve board, Student's result in Class twelve board examination)
CATEGORY	Category of student – General, SC, SCT, OBC, NCC.
HOSTELLERS	Whether students opt for Hostel, or they want their own arrangement of accommodation
The internal marks taken from various courses	Subject name is mentioned in Table 2
The end term examination marks taken from various courses	

Table 2 lists the courses offered in each semester as per the course curriculum followed in Computer Science and Engineering domain under BTech.

Table 2 Courses offered in four-year undergraduate program

Semester	Course information (Internal + End term marks in each subject)
1	Physics, Mathematics, Engineering Graphics and Design, Engineering Mechanics, Sociology, Physics-101 Lab, EM Lab Sessional
2	Problem Solving through Programming using C, Basic Electrical Engineering, Communication and Professional Skill, Chemistry, Mathematics-II, Chemistry-201 Lab, Workshop, Basic Electrical Engineering Lab
3	MATHEMATICS-III, Object Oriented Programming, Data Structure, Data Structure and Algorithm, BSS, COI (0 CREDIT), OOP Lab, DSA Lab, INTERNSHIP
4	DISCRETE MATHEMATICS(DM), Computer Organization and Architecture, Operating System, JAVA PROGRAMMING, GRAPH THEORY, ENVIRONMENT SCIENCE (0 CREDIT), Operating System Lab, IT WORKSHOP WITH PYTHON
5	DBMS, Design and Analysis of Algorithm, Formal Language and Automata Theory, PE1, ENGINEERING ECONOMICS, DBMS LAB, Web Programming Lab, Internship-II (SAI –Academia)
6	CD, CN, PE2 (Data Mining), PE3 (GRAPH THEORY), OE1 (software engineering), ACCOUNTANCY, Compiler Design Lab, Computer Networks Lab, MINI PROJECT
7	PE4, OE2, OE3, Principles of Management, PROJECT1, Internship-III(SAI-Industry)
8	PE5, PE6 (NNDL), OE4, OE5, PROJECT-II

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

As number of courses offered in various semesters of a BTech program vary, the dataset under this study is a multivariate and sequential dataset. As listed in Table 2, the courses with acronyms PE and OE stand for Program Electives and Open Electives. The CGPA_GRADE is derived from the CGPA, as demonstrated in Table 3.

Table 3 Conversion of CGPA to CGPA_GRADE

CGPA (in 5.0 scale)	CGPA_GRADE
Greater than 4.0 and less than or equal to 5.0	A
Greater than 3.0 and less than or equal to 4.0	B
Greater than 2.0 and less than or equal to 3.0	C
Greater than 1.0 and less than or equal to 2.0	D
Greater than 0.0 and less than or equal to 1.0	E

This conversion from CGPA to CGPA_GRADE assists the explainability module in the later stage. There are five categorical features used as a part of the input variables. Table 4 shows the five categorical variables and their numerical equivalents.

Table 4 Categorical features and corresponding numerical values

Sl. No.	Categorical features	Numerical Values
1	Regular_lateral	'L': 0, 'R': 1
2	Gender	'F': 0, 'M': 1
3	CATEGORY	'GEN': 0, 'OBC': 1, 'SC': 2, 'STP':3, 'STH':4, 'OS':5, 'FF':6, 'NCC':7
4	HOSTELLERS	'N': 0, 'Y': 1
5	CGPA_GRADE	'A':0, 'B':1, ‘C’:2, ‘D’:3,'E':4

The categorical variables are converted to numerical values though **Label Encoding** as the machine learning models can recognize only numbers. In ‘Regular_lateral’ feature, ‘R’ represents Regular and ‘L’ represents Lateral students. Students from both genders - Male (M), and Female (F) are analyzed in this study. Different categories of students are available in the dataset i.e. General (GEN), Other Backward Class (OBC), Scheduled Tribes Plains (STP), Scheduled Tribes Hills (STH), Scheduled Caste (SC), Other States candidates (OS), Freedom Fighter (FF), and NCC Cadet (NCC). The feature 'HOSTELLERS' has two choices – Hosteller (Y) and Non-hosteller (N). The CGPA_GRADE contains five options - 'A':0, 'B':1, ‘C’:2, ‘D’:3,'E':4.

A model to measure students' progress over time has been devised and depicted in **Fig. 1**. It is a time-series prediction model based on CNN and Bi directional LSTM that learns from long-term dependencies of eight consecutive semesters of BTech program along with the pre-enrolment records of the student. LSTM models are explicitly used for time-series forecasting and it works only for fixed length sequential data. The dataset used in this study is a multi-variate dataset where each record represents as a sequence of data portraying academic journey of a student. Due to the variations in the number of courses in each semester, the length of each sequence varies. The variable length dataset is transformed into a fixed length data input by the padding layer. Any numerical value is used as a mask to the padded values by the masking layer. In this context, Padding and Making layers function as two preprocessing layers for the CNN.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

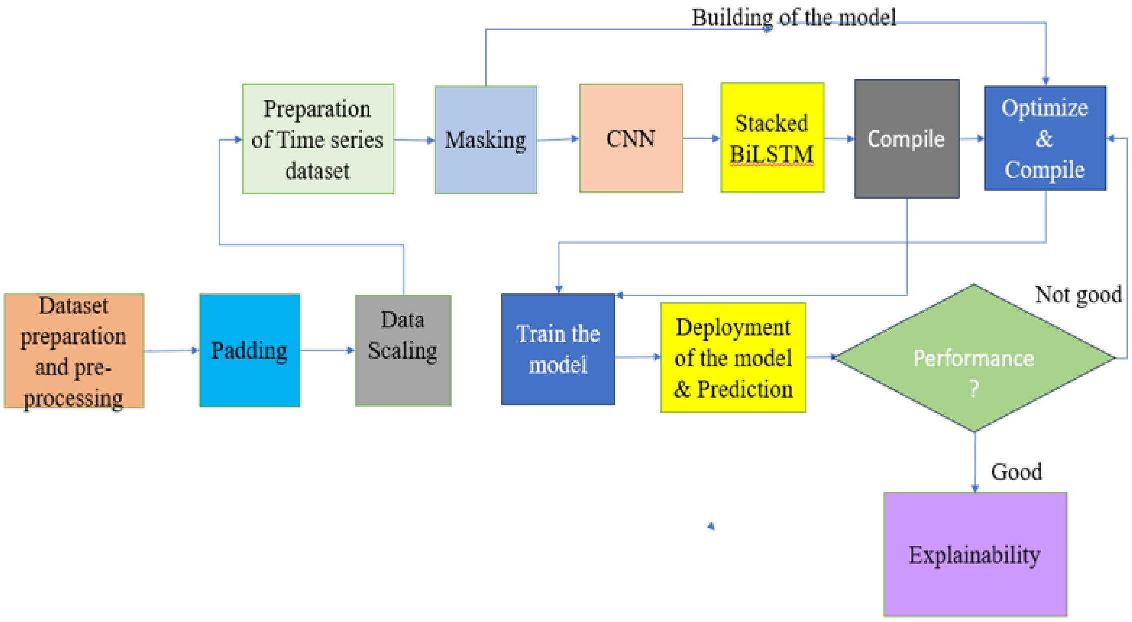

Fig. 1 Interpretable and Deep Learning based time-series prediction framework

The CNN provides exceptional performance in regression analysis by automatic feature selection from the time series dataset. The output generated by the Masking layer is fed to 1D CNN layer which is typically used for time series forecasting by effectively capturing local temporal patterns. The weight of each selected feature by CNN is measured by a custom attention layer that serves as one of the hidden layers of the deep learning model by improving the model's performance. By automatically extracting features from the time series data, 1D CNN directs its output to the next layer of Bi-LSTM for regression or classification. The features suggested by attention mechanism i.e. attention-weighted features are passed through a BiLSTM layer so that only the features having more weight get more value in the prediction process. By incorporating information from both directions, Bi-LSTMs enhance the model's ability by capturing long-term academic records of the students from their pre-enrolment period till the end of BTech 8th semester. Bi - LSTM is followed by a Dense layer that acts as the output layer making the final prediction. The framework is compiled with two types of hyper parameters viz. Loss function and optimizer. The loss functions in neural network compare the predicted value with actual value. The two types of loss functions are used in the proposed framework - Mean Squared Error (MSE) and Mean Absolute Error (MAE). The Adam optimizer is used that adjusts the weights of the neuron in an adaptive way. The model is trained over multiple number of epochs where each epoch comprises of one or more batch of training data. When validation error does not improve over the epochs, 'Early stopping' is used to stop training the model the model training after certain number of epochs. Based on the training and validation loss, the model is tunned using various hyper parameters. For example - Number of hidden layers, number of nodes in each hidden layer, number of epochs etc. To improve the performance of the proposed model, every hidden layer is accompanied by batch normalization, drop out and activation function. In order to increase the efficiency of the proposed model, a pair of custom activation functions is proposed in this study. Activation functions are crucial that introduces nonlinearity in the neural network. However, linear activation functions are preferred for regression **[19]**. Both the custom activation functions are based on Taylore series as represented in **Equation -1** and can be represented as a polynomial of Taylor series **[20].**

$$\textbf{Tailor series } e^x = 1+x+x^2/!2+x^3/!3+\ldots\ldots\ldots \quad [1]$$

The even part of the Tailor series is hyperbolic Cosine function, as represented in **Equation -2**. The shape of cosh(x) is an uneven parabola that expands exponentially.

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$\cosh(x) = 1+x^2/!2+x^4/!4+\ldots\ldots\ldots \; [2]$$

The odd part of the Tailor series is hyperbolic Sin function, as represented in **Equation -3.** The odd part is initially linear and grows exponentially.

$$\sinh(x) = 1+x+x^3/!3+ x^5/!5\ldots\ldots\ldots \; [3]$$

The first custom activation function named as **selu_exp_sinh** is represented by the **Equation -4**.

$$f(x) = (k.\ e^x- a).\ (e^x - e^{-x}/2) \text{ for } x>0 \text{ --------------- } (4)$$

The proposed **selu_exp_sinh** activation function is applied in the CNN layer as shown in **Fig. 2. The selu_exp_sinh** is based on Scaled Exponential Linear Unit (SELU), one of the existing activation functions. SELU is a faster and self-normalizing activation function that helps the neural network to overcome its vanishing gradient problem.

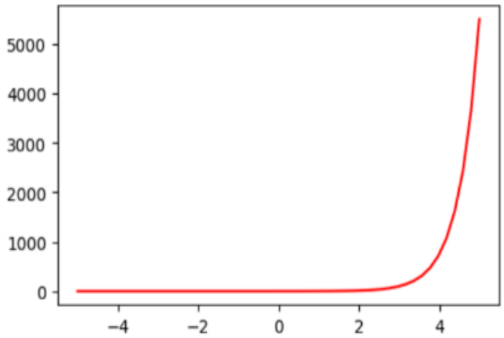

Fig. 2 selu_exp_sinh

Another custom activation function is proposed for this framework named as **selu_exp_sinah.** It is specifically built for the BiLSTM represented by the **Equation -5.**

$$f(x) = k.\ (e^x - e^{-x}/2).\ (e^x + e^{-x})\ /2 \text{ ------------------------}(5)$$

The **selu_exp_sinah** is represented by **Equation -5** is based on Sine hyperbolic function as shown in Fig. 3. The value of k in Equation - 4 and Equation -5 varies from [0.2-0.6].

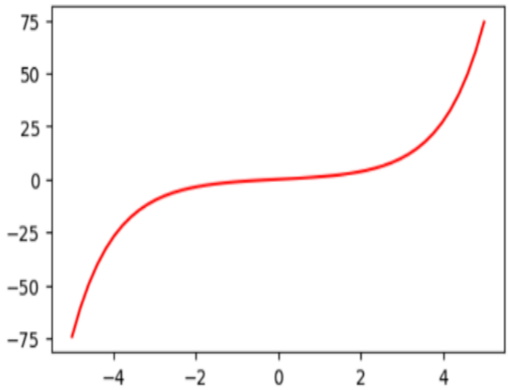

Fig. 3 selu_exp_sinah

The proposed model is integrated with an interpretability module helping the model to identify the most critical features. It enables the model to achieve its expected outcome in the overall prediction. It also explains the contribution of individual feature on the overall prediction process. The effectiveness of the proposed model is discussed in the later section as Results and Discussion.

4. Results and discussions

The performance of the framework developed in this study is compared with two other state -of-the art DL models: CNN-LSTM and CNN-GRU. The framework is implemented on four series of data in a cumulative manner as listed below.

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Series 1: Pre-enrollment and first year examination records.
Series 2: Pre-enrollment, first year and second year examination records.
Series 3: Pre-enrollment, first year, second year and third year examination records.
Series 4: Pre-enrollment, first year, second year, third year and fourth year examination records.
For easier analysis, the results of the proposed framework are elaborated in two steps - without optimization and with optimization.

4.1 Performance evaluation without optimization

The performance of the conventional framework that works on LeakyReLU activation function and ADAM optimizer is assessed using two loss functions - MAE and MSE as shown in Table 5.

Table 5 Performance of the CNN-BiLSTM framework without any optimization

MSE (CGPA prediction)			MAE (CGPA prediction)		
Models	Train	Test or Validation	Models	Train	Test or Validation
CNN-BiLSTM +ADAM+ LeakyReLU			CNN-BiLSTM+ ADAM+ LeakyReLU		
Series 1	2.9117	3.932	Series 1	1.36	0.9597
Series 2	2.797	1.626	Series 2	1.358	0.791
Series 3	2.2658	1.372	Series 3	1.228	1.077
Series 4	**1.1082**	**1.113**	Series 4	**0.915**	**0.5112**

Table 5 displays how all four models with four different datasets i.e. Series 1, Series 2, Series 3, and Series 4 can be used in the early prediction of CGPA in an incremental approach. It can be observed that the losses that occur in the framework with Series 1 i.e. after 1st year, are the highest. However, the loss is gradually reduced with the addition of more information in Series 2, Series 3 and Series 4. The framework has shown the best performance at the end of the fourth year (with Series 4) with the least amount of performance loss.

4.2 Performance evaluation with optimization

The prediction performance of the optimized framework is evaluated after implementing different strategies like attention mechanism, custom activation functions - selu_exp_sinh and selu_exp_sinah, batch normalization, drop-out percentage etc. The following sub-sections discuss about optimization steps used in the framework.

.

4.2.1 Attention mechanism

Table 6 demonstrates the difference in the model's performance without and with the attention module.

Table 6 Performance comparison of the proposed framework with respect to attention mechanism (CGPA prediction)

Dataset	CNN- BiLSTM (Without attention) +Adam+ LeakyReLU				CNN- BiLSTM (With attention) +Adam+ LeakyReLU			
	MSE Train	MSE Test or validation	MAE Train	MAE Test or validation	MSE Train	MSE Test or validation	MAE Train	MAE Test or validation
Series 1	2.9117	3.932	1.36	0.9597	0.6123	0.0445	0.6431	0.1491
Series 2	2.797	1.626	1.358	0.791	0.6276	0.3709	0.655	0.609
Series 3	2.2658	1.372	1.228	1.077	0.5946	0.3954	0.6194	0.6188
Series 4	**1.1082**	**1.113**	**0.915**	**0.5112**	**0.5942**	**0.3611**	**0.6264**	**0.6009**

Table 6 has shown that the framework provides the best performance at the end of 4th year in terms of MSE and MAE. From all four series data, it can be confirmed that the prediction performance of the framework is enhanced with the addition of the "Attention" module.

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4.2.2 Custom activation function

A pair of custom activation function (Fig 2 and Fig 3) is integrated with the framework to enhance its prediction performance. The performance of the models with several optimization steps such as attention mechanism, custom activation functions i.e. selu_exp_sinh and selu_exp_sinah, batch normalization, and drop-out is displayed in Table 7.

Table 7 Performance of the proposed framework with attention mechanism and custom activation function

MSE (CGPA prediction)			MAE (CGPA prediction)		
Models	Train	Test or Validation	Models	Train	Test or Validation
CNN-BiLSTM+Attention+Custom Activation+Adam			**CNN-BiLSTM+Attention+Custom Activation**		
Series 1	**0.5838**	**0.0466**	Series 1	**0.621**	**0.1601**
Series 2	0.6039	0.1969	Series 2	0.6306	0.4437
Series 3	0.6012	0.6003	Series 3	0.6412	0.5748
Series 4	0.6234	0.9139	Series 4	0.6491	0.9561
CNN-BiLSTM+ Attention+ReLU+Adam			**CNN-BiLSTM+Attention+ReLU Activation**		
Series 1	0.6036	0.0458	Series 1	0.6414	0.159
Series 2	**0.5992**	**0.3531**	Series 2	**0.6319**	**0.5942**
Series 3	0.6326	0.3589	Series 3	0.6636	0.5891
Series 4	0.6276	0.3728	Series 4	0.6537	0.5934
CNN-BiLSTM+ Attention+LeakyReLU+Adam			**CNN-BiLSTM+ Attention+LeakyReLU+Adam**		
Series 1	0.6123	0.4445	Series 1	0.6431	0.1491
Series 2	0.6076	0.3709	Series 2	0.655	0.609
Series 3	**0.5926**	**0.3954**	Series 3	**0.6194**	**0.6288**
Series 4	0.5932	0.3611	Series 4	0.6264	0.6009
CNN-BiLSTM+ Attention+GELU+Adam			**CNN-BiLSTM+ Attention+GELU+Adam**		
Series 1	0.6416	0.0446	Series 1	0.6681	0.149
Series 2	0.6126	0.3651	Series 2	0.6422	0.6042
Series 3	0.6094	3.32442E+11	Series 3	**0.634**	**0.2031**
Series 4	**0.6082**	**0.3738**	Series 4	0.6374	0.6114
CNN-BiLSTM+ Attention+SELU+Adam			**CNN-BiLSTM+ Attention+SELU+Adam**		
Series 1	0.6426	0.0446	Series 1	0.6993	0.1496
Series 2	0.6357	0.573	Series 2	0.7027	0.757
Series 3	0.6248	0.7523	Series 3	0.6835	0.8663
Series 4	**0.5989**	**0.4053**	Series 4	**0.6467**	**25.6756**

The comparative analysis of the model's performance using custom activation functions with four existing activation functions- ReLU, LeakyReLU, GELU, and SELU are deliberated in Table 7. The framework with ReLU gives better performance after the second-year examination. The same framework with LeakyReLU shows better performance after the third-year examination. While used with GELU and SELU, framework shows better performance after the fourth-year examination. On the other hand, in terms of MSE and MAE, CNN-BiLSTM with two custom activation functions and attention mechanism gives better performance after first year examination.

4.3 Interpretation of the contributing factors in model's prediction

The DL-based models are hard to interpret due to their complex architecture. The interpretability module helps the proposed framework to assess different categories of courses offered in the four years of the BTech program. The model's Interpretability helps to explain the reasons behind the decisions taken while choosing the primary contributing features in the prediction process. To understand the feature's influence in the predicted outcome, two widely used explainable AI tools are

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

used in the current study - SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). SHAP is one of the most commonly used summary-based and instance-based explainability techniques that leverages game theory concept to measure the impact of each feature on the prediction. LIME is a model-agnostic explainability technique.

4.3.1 Local interpretation by LIME for a specific instance after 1st year

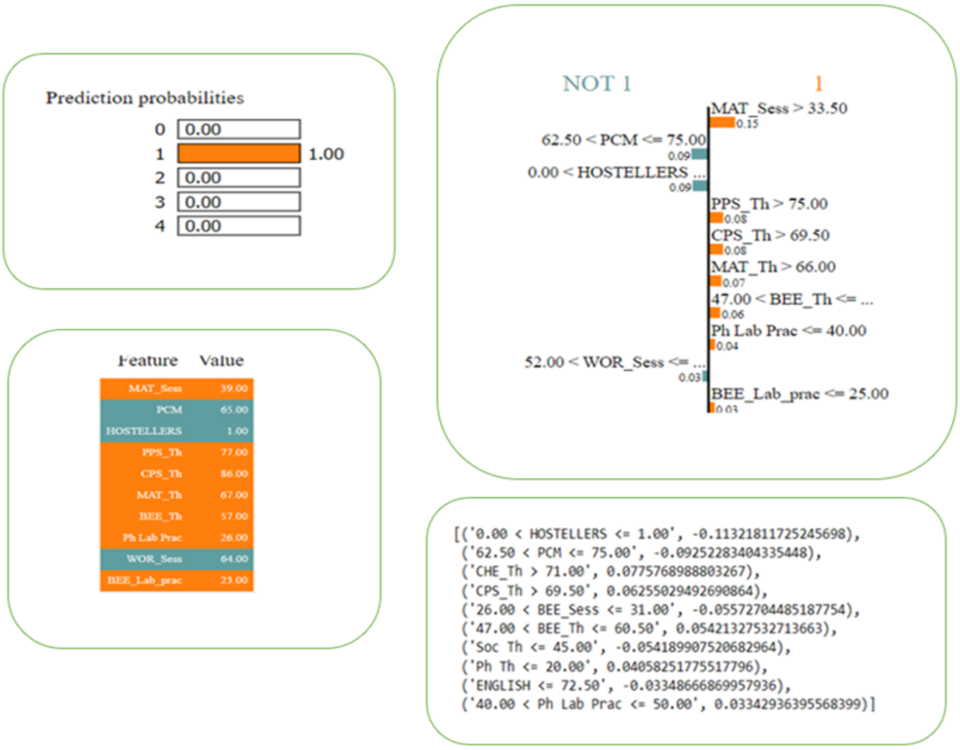

Fig. 4 Contribution of important features in pre-enrolment academic days and first year of BTech, Target – CGPA_GRADE as 1

From Fig. 4, it can be interpreted that being a student from Computer Science and Engineering background, Mathematics sessional and Programming and problem-solving theory (PPS) have higher weightage for better CGPA as shown in the given figure (Fig. 4). The contribution of individual subject in the prediction process is as follows –

Mathematics sessional:15%, Programming and problem-solving (PPS) Theory: 8%, Communication and Professional Skill (CPS) Theory: 8%, Mathematics Theory: 7%, Basic Electrical Engineering (BEE) Theory 6%, Physics lab practical:4%, BEE lab practical:4%. The subjects having no contribution in this prediction (green background): PCM marks: 9%, Hostel accommodation: 9%.

4.3.2 Global interpretation by SHAP bar plot (After 4th year)

Received: July12,2025

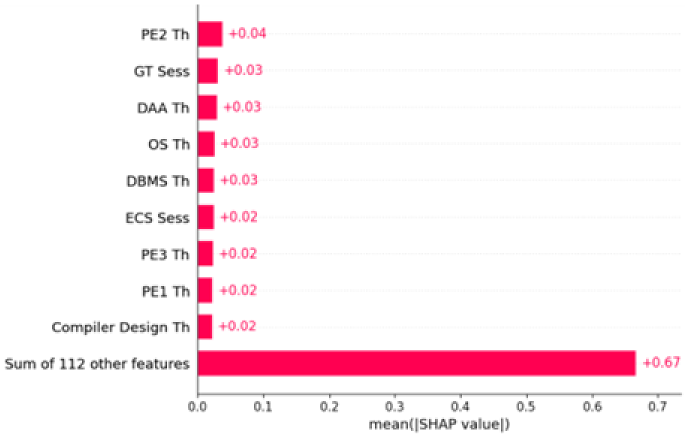

Fig. 5 Global interpretation by SHAP bar plot

The feature importance of each feature taken after 4th year are as flows –

PE2 Theory marks:4%, Graph Theory (GT) Sessional, DAA Theory, Operating System (OS) Theory, Data Base Management System (DBMS) marks:3%, ECS Sessional, PE3 Theory, PE1 Theory, Compile Design Theory marks:2% and so on.

5. Conclusion

Student's performance evaluation at the time of graduation is a continuous evaluation process. The process requires systematic and continuous monitoring of academic performance in a planned manner. The proposed DL based time-series prediction framework acts as an alarming as well as counselling system for students indicating their probable grade towards the end of the final semester. Although it is implemented on student's academic records taken from only Computer Science and Engineering background nonetheless, such approach may be employed in performance prediction of students from diverse academic backgrounds such as basic sciences, management, social sciences and so on. As a future scope, more emphasis would be given to improve the performance of the proposed model using different trending deep learning techniques like Cognitive, Genetic algorithms and highly predictive algorithms.

Disclosure statement

The authors report there is no conflict of interests to declare.

References

1. Al-Sudani, S., Palaniappan, R.: Predicting students' final degree classification using an extended profile, Education and Information Technologies, vol. (24), 1-13, (2019).
2. Aydoğdu, Ş.: Predicting student final performance using artificial neural networks in online learning environments, Education and Information Technologies, vol (25), 1-15, (2020).
3. Guleria, P., Sood, M.: Explainable AI and machine learning: performance evaluation and explainability of classifiers on educational data mining inspired career counseling, Educ Inf Technol, vol. (28), 1081–1116. (2023).
4. Iatrellis, O., Savvas, I.K., Fitsilis, P.: A two-phase machine learning approach for predicting student outcomes, Educ Inf Technol, vol. (26), 69–88 (2021).

5. Costa-Mendes, R., Oliveira, T., Castelli, M., Cruz Je-sus, F.: A machine learning approximation of the 2015 Portuguese high school student grades: A hybrid approach, Education and Information Technologies, vol. (26), 1-21, (2021).
6. Elmasry, M.: Applying Deep Learning Neural Networks in Predicting Students' Cumulative Grades, International Journal of Engineering and Advanced Technology, vol. (10), 293-297, (2020).
7. Kumar, N., Krishna, H., Shubham. S., Rout, P.P.: Automated categorization of student's query, International Journal of Computers and Applications, vol. (44), 1056-1064, (2022).
8. Tianchang, W.: Online student achievement prediction model based on Data Envelopment Analysis (DEA) for multidimensional educational internet of data mining, International Journal of Computers and Applications, vol. (43), 315-319, (2021).
9. Sekeroglu, B., Dimililer, K., Tuncal, K.:Student Performance Prediction and Classification Using Machine Learning Algorithms, ICEIT 2019: Proceedings of the 2019 8th International Conference on Education-al and Information Technology, pp. 7-11, (2019).
10. Çetinkaya, A., Baykan, O.: Prediction of middle school students' programming talent using artificial neural networks, Engineering Science and Technology, an International Journal, vol. (23), (2020).
11. Tadayon, M., Pottie, G. J.: Predicting Student Performance in an Educational Game Using a Hidden Markov Model, IEEE Transactions on Education, vol. (63), pp. 299-304, (2020).
12. Dien, T. T., Nguyen, T.-N., Nguyen, H., Luu, S.: Deep Learning with Data Transformation and Factor Analysis for Student Performance Prediction, International Journal of Advanced Computer Science and Applications, vol, (11), 711-721, (2020).
13. Poudyal, S., Mohammadi-Aragh, M., Ball, J.: Pre-diction of Student Academic Performance Using a Hybrid 2D CNN Model, Elec-tronics, vol. (11), (2022).
14. Jung, S., Moon, J., Park, S., Hwang, E.: An Attention-Based Multilayer GRU Model for Multistep-Ahead Short-Term Load Forecasting, Sensors, vol. (21), (2021).
15. Aljaloud, A., Uliyan, D., Alkhalil, A., Elrhman, M., Alogali, A., Altameemi, Y., Altamimi, M., Kwan, P.: A deep learning model to predict students' learning outcomes in LMS using CNN and LSTM, IEEE Access, vol. (10), (2022).
16. Biswas, K., Kumar, S., Banerjee, S., Pandey, A.: TanhSoft -- a family of activation functions combining Tanh and Softplus.
17. Trottier, L., Gigu, P., Chaib-draa, B.: Parametric exponential linear unit for deep convolutional neural networks, in: Machine Learning and Applications (ICMLA), 16th IEEE International Conference on, IEEE, pp. 207–214, (2017).
18. https://www.datacamp.com/tutorial/introduction-to-shap-values-machine-learning-interpretability, last accessed on 2025/01/25.
19. https://encord.com/blog/activation-functions-neural-networks/, last accessed on 2025/01/ 25.
20. https://betterexplained.com/articles/hyperbolic-functions/, last accessed on 2025/01/25, last accessed on 2025/01/25.