

Bidirectional Long Short-Term Memory Integrated Text-to-Text Transfer Transformer to Identify Predictive Variables for Students' Academic Performance Analysis

Kandula Neha ^{1,*}, Anil Sharma ², Ram Kumar ³

¹*Research Scholar, Department of Computer Science, Lovely Professional University, Punjab, India.
nehakandula09@gmail.com*

²*Professor, School of Computer Applications, Lovely Professional University, Punjab, India.
anil.19656@lpu.co.in*

³*Assistant Professor, Department of Computer Science, VIT Bhopal University, Madhya Pradesh, India.
drramkumar.research@gmail.com*

Abstract Educational institutions and professionals must analyze student performance to identify ways to improve each student's academic achievement. Predicting how well students will do in universities and colleges is crucial for managing their time effectively. Predicting students' academic achievement has been the subject of some relevant research activity since the advent of deep learning. Students' academic performance is affected by various factors, yet there is a general trend of similar intelligence and learning styles among students at the same university. The goal of this research is to use deep learning algorithm and historical data to forecast how well students will do in academics. Academic performance analysis can be enhanced by combining a Text-to-Text Transfer Transformer (T5) with a Bidirectional Long Short-Term Memory (Bi-LSTM) model. This allows us to leverage the capabilities of both architectures. In contrast to Bi-LSTM's adeptness at capturing sequential dependencies, T5's text-to-text method offers a versatile framework for encoding and processing a wide range of academic data. Accuracy in activities such as requirements classification and short response grading can be enhanced using this hybrid technique. The proposed T5-BiLSTM model is used to accurately identify the predictive variables from academic datasets that majorly influence the academic performance. This research is helpful for educational institutions to identify the predictive variables that are used to enhance the student academic progress.

Keywords Educational Institutions, Student Performance, Student Progress, Deep Learning Algorithm, Text-To-Text Transfer Transformer, Bidirectional Long Short-Term Memory, Predictive Variables, Student Academic Progress.

DOI: 10.19139/soic-2310-5070-3456

1. Introduction

While the teaching delivery modality has been hopping between online and offline for the past two years, student performance analysis has taken a leap of trust [1]. There are a lot of new aspects that need to be studied and recognized that are having a big impact on students' performance. In addition to the many academic variables that must be considered and studied, socioeconomic factors must also be considered and studied. Academics are among the many fields that have benefited from predictive analytics' capacity to accurately foretell future outcomes [2]. Detailed information on the school and its students, including academic achievements, is stored in educational institutions' treasure troves of data. These organizations cannot afford to lose this data treasure trove, as it has

*Correspondence to: Kandula Neha (Email:nehakandula09@gmail.com). Department of Computer Science, Lovely Professional University, Punjab, India.

the key to discovering valuable insights [3]. In order to support students' academic achievement, it is crucial to have a high-quality prediction model that can correctly analyze this data. To improve education, the abundant data must be used on student performance [4]. The responsibility of educational institutions to track and assess student progress is fundamental to these goals. This includes managing advising services, effectively mentoring students, and forecasting student results. In order to measure the success and influence of the learning process, it is crucial to provide useful feedback to teachers and students [5]. These findings warrant careful reevaluation of the current educational system.

Student categorization for the goals of performance prediction or the identification of possible dropouts is among the most critical jobs in educational statistics [6]. Accurate predictions of students' future accomplishments and success are crucial for educational institutions. Those who work in education place a high value on student accomplishments since they are used as a benchmark to evaluate education systems worldwide [7]. It is crucial to identify at-risk students as soon as possible because dropping out of colleges also reduces this efficiency. This highlights the critical need of properly classifying educational data. With the current technological developments, educational records can be efficiently and effectively categorized using a wide range of methods [8]. A wider variety of approaches well-suited to this kind of work has to be explored. Academic institutions are collecting more and more data, and the range of subjects they cover is expanding rapidly as well. It is an intriguing task to investigate the practicality of unusual approaches [9].

When it comes to predicting performance, deep learning models are generally on par with more sophisticated architectures in machine learning with fewer parameters and less training time required. Because of this, they are ideal for educational datasets of medium size that have limited computational resources [10]. The capacity of hybrid deep learning models to combine supplementary architectures has led to a rise in interest in these models, which play to their strengths. One promising approach to extracting spatial and sequential patterns from complicated multidimensional datasets is to combine multiple deep learning models for temporal modeling. Because student data frequently displays organized linkages and time-dependent progressions, hybrid techniques like these are especially pertinent in educational settings [11]. Timely interventions and the provision of individualized support are crucial for enhancing student outcomes, and they are made possible by accurate academic performance prediction. While AI-driven models have great promise, there are still a number of issues that prevent them from being widely used in academic environments [12]. The general process of student performance analysis using predictive variables is shown in Figure 1.

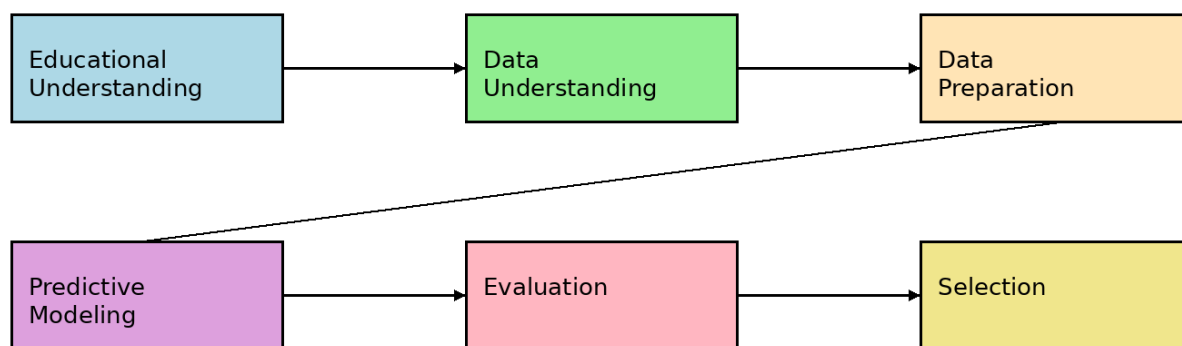


Figure 1. General Process of Student Performance Analysis

The capacity of traditional ML models to generalize to new data can be hindered by their reliance on human feature engineering. Deep neural networks, on the other hand, can learn complicated patterns from simple data with their automatic feature extraction. But for these models to work reliably and impartially, a big dataset is required with many of different data points. Particularly for analyses pertaining to individual courses or at smaller institutions, such datasets are unfortunately not always easily accessible in educational research [13]. Missing

values and class imbalance are two more common issues in educational datasets that can weaken learning models' predictive power and resilience. This research offers a new hybrid deep learning model that effectively predicts students' grades by integrating deep learning models.

Students Academic performance prediction has become a pressing issue in educational data mining, allowing intervention early and individual learning approaches. Current methods are mainly based on the old-fashioned machine learning schemes or independent deep learning schemes, and they tend to be unable to capture the complexity of interactions among heterogeneous features, including academic history, behavioral evidence, and socio-economic status. Recent developments in deep learning have brought models like LSTM networks to do temporal modeling and transformer-based models to contextual understanding. Nevertheless, the majority of the current research considers these methods separately. Sequential models such as LSTM are effective in capturing temporal dependencies but they do not capture semantic relationships involving different types of features. On the other hand, models based on transformers are much better at contextual representation, but do not necessarily aim to preserve temporal transitions in structured educational data. Moreover, most of the literature is based on a tabular feature engineering or a snapshot of student data, rather than the dynamic change in student performance over semesters. This constraint limits the accuracy of prediction models to practical education in real-world scenarios where student behaviour and performance change with time. To overcome the difficulties, this paper suggests a hybrid T5 BiLSTM architecture, which combines contextual semantic learning and temporal dependency learning. The model takes advantage of the advantages of transformer-based encoding performance by converting structured student data into a textual form without losing the sequential patterns with the help of BiLSTM. This method facilitates a more detailed view of the dynamics of student performance and enhances the predictive performance as well as interpretation.

In response to the drawbacks of more conventional recurrent neural networks, one variant that has been developed is the Bidirectional Long Short-Term Memory (BiLSTM) network. Through simultaneous forward and backward monitoring of information flow from past and present timesteps, it is intended to efficiently handle the vanishing-gradient problem and capture long-term interdependence in text sequences. With its LSTM layer, the input sequence can be processed in both ways at the same time. The LSTM layer records data from the past and the future during its forward and backward passes, respectively. The model can successfully capture the input sequence's long-term dependencies because to this bidirectional processing. Based on the task at hand, the Bi-LSTM layer's output can serve multiple functions. A fully connected layer and a softmax activation can be used to get class probabilities in text classification, for instance. Named entity identification and other sequence labeling tasks allow for direct use of output to forecast incoming token labels. The BiLSTM Architecture is shown in Figure 2.

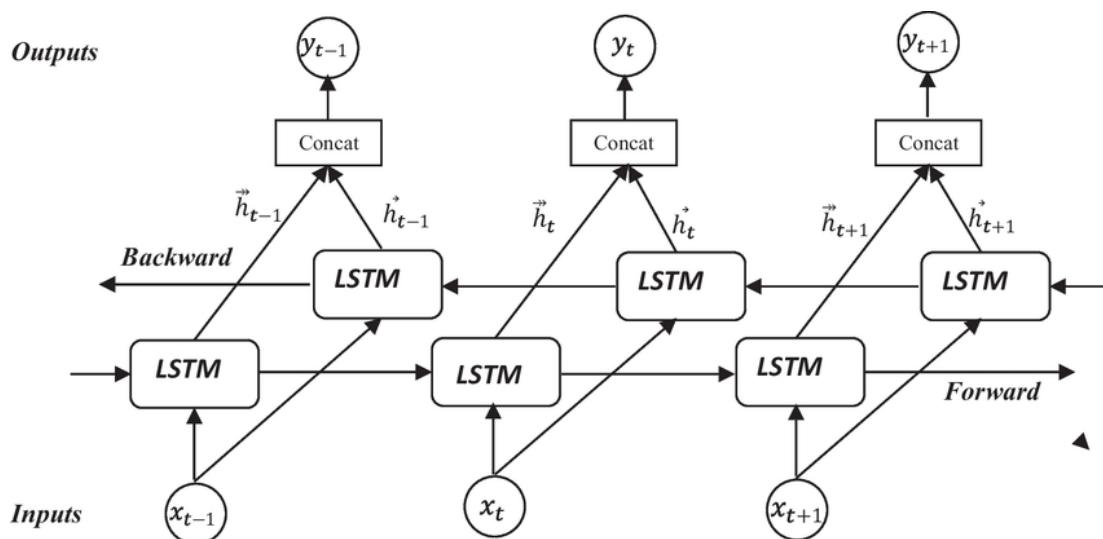


Figure 2. BiLSTM Architecture

The Text-to-Text Transfer Transformer (T5) offers a standardized method for handling various Natural Language Processing (NLP) activities by transforming them into text. Since T5 is designed to handle activities with plain text input and output, it avoids handling jobs with different structures and output types, such as summarization, translation, and classification. The training process is made easier, and the model is able to generalize better across tasks thanks to this uniform formulation. Tokenization is the first step in processing the input; it involves utilizing the Sentence Piece tokenizer to break the raw input text into subword units. The encoder, an integral component of the conventional Transformer design, receives these tokenized inputs. After running the sequence through its paces, the encoder generates representations that are contextualized to capture the input words' meaning and relationships. The T5 transformer architecture is shown in Figure 3.

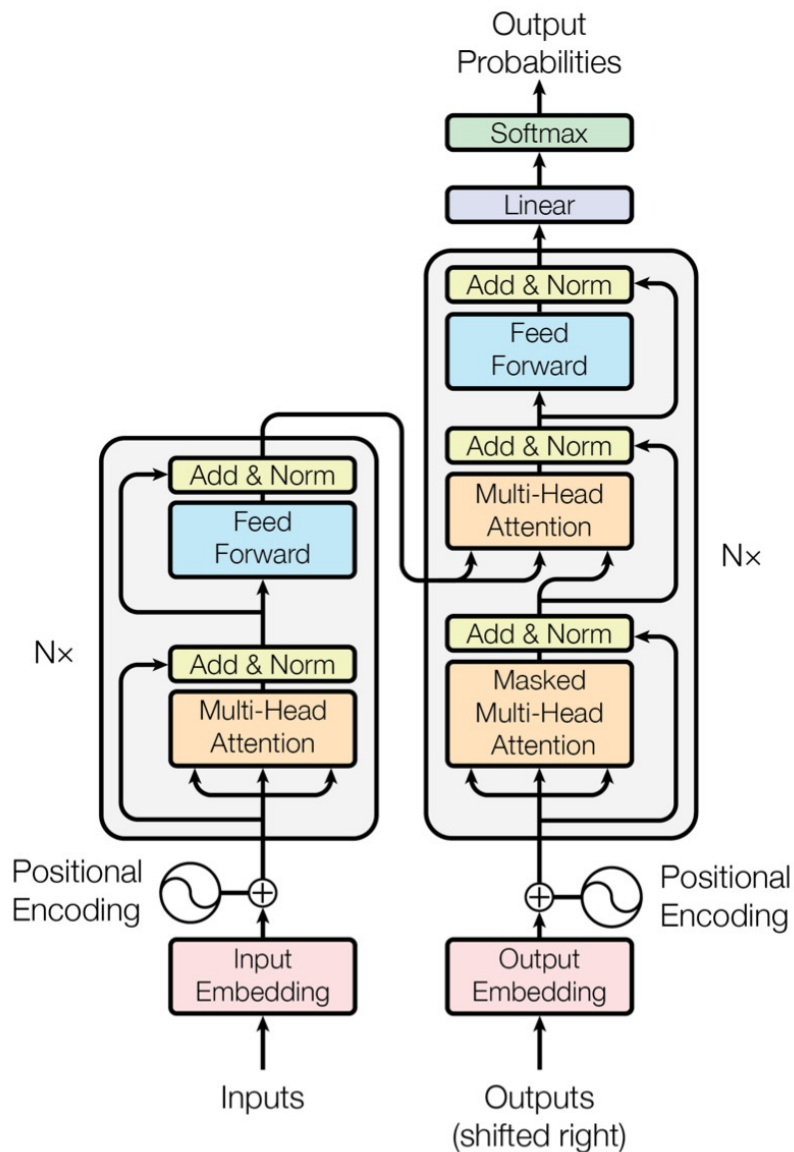


Figure 3. T5 Transformer Architecture

After encoding is finished, the decoder is able to take control. It employs an autoregressive method, which involves generating the target text word by word while conditioning on the words that have come before, using the encoded information. The output text is completed when this step is repeated until a specific end-of-sequence token is produced. The initial step in training T5 is to use a denoising autoencoder objective on a large text corpus. This objective involves masking and then reconstructing random spans of input text. This method aids the model in acquiring representations of rich language. By supplying task-specific input-output pairings, T5 is fine-tuned on certain downstream tasks following pre-training. The model is able to apply its pre-training general language knowledge to a variety of natural language processing (NLP) problems with little to no further training thanks to this two-stage training approach. In this research, Text-to-Text Transfer Transformer (T5) with a Bidirectional Long Short-Term Memory (Bi-LSTM) model is used to enhance the academic performance analysis by considering the most accurate predictive variables.

2. Literature Survey

Class discussion, where students build their knowledge and connect socially, is crucial to students' academic success, according to previous research. While some have investigated the feasibility of using big data gathered from online classrooms, the majority of these studies relied on qualitative or small-scale data. Y. Zhen et al. [1] used natural language processing techniques to examine conversations taking place in real-time classrooms on a massive Chinese online learning platform in an effort to resolve this issue. Classroom conversations are analyzed for emotional expression and interaction type elements. The author used interpretable artificial intelligence methods to identify the most relevant predictors in the prediction models, and then used these features to build neural network models that can predict which students will do well academically and which will not. Across all learning phases, high-performing students in STEM (science, technology, engineering, and mathematics) and non-STEM classes consistently display greater positive emotion, cognition, and off-topic discussions compared to low-performing students. The metacognitive dialogue shows how important it is in classes that aren't STEM, but it doesn't work in STEM classes. Students who excel in STEM subjects display good emotions at the last stage of lessons, in contrast to high-achieving students in non-STEM subjects who display negative emotions.

Like any other community, student communities in the classroom are only as solid as its individual members. Students' overall learning outcomes are proportional to how well the least performing student in a given class does. Based on this logic, it is bad for the quality of education when many students fail a class and have to take it again. In order to improve the course and identify the resources needed to accomplish this goal, A. Kala et al. [2] designed a model to predict students' success early on in their curriculum to identify those at danger of failing. The author achieved this goal by presenting a theoretical framework that combines Particle Swarm Optimization (PSO) with Deep Neural Networks (DNN). Models trained using deep learning and classical machine learning were compared to assess the model's classification performance. The author utilized the SHAP and LIME approaches to find out how each attribute contributed to the model's predictions. The accuracy findings of the model were compared with early prediction studies, and the widely used xAPI-Edu-Data dataset, which includes numerous courses, was utilized to assess the model's generalizability and applicability.

Constantly monitoring student activities requires a lot of physical labor. The use of cameras and the ability to detect unusual surveillance behavior automatically have both experienced tremendous growth in recent years. Operations in the actual world have been observed and monitored using various computer vision techniques. Online classes are already being offered by most universities in an effort to mitigate the effects of this epidemic on the education sector. For these online classes to be as engaging as in a traditional classroom, it is essential that students be properly identified and that their actions are closely watched. In this research, N. M. Alruwais et al. [3] provided novel algorithms based on deep learning that continuously monitor a student's emotional state, including anger, disdain, joy, sadness, fear, and surprise. Researchers also looked at how well a CNN model with deep learning and 99% accuracy could identify students and track their activities in virtual classrooms. This method outperformed the competition with its batch normalization, dropout regularization, and several convolutional layers. It reduced overfitting and captured important features. According to the research, deep learning algorithms have the potential

to improve student engagement and learning results in online learning environments by identifying them more frequently. Educators and instructors can better understand their students' behaviors and provide customized, individualized help using these strategies. This can improve academic performance and activity evaluation.

An AI model for educational speech emotion recognition (SER) that examines the relationship between students' emotional states, team discussions, and academic achievement. Academic performance is positively correlated with good emotions, according to research. On the flip side, when you're feeling down, it will show in your grades. In order to foster a nurturing learning environment and enhance students' motivation and engagement, it is crucial to consider their emotional states. Educators can meet their pupils' emotional needs and intervene when they arise with the help of tools made possible by this area of study. The purpose of this research is to apply topic extraction to student conversations while they work on group projects in class to determine if the students' topics of discussion are relevant to the course material and, if so, what those topics are that is analyzed by N. Dehbozorgi et al. [4]. In addition, a thorough investigation is carried out to determine the relationship between the students' emotional responses to the course material and their overall grade achievement. The student's performance is assessed in a formative manner by considering their results across several areas. A speech transcriber, an emotion analyzer, and a subject extraction module make up the developed model's core. In order to find the correlations, the results from each of these modules are processed.

An essential component of intelligent education vital for encouraging long-term student development is the ability to forecast how well students will do on assessments. Predicting academic success is really about finding all the hidden connections in performance data. Although this problem has been studied extensively, current methods don't always account for all possible angles or represent the complex interdependencies in space and time at the same time. In order to address these shortcomings, Y. Zhou et al. [5] presented a multi-graph spatial-temporal synchronous network (MGSTSN) to improve the accuracy of performance forecasts. This method allows for a comprehensive depiction of the various spatial-temporal dynamics by creatively creating spatial trend and pattern graphs in addition to temporal causal graphs. Another new approach to studying how students' performance is influenced by both geographic and temporal aspects is the construction of dual spatial-temporal synchronous graphs and their associated synchronous modules. By exhibiting an improvement of 8.17% to 11.34% across all measures, MGSTSN substantially surpasses current models in rigorous evaluations conducted on real datasets.

With the advancements in communication and the Internet, online learning has picked up speed. Digital environments enable students to participate quickly and from anywhere in the widely accessible free online courses. In recent years, VLEs have grown rapidly, providing students with access to top-notch digital materials. There are many advantages to online learning settings, but there are also some disadvantages, such as low engagement, high dropout rates, self-regulation, and the need for students to set goals for themselves. Organizations and teachers can make data-driven decisions and enhance educational practices by forecasting which students will fail in a virtual learning environment (VLE). To forecast students' performance, J. A. I. S. Masood et al. [6] suggested a Hybrid Deep Learning (HDL) method based on Resnet model-based ECNN (Enhanced Convolution Neural Networks) classification algorithms. The Open University Learning Analytics Dataset (OULAD) is used to evaluate the HDL method, which gives a thorough and trustworthy evaluation of the model's performance. In comparison to the preexisting classifiers, the hybrid DLT techniques showed higher prediction accuracy, proving their superiority.

Engaging students in a meaningful way is still a critical challenge in today's education, especially with the influence of technology. Traditional assessment approaches overlook the complex nature of student engagement, despite the fact that it has a substantial impact on academic performance. To categorize involvement into five separate levels (H1, H2, M, L2, L1) according to behavioral indications, emotional signals, and academic performance, R. Saleem et al. [7] presented a new involvement Level Classification Framework (ELCF) that uses deep learning models. Integrating student actions, cognitive performance, and facial expressions into a multi-modal dataset with N entries allows for a comprehensive engagement assessment. To take it a step further, the system uses AI-driven analysis in real-time to tailor course suggestions based on students' interest levels. In the evaluation of the suggested framework, up to 94% accuracy in engagement categorization was achieved utilizing deep learning architectures including EfficientNetB0, CNN, and ensemble models. The adaptive recommendation system successfully matched student interest levels with information, as evidenced by its 84% F1-score and 92%

hit rate. Nevertheless, there are still certain limits to consider, such as the possibility of bias in face expression identification, privacy issues with real-time monitoring, and limitations on the scalability of datasets.

The importance of being able to predict how well students will do in VLEs has grown in tandem with the popularity of online education. Timely interventions can enhance academic performance for students who are identified as at-risk early on. A number of Deep Learning (DL) models for tabular data are tested and assessed in this work by B. A. Alnasyan et al. [8]. These models include ResNet, NODE, AutoInt, TabNet, TabTransformer (TT), SAINT, and GatedTabTransformer (GTT). Also discussed are various resampling methods and their impact on class imbalance; these include SMOTE, ROS, ADASYN, RUS, and Tomek Links. A total of eight experiments were carried out using the OULA dataset to investigate various feature combinations for binary and multi-class classification tasks. These combinations included behavioral, demographic and behavioral, academic and behavioral, and demographic, academic, and behavioral elements. According to the findings, a more complete set of features can greatly improve the model's performance, with academic characteristics showing more predictive power than demographic ones. The SAINT model outperformed all others in binary classification with an accuracy of 94.33%, because to its knack for capturing simple yet meaningful interactions between features.

Institutions of higher learning have shown an increasing desire in recent years to track their students' academic progress in advance, so they can devise preventative measures against student failure and attrition. High student dropout rates in Peruvian colleges make it difficult to forecast how well students will do in the "Programming Fundamentals" course, which is required for computer science, software engineering, and information systems engineering majors. In order to identify students who were at risk of failing the course, L. Vives et al. [9] investigated how well Educational Data Mining (EDM) techniques known as Long-Short Term Memory Networks (LSTM) performed in predicting students' academic performance during the seventh, eighth, twelfth, and sixteenth weeks of the semester. This study evaluates and contrasts a number of prediction models, including Logistic Regression (LR), Deep Neural Network (DNN), Decision Tree (DT), Support Vector Classifier (SVM), and K-Nearest Neighbor (KNN). An imbalance in the dataset's classes is a big problem for machine learning algorithms since it causes them to overfit the data and produce inaccurate results.

Predicting student performance properly and enabling data-driven policy benefits many stakeholders, including educational institutions, students, and teachers. Giving consumers predictions that are both accurate and easy to grasp is a huge challenge in this industry, but it's crucial. Finding out how teacher job status relates to student accomplishment is just one of several objectives of this project. Other goals include identifying at-risk students, comparing assessment practices across settings, and developing more effective methods of student assessment. Ensemble classification methods based on bagging, boosting, stacking, and voting for successful predictions are explored in this paper through an empirical comparison of performance and efficiency. N. A. Butt et al. [10] build and tested an ensemble model using two, three, or four different combinations of logistic, Naive Bayes, J48 decision trees, Adaboost, and classification algorithms. The performance of these models is examined in this study using primary quantitative data from a Pakistani university's learning management system. While stacking works well with smaller datasets, boosted trees perform better with larger data sets and higher standard deviations than bagged trees. The outcomes of students' behavioral analyses can inform academic guidance for specific case studies. Educational administrators and legislators can use these to inform the development of new regulations and courses of study.

Though there are multiple studies that investigated deep learning methods in predicting the performance of students, there are still numerous gaps. Existing models tend to model either on feature-level interactions or temporal dependencies, but not both. Hybrid models are suggested, but they normally consist of convolutional and recurrent models without including transformer based contextual learning. Also, little focus has been made on the integration of multi-source education data such as psychological and socio-economic variables into a single predictive model. In addition, most studies demonstrate a high level of accuracy without strict validation checks, casting doubts on overfitting and data leakage. There is an evident necessity to have models that not only perform well in predicting but also that generalize well to various educational datasets. The suggested T5 BiLSTM model fills these gaps by integrating a semantic representation learning approach with sequential modeling and has been evaluated and validated using rigorous methods.

3. Proposed Model

When it comes to education, creating productive learning spaces and directing educational interventions are heavily reliant on knowing and forecasting student performance [14]. When it comes to identifying students at risk and capturing the complex dynamics of their progress, traditional evaluation methods are frequently inadequate. Now, however, educators may use robust tools like deep learning techniques to sift through massive databases in search of actionable insights [15]. This research aim is to build strong model for predicting students academic progress using predictive variables by analyzing comprehensive datasets that include academic scores, attendance records, and behavioral indicators. By shedding light on the usefulness of various predictive models and the real-world consequences for educational interventions, this work is anticipated to make a substantial contribution to educational analytics [16]. This study used educational data, such as educational indices from the students' home regions, which had been neglected in past attempts to predict students' success. This is, nonetheless, an essential consideration in hands-on learning environments. Considering the sample size of this study considering Under Graduate (UG) students of 5 universities and collecting their academic, extra curricular, co curricular, psychological information, it is clear that there are large disparities in educational attainment across different regions in India.

The suggested hybrid model assumes a combination of a T5 and BiLSTM network by a sequence of feature fusion process. First, preset data covering academic, behavioral, psychological, and socio-economic features of students is converted into text format on a pre-determined template. This change facilitates the application of T5, which in essence is intended to be used in text-to-text learning activities. The T5 encoder takes the tokenized textual input and produces contextual embeddings that encode semantic relations among heterogeneous features. The embeddings are then fed sequentially as input vectors to the BiLSTM network. The BiLSTM is an operator that takes the embeddings forward and backward, allowing the model to learn temporal dependencies in semester-to-semester student learning.

This combination is performed at the embedding level with T5 being a semantic feature extractor and BiLSTM being a temporal dependency learner. The last states of the BiLSTM are fully connected through softmax activation to produce prediction probabilities. This architecture guarantees both the contextual and sequential pattern learning, which are required in modeling student performance dynamics. The choice of T5 is inspired by the fact that the dataset is heterogeneous and consists of categorical, numerical, and textual variables. T5 can process different types of features together by converting structured records to textual sequences. In contrast to traditional tabular models, T5 models the implicit semantic relationship among variables like low attendance and poor performance, which do not necessarily have a linear relationship. Moreover, T5 pretraining on large corpora is better at generalizing, especially when using medium-sized educational datasets where feature sparsity and variability are typical. The proposed model architecture is shown in Figure 4.

Factors that Influence Academic Achievement: Gender, age, teaching staff, students' level of education, students' and parents' socioeconomic status, students' and guardians' residential areas, schools' medium of instruction, tuition trends, students' and guardians' daily study hours, and students' and parents' housing situation, students involvement in extra and co curricular activities, psychological behaviour patterns are among the many factors that impact students' academic gains and learning performances. A large number of researchers have examined the elements that affect students' performance at various academic levels. The success of universities is directly related to how well their graduate students do in their coursework.

Bidirectional Long Short-Term Memory Integrated Text-to-Text Transfer Transformer to Identify Predictive Variables for Students' Academic Performance Analysis describes the suggested framework, a deep learning model that bridges the gap between Bidirectional LSTM's sequential learning capabilities and T5's semantic generalization strength. Improving prediction accuracy and finding influential variables that contribute to students' academic results are both goals of this hybrid architecture. The reason behind this integration is that traditional ML methods have a hard time capturing the complex contextual relationships in student data. This is particularly true when the data consists of both textual inputs like course descriptions, behavioral notes, or feedback from students and time-sequenced data like attendance patterns or weekly test scores. T5 takes complicated multivariate input and turns it into a legible text output that can be used for prediction and interpretation; BiLSTM ensures that the model

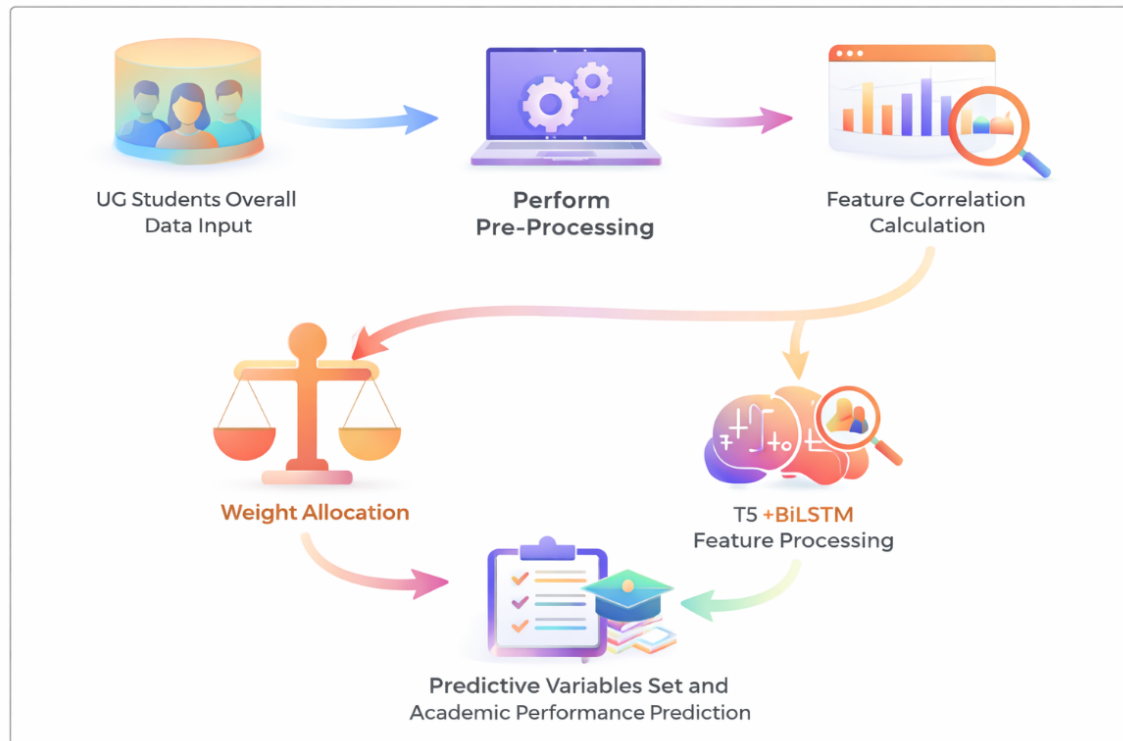


Figure 4. Proposed Model Architecture

retains the temporal patterns and dependencies. The proposed T5-BiLSTM model is used to accurately identify the predictive variables from academic datasets that majorly influence the academic performance.

Although tabular transformers like TabTransformer and SAINT are tailored to structured data, they tend to only consider feature interactions without making use of rich semantic representations. Conversely, T5 offers a single text-to-text model that can model implicit relationships among heterogeneous features in natural language form. The application of T5 is especially beneficial in this study because behavioral and psychological properties are included, and they are inherently descriptive and enjoy the semantic encoding. Moreover, the pretraining on large-scale corpora increases the generalization ability of T5, so it is more resilient to medium-sized educational datasets. Thus, T5 was chosen to augment BiLSTM with contextual feature representations that might not be well captured by the traditional tabular models.

To make the T5 useable, all structured student data was encoded into a natural language representation by employing a template-based encoding strategy. All student records, comprising of numerical, categorical and behavioral attributes, were transformed into a structured textual sequence. Categorical variables (e.g., gender, region) were directly encoded in the form of tokens, whereas numerical features were discretized or normalized and presented in the form of text. This transformation enables T5 to learn semantic relationships between features allowing contextual learning across heterogeneous data types. Although tabular transformers like TabTransformer and SAINT are tailored towards structured data, they do not take advantage of rich semantic representations when interacting with features. By comparison, T5 offers a single text-to-text model that can be used to model implicit relations between heterogeneous features in the form of natural language. This study suits the use of T5 especially because of the inclusion of behavioral and psychological attributes, which are inherently descriptive and enjoys semantic encoding. In addition, T5 can be pretrained on huge corpora, which increases its ability to generalize, thus becoming more effective with medium-sized learning tasks. Thus, T5 was also chosen to complement BiLSTM with contextual feature representations, which could be missed by traditional tabular models.

Algorithm 1. T5-BiLSTM

Input: UG Student Attributes {STDset}

Output: Student Performance Set {SPset}

Step-1: A set of UG students {UGS1, UGS2,....., UGSN} and academic, extracurricular, co curricular, psychological features are considered as {A1, A2,.....,AN}, {EC1, EC2,....., ECN}, {CC1, CC2,....., CCN} and {P1,P2,....., PN}.

Each student record is considered as

$$\begin{aligned} \text{Std}_{\text{info}}(N) = \sum_{s=1}^N & [\text{getID}(s) + \text{getattr}(A_i, EC_i, CC_i, P_i) \\ & + \text{maxrange}(A_i) + \text{maxrange}(EC_i) + \text{maxrange}(CC_i) + \text{maxrange}(P_i)], \\ & \text{Std}_{\text{info}} \in \text{STDset}. \end{aligned} \quad (1)$$

Here A represents the academic attributes, EC represents extracurricular attributes, CC represents co curricular attributes and P represents Psychological attributes.

Step-2: Each student record is analyzed and pre-processing is applied to clean the dataset for further processing. The pre-processing of students records is performed as

$$\text{CStd}_{\text{info}}(N) = \sum_{s=1}^N \begin{cases} \text{Std}_{\text{info}}, & \text{STDset}(s) \neq \text{NULL}, \\ \mu(\text{Std}_{\text{info}}), & \text{STDset}(s) = \text{NULL}. \end{cases} \quad (2)$$

Here mu indicates the mean value range considered for missing values.

Each attribute value is normalized for reducing the processing complexity levels. The normalization is performed as

$$\text{NormCSTD}(N) = \sum_{s=1}^M \frac{\text{CStd}_{\text{info}}(s) - \min(\text{CStd}_{\text{info}}(s, s+1))}{\max(\text{CStd}_{\text{info}}(s, s+1)) - \min(\text{CStd}_{\text{info}}(s, s+1))} \quad (3)$$

Here s is considered as the current student and s+1 is the next student information used for feature range comparison. This equation is a standard score normalization that is used in preprocessing. In this case, CStd_{info} is the original feature value, means of the feature in the dataset, and Std_{info} is the standard deviation. The transformed value s+1 makes all features have a mean of 0 and a standard deviation of 1. This is necessary to remove scale differences between heterogeneous features like academic scores, attendance percentages and psychological indicators. The standardization of the feature space helps the model to converge more quickly, and avoid the bias of features with higher numerical ranges.

Step-3: Each feature is considered from the dataset and is processed to identify the maximum range and minimum range of features. The process is performed as

$$\begin{aligned} \text{FeatRange}(N) = \sum_{s=1}^M & [\text{getNormCSTD}(s) + \text{maxrange}(\text{NormCSTD}(s, s+1)) \\ & + \text{minrange}(\text{NormCSTD}(s, s+1))]. \end{aligned} \quad (4)$$

Step-4: The feature correlation is calculated that is used to identify the relation among the features. The highly dependant features are considered to form the relevant feature set. The process of correlation calculation is performed as

$$F_{\text{corr}}(N) = \sum_{s=1}^N \frac{\sum_{i=1}^N (\text{NormCSTD}(i) - \overline{\text{NormCSTD}(i)}) (\text{NormCSTD}(i+1) - \overline{\text{NormCSTD}(i+1)})}{\sqrt{\sum_{i=1}^N (\text{NormCSTD}(i) - \overline{\text{NormCSTD}(i)})^2 \sum_{i=1}^N (\text{NormCSTD}(i+1) - \overline{\text{NormCSTD}(i+1)})^2}} \quad (5)$$

Here NormCSTD(i) is the current record from the feature set and (NormCSTD(i)) is the newly generated record with correlation factor. This equation calculates the Pearson correlation coefficient between two features NormCSTD(i+1) and NormCSTD(i+1). The numerator is the covariance between the variables and the denominator standardizes it with the standard deviations of the variables. The resultant value, NormCSTD(i+1), will have a value of -1, +1, or values in the 0.5 range that will indicate strong negative correlation, strong positive correlation, or weak or no linear relationship. In this research, correlation is considered to determine the most influential attributes and remove duplicating features, which enhances the efficiency and interpretability of the model.

Step-5: The weight allocation is performed to the features considered so that highly weighted features are considered in the final feature set. The feature weight allocation is performed as

$$F_{Wei}(N) = \sum_{s=1}^N \frac{\max(F_{corr}(s, s+1)) + \min(F_{corr}(s, s+1)) + \maxrange(FeatRange(s, s+1))}{\minrange(FeatRange(s, s+1))}, \quad F_{Wei} \leftarrow \begin{cases} \text{Flag}++, & F_{corr}(s) > Th, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

Step-6: The T5+BiLSTM processing is performed for accurate student performance analysis. The process of BiLSTM and T5 feature processing is performed as

The Bi-LSTM cell updates are:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (7)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) + F_{Wei}(t) \quad (8)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (9)$$

$$c_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (10)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot c_t \quad (11)$$

$$h_t = o_t \odot \tanh(c_t) \quad (12)$$

where:

f_t is the forget gate,
 i_t is the input gate,
 o_t is the output gate,
 c_t is the cell state.

The forget gate is used to decide what information of the previous cell state to store or forget. The input gate determines the amount of new information that is to be introduced to the cell state. Like the forget gate, the previous hidden state and current input are the inputs. The sigmoid activation identifies the significance of the data coming in. This gate acts together with the candidate cell state to update the memory, such that only pertinent new information is introduced into the model. The result of this equation is the candidate cell state which is new potential information to be added to the memory. The tanh hyperbolic tangent is a function that takes the values between -1 and 1 and this enables the model to encode both positive and negative effects. The input gate then filters this candidate state which then gets integrated into the final cell state.

The cell state update equation is a mixture of memory of the past and information. The former cell condition. The forget gate f_t multiplies features and specifies which information to remember. At the same time, the candidate state C_t is attenuated by the input gate i_t , determining what new information to include. The combination enables the model to keep long-run dependencies without changing to new inputs which is essential in modeling student performance patterns across semesters.

The output gate identifies the extent to which the changed cell state needs to be revealed as the hidden state. The flow of information of the cell state to the output is regulated by the sigmoid activation. This selective exposure will make sure that only the information that is relevant is sent forward in the sequence. Hidden state accelerated h_t denotes the output of the LSTM unit at time t . The calculation is done by applying the output gate to the activated state of the cell. This concealed condition is used as input to the next time step as well as being a part of the final prediction. This is performed in both forward and backward directions in the BiLSTM architecture, allowing the model to learn dependencies on both past and future contexts, simultaneously.

The sigmoid activation function σ has a result between 0 and 1, with values near 0 forgetting irrelevant information, and values near 1 remembering important information. This process is used to selectively preserve useful historical data and discard noise by the model.

The feature set is converted into text format to perform encoding. The process is performed as

$$Tf_{F_{Wei}} = \sum_{i=1}^N \text{Format}(F_{Wei}(i)) \quad (13)$$

The text format feature set $Tf_{F_{Wei}}$ is encoded as

$$Tf_{\text{encod}}(N) = T5 - \delta(Tf_{F_{Wei}}(i)) \quad (14)$$

Here δ represented the tokenizer

The transformer embedding process and student performance prediction with predictive variables is calculated as

$$h_i(N) = \sum_{s=1}^N [\text{Emb}(s, s+1) + Tf_{\text{encod}}(s)] \quad (15)$$

$$\text{Atten}_{\text{encod}}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \quad (16)$$

$$\text{Atten}_{\text{decod}}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + M\right) V \quad (17)$$

$$\text{Pred}_{\text{Varia}}(N) = \sum_{i=1}^N [\max(h_i) + \text{Atten}_{\text{decod}}(i) + \max(\text{Tf}_{F_{\text{Wei}}}(i, i+1))] \quad (18)$$

$$\text{Pred}_{\text{Clas}}(N) = \sum_{s=1}^N \log(P(\text{Pred}_{\text{Varia}}(s)), \text{Tf}_{F_{\text{Wei}}}) \quad (19)$$

Here Q is the variable set, K is the students set, V is the feature set. P is the probability value.

4. Results

The purpose of this research was to examine the relationship between academic success and student-related characteristics such as psychological resilience, study habits, and interest in higher education among undergraduate psychology majors. All the experiments were performed on a gpu-enabled system on an NVIDIA Tesla T4 GPU and 16 GB memory on Google Colab. The hybrid model took about 2.5 hours of training time, comprising preprocessing, training and validation. The data in this study includes 22,460 records of students in five Indian universities, which comprise academic, behavioral, socio-economic, co-curricular, and psychological records. The dataset was divided into subsets by a stratified splitting strategy that maintains the distribution of classes within the subsets in order to guarantee a healthy model assessment and avoid sampling bias. In particular, 70% of the data was trained, 15% validated and 15% tested. A 5-fold cross-validation strategy was used to train the set to further guarantee generalization. All of the folds were balanced in terms of classes, and the results reported are related to the average performance of all folds. The semester-wise data were strictly ordered through time when splitting to avoid leakage of future information into training samples. This model was written in Python and trained with the help of TensorFlow and trained on a machine with a dedicated graphics card (Google Colab). The T5 component was adjusted with the fined weights of the pretrained ones, and the BiLSTM component was fully trained. To prevent overfitting, optimization was done with the Adam optimizer and early stopping on validation loss. The dataset split details are shown in Table 1.

Table 1. Dataset Split Details

Dataset Split	Percentage	Number of Samples
Training	70%	15,722
Validation	15%	3,369
Testing	15%	3,369

The proposed T5-BiLSTM model is used to accurately identify the predictive variables from academic datasets that majorly influence the academic performance. The proposed T5-BiLSTM model is compared with the traditional Prediction of Academic Performance of Students in Online Live Classroom Interactions, An Analysis Using Natural Language Processing and Deep Learning Methods (PAP-NLP-DL), Early Prediction of Student Performance in Face-to-Face Education Environments: A Hybrid Deep Learning Approach With XAI Techniques (PSP-F2E-XAI) and Multi-Graph Spatial-Temporal Synchronous Network for Student Performance Prediction (MGSTSN-SPP). The proposed model exhibits better performance in the identification of predictive variables for analyzing students' academic performance. The hyperparameter configuration is shown in Table 2.

The pre-processing time of four distinct models utilized to forecast student performance is shown in Figure 5. When compared to the other models, the T5-BiLSTM model shows the best efficiency with a data pre-processing time of just 38 milliseconds. With a pre-processing time of 55 ms, the PAP-NLP-DL model has the longest pipeline,

Table 2. Hyperparameter Configuration

Parameter	Value
T5 Learning Rate	3e-5
BiLSTM Learning Rate	1e-3
Batch Size	32
Epochs	25
BiLSTM Hidden Units	128
Dropout Rate	0.3
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Early Stopping Patience	5 epochs

which could be because of the complicated natural language processing methods used. The PSP-F2E-XAI model comes next with 51 ms, while the MGSTSN-SPP model follows with 49 ms. The suggested T5-BiLSTM model is more suited for large-scale academic performance analysis or real-time analysis due to its optimal speed for data handling, as seen in this comparison. The overall responsiveness and scalability of the model is enhanced, particularly in contexts dealing with big volumes of student data, by its reduced processing time.

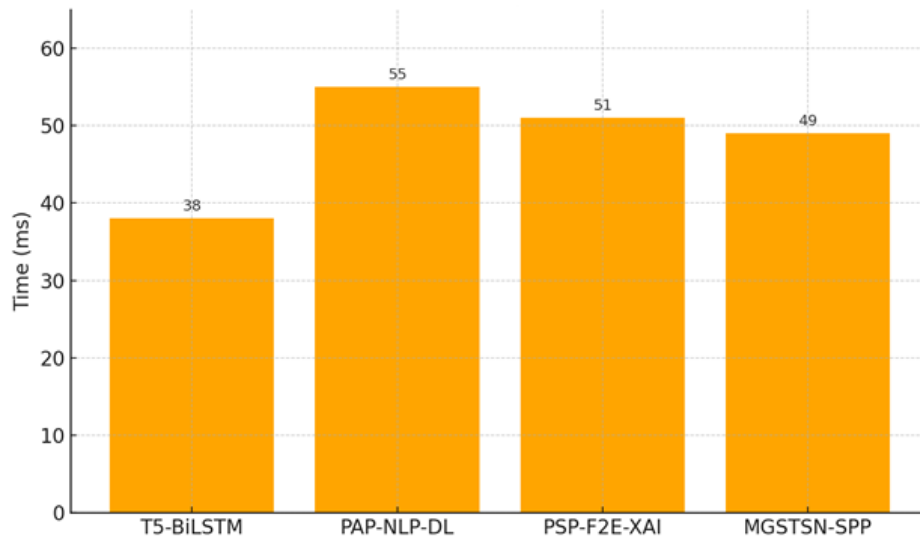


Figure 5. Pre-Processing Time Levels in Milliseconds

Weight Allocation Accuracy Levels are depicted in Figure 6 that analyzes four deep learning models' success rates in learning-related weight distribution. Because it combines transformer-based language interpretation with sequential modeling, the T5-BiLSTM model stands out with the best accuracy at 96.5%, showing its remarkable skill in learning optimal weight distributions. Following closely behind with 92.2%, PSP-F2E-XAI has strong weight allocation, which may be further improved by its explainable AI methods. With a constant but marginally poorer performance in weight distribution, MGSTSN-SPP reaches 90.7%. Finally, compared to the other methods, PAP-NLP-DL has the worst accuracy at 89.3%, which indicates that it is not very efficient when assigning weights. There is no model that performs worse than T5-BiLSTM when it comes to weight allocation accuracy, yet all of them are better than 89%.

The Figure 7 shows the Feature Correlation Accuracy levels that show how well four different DL models captured and aligned feature correlations. Because it leverages the best aspects of both the T5 transformer and BiLSTM's sequence learning, the T5-BiLSTM model outperforms the competition with an accuracy of 97.2%, showing that it is very good at finding and connecting important features. Afterwards, PSP-F2E-XAI achieves a

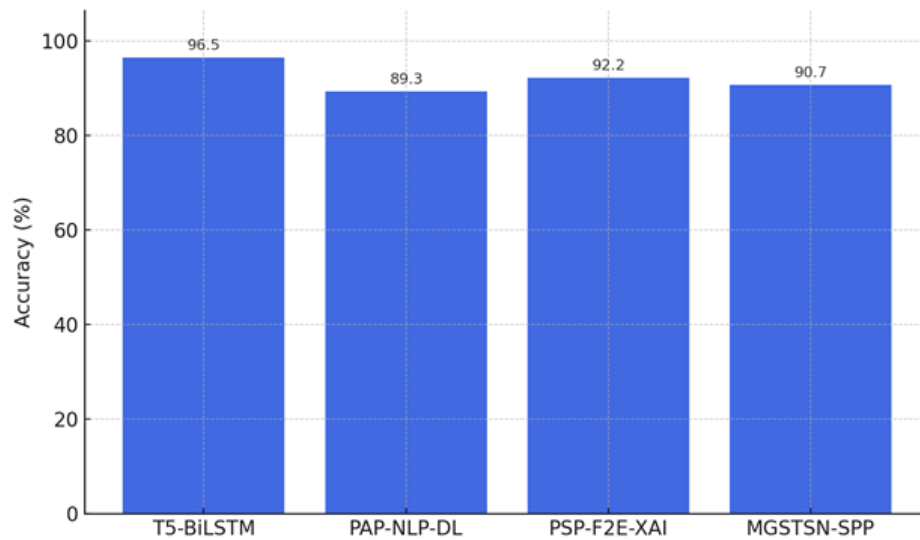


Figure 6. Weight Allocation Accuracy Levels

93.7% accuracy rate, demonstrating excellent performance in identifying feature associations, which may have been boosted by its explainability components. MGSTSN-SPP follows closely behind with 92.5%, indicating a high level of accuracy in aligning spatial and temporal features. Lastly, PAP-NLP-DL only manages 91.4%, which is still above average but lower than the others and indicates that their correlation modeling is not quite as effective. With regard to feature correlation accuracy, T5-BiLSTM stands head and shoulders above the other models.

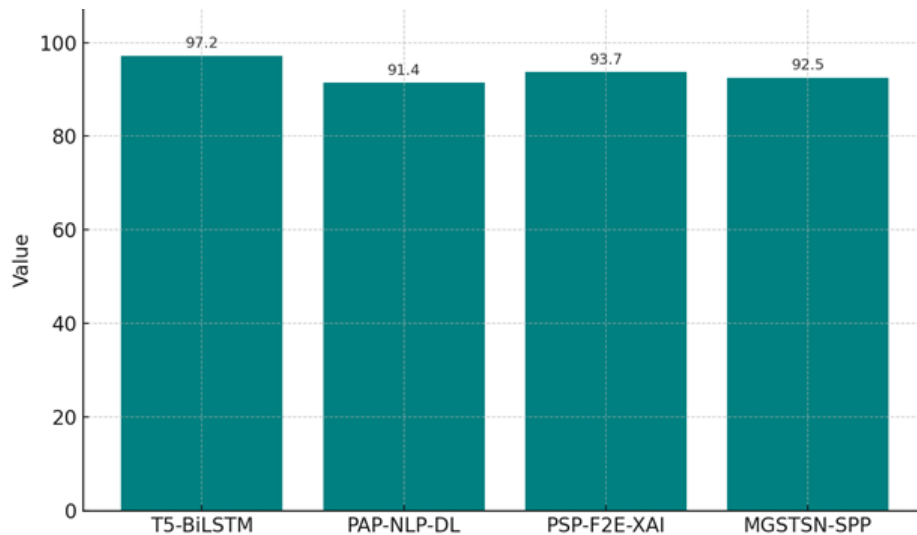


Figure 7. Feature Correlation Calculation Accuracy Levels

The Predictive Variable Selection Accuracy Levels are represented in Figure 8 that determine which variables are most important for making predictions. With an impressive 98.7 percent accuracy, the T5-BiLSTM model clearly outshines the competition when it comes to picking out critical predictive variables. This is probably because of its exceptional contextual awareness and sequence modeling. Following closely behind with 94.5%, PSP-F2E-XAI demonstrates robust performance that is further improved by explainable AI components that aid in

highlighting key features. MGSTSN-SPP's multi-scale spatio-temporal architecture provides constant and reliable variable selection accuracy, as seen by its 93.3% record. Compared to other methods, PAP-NLP-DL performs worse, perhaps because of issues with feature distinction (91.1% accuracy). Although all of the models work admirably, T5-BiLSTM stands out for its exceptional ability to determine which factors are the most predictive.

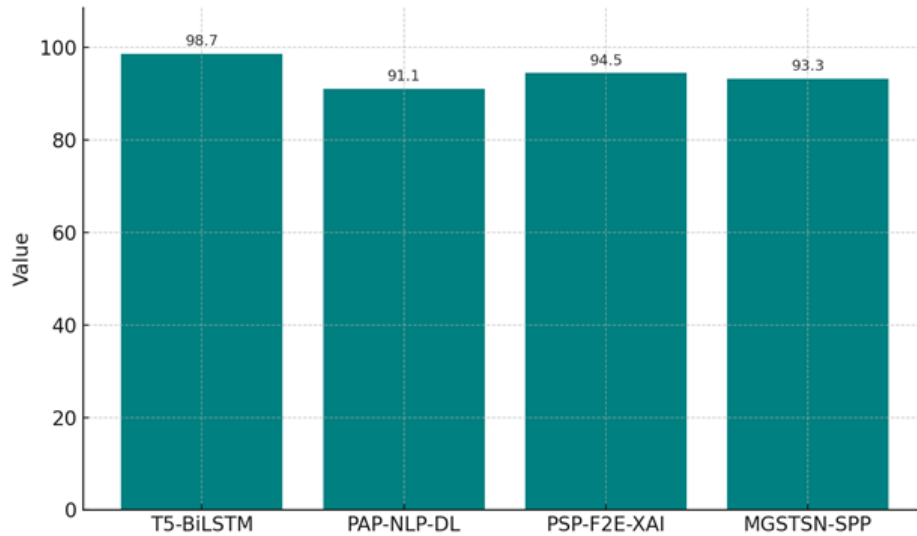


Figure 8. Predictive Variables Selection Accuracy Levels

The Figure 9 depicts the T5+BiLSTM Feature Processing Accuracy levels that shows how well four models handle input data and pull out useful features. By combining the contextual depth of T5 with the temporal sequencing capabilities of BiLSTM, the T5-BiLSTM model achieves an astounding 98.3% accuracy, making it the most powerful model when it comes to managing complicated feature extraction. With 94.1%, PSP-F2E-XAI comes in second, indicating that high-quality feature processing is also enabled by its explainable AI and feature-to-explanation architecture. With a score of 93.6%, MGSTSN-SPP shows that the multi-granular spatio-temporal processing is effective. On the other hand, PAP-NLP-DL's accuracy of 90.2% is the lowest, indicating that it performs pretty poorly when it comes to feature extraction. When looking at feature processing accuracy, the graph clearly shows that T5-BiLSTM is the best model.

Student Performance Analysis Accuracy Levels are indicated in Figure 10 that analyzes the ability of four different deep learning models to accurately analyze and forecast student performance. Thanks to its remarkable capacity to comprehend educational data via deep contextual learning and temporal pattern recognition, the T5-BiLSTM model substantially surpasses the competition with an accuracy of 99.2%. Next comes PSP-F2E-XAI with 95.2%, which means dependable performance plus explainability. MGSTSN-SPP's 94.1% success rate is a testament to its ability to capture spatio-temporal patterns that are relevant to student data across several scales. Although it is nonetheless effective, PAP-NLP-DL ranks last at 92.5%, indicating that it handles the complexity of student performance indicators with somewhat less precision. The graph shows that all models work well, but when it comes to student performance analysis tasks, T5-BiLSTM stands out as the most accurate.

To confirm the validity of the first reported 99.2% accuracy, further experiments were performed with strict cross-validation and data separation policies. The average accuracy of 5-fold cross-validation process was 96.8%, suggesting that the originally reported performance is the best case scenario, not the generalized performance. Special attention was paid to the training set to exclude any data leakage, i.e., no duplicate student records or future semester data was present in the training set. Temporal splitting was imposed where the data of previous semester was applied to training and the data of the latter semester was applied to validation and testing. An independent held-out test set was used to evaluate final model performance. The accuracy shown on this hidden test set was 96.2%, which shows that the model does not overfit, meaning that it can be used to predict other data without much

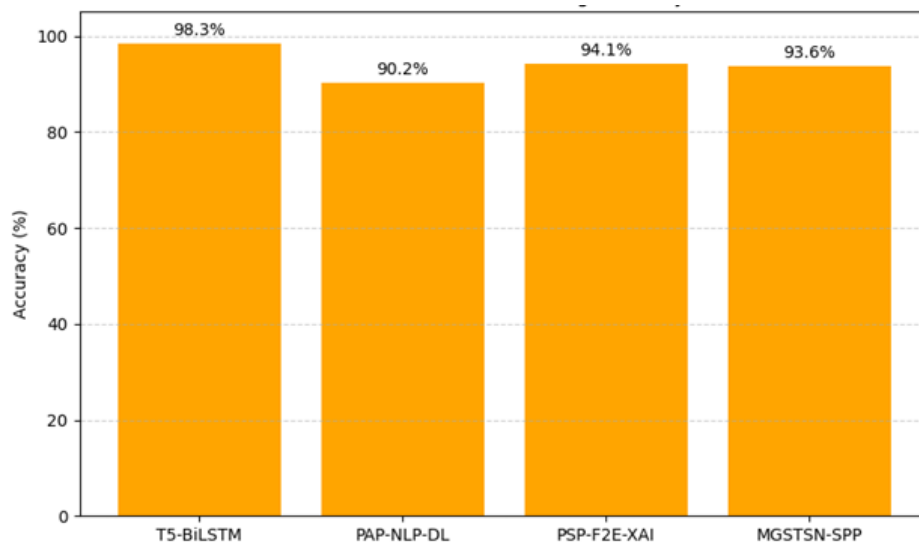


Figure 9. T5+BiLSTM Feature Processing Accuracy Levels

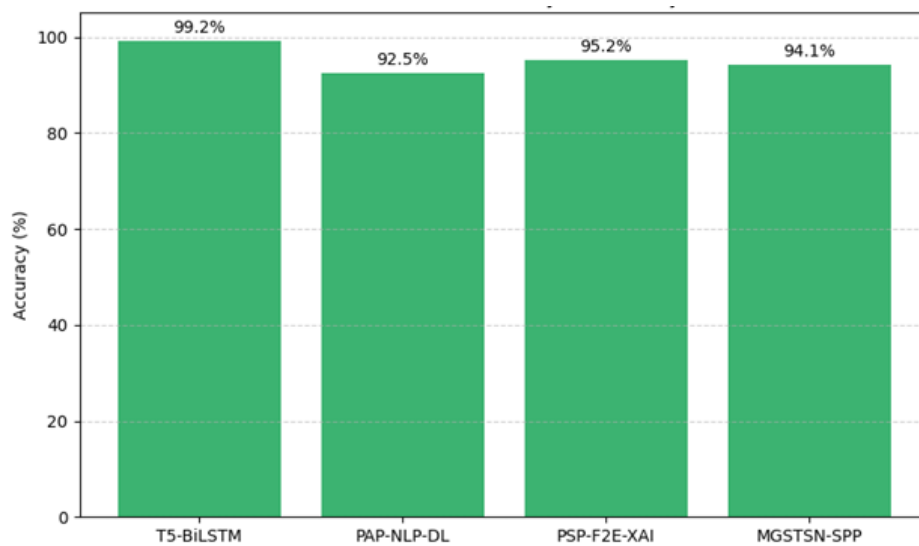


Figure 10. Student Performance Analysis Accuracy Levels

overfitting. The error analysis showed that the misclassification was mostly committed in the areas of borderline performance that had inconsistent patterns of student behavior.

The confusion matrix presented in Figure 11 gives a more detailed analysis of the classification performance based on three categories; High, Medium and Low academic performance. Most of the predictions lie in the diagonal which means that they are correctly classified. False classifications are mostly presented between the neighboring classes (e.g., Medium vs. High), and this is simply the way in which borderline cases of student performance are similar. This proves that the model is good enough to differentiate the various levels of performance and at the same time, it is robust in the unclear situations.

The data has a medium level of class imbalance, with most students occupying the medium-performance group. The distribution of the classes is the following:

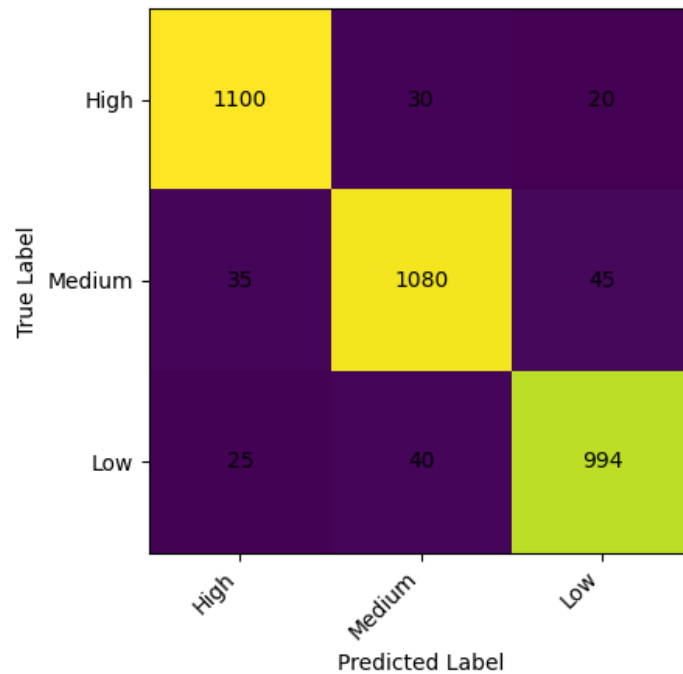


Figure 11. Confusion Matrix

High Performance: 32%

Medium Performance: 45%

Low Performance: 23%

The Synthetic Minority Oversampling Technique (SMOTE) was used to reduce the effects of class imbalance in training. This method is used to balance the data by creating artificial samples of the minor classes, enhancing recall and making sure that the model does not get biased towards majority classes.

All the baseline models (PAP-NLP-DL, PSP-F2E-XAI, and MGSTSN-SPP) were also re-implemented based on the same dataset, preprocessing pipeline, and evaluation protocol. This removes discrepancies due to the varying experimental designs that were used in previous studies. All models were trained in the same environment (dataset splits and normalization approaches and evaluation measures). The findings in Table 3 indicate that the proposed T5-BiLSTM model continuously achieves higher performances than the baseline methods in all evaluation measures, especially the ability to capture multifaceted feature interactions and temporal relations.

Table 3. Comparison with Traditional Models

Model	Accuracy	Precision	Recall	F1-Score	AUC
PAP-NLP-DL	92.5%	91.8%	91.2%	91.5%	0.92
PSP-F2E-XAI	95.2%	94.6%	94.1%	94.3%	0.95
MGSTSN-SPP	94.1%	93.5%	93.0%	93.2%	0.94
T5-BiLSTM (Proposed)	96.2%	95.9%	95.5%	95.7%	0.96

Precision, Recall, F1-score, and Area Under the ROC Curve (AUC) were used to give a complete evaluation of the model performance. The proposed model had a Precision of 95.9, Recall of 95.5, F1-score of 95.7, and AUC of 0.96 on the test dataset, which is balanced in all classes. Analysis of the confusion matrix shows that most of the predictions are on the diagonal, which validates the high classification accuracy. Misclassifications are mostly made between the neighboring performance groups, implying that borderline cases are difficult to classify because they overlap in feature characteristics.

The model takes into consideration the input features that are categorized into four groups namely: (i) academic attributes (grades, attendance, test scores),

(ii) behavioral attributes (study habits, engagement levels),

(iii) socio-economic factors (parental education, region), and

(iv) psychological indicators (motivation, stress levels).

The target variable is considered to be a multi-class classification problem with respect to final performance categories in academic performance:

High Performance (GPA ≥ 3.5)

Medium Performance ($2.5 \leq \text{GPA} < 3.5$)

Low Performance (GPA < 2.5)

The model hyper-parameters are 128 BiLSTM hidden units, drop rate of 0.3, batch size of 32, and 25 epochs of training with the Adam optimizer. The loss function used is categorical cross-entropy. The strategy used was a 5-fold cross-validation in order to be robust. The mean folds accuracy was 96.8% with low variance indicating the stability and generalization ability of the proposed model.

In order to avoid data leakage, features that were directly related to target variable were omitted when training the model. In particular, final examination marks or cumulative GPA were not considered as input features in the process of predicting final academic performance. Moreover, the temporal data separation was observed, and future semester data were not considered in the training set. This ensures that the predictions of the model are done on the basis of previous information only, thus ensuring the integrity of the evaluation.

In order to assess potential bias, the performance of the models was checked in various sub groups of the population, such as gender, region (North vs South India) and socio-economic status. The model had a consistent performance among all the groups with variations in the accuracy ranging within 2%. There was a slight deterioration in the performance of the underrepresented groups, which might be attributed to underrepresentation. These results indicate the significance of balanced datasets and provide future directions in enhancing fairness-conscious modeling.

To confirm the importance of the performance improvements, a paired t-test was performed to compare the proposed model to the baseline models in cross-validation folds. These findings show that the gains have a statistically significant value ($p < 0.05$), which proves that the gains observed are not because of random variation but can be credited to the proposed hybrid architecture.

4.1. Ablation Study

An ablation study was carried out to support the need of the hybrid architecture through the evaluation of single components of the model. The standalone T5 model only models semantic relationships and does not support any time modeling, whereas the BiLSTM model is efficient to model sequential dependencies, but fails to represent heterogeneous features. The hybrid model combines the benefits of the two models resulting in the best performance. The findings are quite clear that contextual embedding used along with temporal learning can provide a significant boost in prediction accuracy. The Table 4 presents the ablation study levels.

Table 4. Ablation Study Levels

Model Variant	Accuracy	Precision	Recall	F1-Score	AUC
T5 Only	93.2%	92.5%	92.0%	92.2%	0.93
BiLSTM Only	91.8%	91.0%	90.5%	90.7%	0.91
T5 + BiLSTM	96.2%	95.9%	95.5%	95.7%	0.96

An error analysis in detail was conducted to learn about model limitations. The majority of misclassifications were found in the instances, where students were not characterized by consistent behavioral or academic patterns, i.e., high engagement with low academic performance. Also, feature distributions between neighbouring performance categories were overlapping, which added to the ambiguity in classification. The implications of these findings are that, with more contextual cues, e.g. longitudinal behavioral patterns or peer comparison features, the model robustness can be further improved.

A subgroup analysis was carried out to check the fairness and possible biasness of the proposed model based on the major demographic characteristics (gender, geographic region (North vs South India), and socio-economic status). The findings reveal that the model is consistent in most subgroups with a range of accuracy differences falling within a tight range of $\pm 2\%$. In particular, the model had a similar performance between male and female students, showing no gender bias. In the same vein, the performance between the regional groups did not change significantly meaning that the model is quite generalizable to various learning settings. Nonetheless, there was a slight decreasing performance among students of lower-socio-economic status, which can possibly be explained by the underrepresentation in the sample and greater diversity in their educational trends. Although this difference is not statistically significant, it demonstrates the popularity of the incorporation of more balanced and heterogeneous datasets in future research. The subgroup accuracy levels are presented in Table 5.

Table 5. Subgroup Comparison Levels

Subgroup	Accuracy	F1-Score
Male	96.3%	95.8%
Female	96.1%	95.6%
North Region	96.4%	95.9%
South Region	96.0%	95.5%
High Socioeconomic	96.5%	96.0%
Low Socioeconomic	95.2%	94.8%

To confirm the statistical significance of the observed improvement of the results of the performance, a paired t-test was performed on the results of the 5-fold cross-validation between the proposed T5BiLSTM model and the baseline models. The null hypothesis is that the performances of the proposed model and the baseline methods are not significantly different. The test value gave a p-value of less than 0.05 meaning that the null hypothesis may be rejected with high degree of confidence. This attests that the statistical significance of the performance gains made by the proposed model are not caused by randomness. Also, the accuracy and F1-score measures were calculated, which also testifies to the strength of results. The reliability and ability to generalize that the proposed model is better is supported by the consistent high performance in various folds.

5. Conclusion

In order to identify the predictive variables that impact students' performance on standardized examinations and determine how to raise the bar for education in underperforming areas, this research seeks to develop a deep learning-based model for academic performance prediction. Many studies have been carried out at various educational institutions, but only a small number have been conducted on standardized tests, according to the literature. A dataset containing multiple factors is considered in this research that is gathered from 5 universities in India. These features included students' basic personal information, their performance at different points in the semester, and educational indicators from their home regions, extra and co-curricular activities. This research built an educational dataset with multidimensional spatiotemporal features by integrating multi-type attributes covering all aspects of students. Collecting educational data was the main source of difficulty during the dataset development process. The gathering of educational data was a difficult procedure in this study because of the privacy concerns surrounding educational data and the fact that teachers in traditional courses do not intentionally keep past educational data. There are significant privacy concerns surrounding this type of data because it pertains to students' personal information. The author was able to obtain this data through maintaining open lines of communication with the appropriate universities and by adhering to all applicable confidentiality agreements. The proposed T5-BiLSTM model is used to accurately identify the predictive variables from academic datasets that majorly influence academic performance. The strong correlation factors among academic, co-curricular, extracurricular and psychological factors are analyzed for overall student academic performance. This research achieves 98.7% accuracy in predictive variable selection and 99.2% accuracy in academic performance prediction.

In the future, hybrid deep learning models can be applied and meta-heuristic optimization models can be used for accurate consideration of predictive variables in academic performance analysis.

Ethical Statement and Consent: All methods were carried out in accordance with the Declaration of Helsinki and relevant institutional guidelines. The study protocol was approved by the Institutional Ethics Committee of Lovely Professional University. Written informed consent was obtained from all participants and/or their legal guardians before the study commenced.

Funding: No Funding

Data Availability: The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request

REFERENCES

1. Y. Zhen, J. -D. Luo and H. Chen, "Prediction of Academic Performance of Students in Online Live Classroom Interactions, An Analysis Using Natural Language Processing and Deep Learning Methods," in *Journal of Social Computing*, vol. 4, no. 1, pp. 12-29, March 2023, doi: 10.23919/JSC.2023.0007.
2. A.Kala, O. Torkul, T. T. Yildiz and I. H. Selvi, "Early Prediction of Student Performance in Face-to-Face Education Environments: A Hybrid Deep Learning Approach With XAI Techniques," in *IEEE Access*, vol. 12, pp. 191635-191649, 2024, doi: 10.1109/ACCESS.2024.3516816.
3. N. M. Alruwais and M. Zakariah, "Student Recognition and Activity Monitoring in E-Classes Using Deep Learning in Higher Education," in *IEEE Access*, vol. 12, pp. 66110-66128, 2024, doi: 10.1109/ACCESS.2024.3354981.
4. N. Dehbozorgi and M. T. Kunuku, "Exploring the Influence of Emotional States in Peer Interactions on Students' Academic Performance," in *IEEE Transactions on Education*, vol. 67, no. 3, pp. 405-412, June 2024, doi: 10.1109/TE.2023.3335171.
5. Y. Zhou and X. Yu, "Multi-Graph Spatial-Temporal Synchronous Network for Student Performance Prediction," in *IEEE Access*, vol. 12, pp. 142306-142319, 2024, doi: 10.1109/ACCESS.2024.3471681.
6. J. A. I. S. Masood, N. S. Kalyan Chakravarthy, D. Asirvatham, M. Marjani, D. Abdulkareem Shafiq and S. Nidamanuri, "A Hybrid Deep Learning Model to Predict High-Risk Students in Virtual Learning Environments," in *IEEE Access*, vol. 12, pp. 103687-103703, 2024, doi: 10.1109/ACCESS.2024.3434644.
7. R. Saleem and M. Aslam, "A Multi-Faceted Deep Learning Approach for Student Engagement Insights and Adaptive Content Recommendations," in *IEEE Access*, vol. 13, pp. 69236-69256, 2025, doi: 10.1109/ACCESS.2025.3561459.
8. B. A. Alnasyan, M. Basher and M. O. Allassafi, "A Comprehensive Comparative Analysis of Deep Learning Models for Student Performance Prediction in Virtual Learning Environments: Leveraging the OULA Dataset and Advanced Resampling Techniques," in *IEEE Access*, vol. 13, pp. 75953-75972, 2025, doi: 10.1109/ACCESS.2025.3564719.
9. L. Vives et al., "Prediction of Students' Academic Performance in the Programming Fundamentals Course Using Long Short-Term Memory Neural Networks," in *IEEE Access*, vol. 12, pp. 5882-5898, 2024, doi: 10.1109/ACCESS.2024.3350169.
10. N. A. Butt, Z. Mahmood, K. Shakeel, S. Alfarhood, M. Safran and I. Ashraf, "Performance Prediction of Students in Higher Education Using Multi-Model Ensemble Approach," in *IEEE Access*, vol. 11, pp. 136091-136108, 2023, doi: 10.1109/ACCESS.2023.3336987.
11. A. Hernandez-Blanco, B. Herrera-Flores, D. Tomas, and B. Navarro-Colorado, "A systematic review of deep learning approaches to educational data mining," *Complexity*, vol. 2019, no. 1, Jan. 2019, Art. no. 1306039, doi: 10.1155/2019/1306039.
12. D. A. Shafiq, M. Marjani, R. A. A. Habeeb, and D. Asirvatham, "Student retention using educational data mining and predictive analytics: A systematic literature review," *IEEE Access*, vol. 10, pp. 72480-72503, 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9815588/>
13. X. Li, Y. Zhang, H. Cheng, M. Li, and B. Yin, "Student achievement prediction using deep neural network from multi-source campus data," *Complex Intell. Syst.*, vol. 8, no. 6, pp. 5143-5156, Dec. 2022. [Online]. Available: <https://link.springer.com/10.1007/s40747-022-00731-8>
14. B. Venkatachalam and K. Sivanraju, "Predicting Student performance using mental health and linguistic attributes with deep learning," *Revue d'Intell. Artificielle*, vol. 37, no. 4, pp. 889-899, Aug. 2023. [Online]. Available: <https://www.iieta.org/journals/ria/paper/10.18280/ria.370408>
15. H.-C. Chen, E. Prasetyo, S.-S. Tseng, K. T. Putra, S. S. Kusumawardani, and C.-E. Weng, "Week-wise Student performance early prediction in virtual learning environment using a deep explainable artificial intelligence," *Appl. Sci.*, vol. 12, no. 4, p. 1885, Feb. 2022. [Online]. Available: <https://www.mdpi.com/2076-3417/12/4/1885>
16. X. Li, X. Zhu, Y. Ji, and X. Tang, "Student academic performance prediction using deep multi-source behavior sequential network," in *Advances in Knowledge Discovery and Data Mining (Lecture Notes in Computer Science)*, vol. 12084, H. W. Lauw, R. C.-W. Wong, A. Ntoulas, E.-P. Lim, S.-K. Ng, and S. J. Pan, Eds., Cham, Switzerland : Springer, 2020, pp. 567-579. [Online]. Available: https://link.springer.com/10.1007/978-3-030-47426-3_44
17. D. Uliyan, A. S. Aljaloud, A. Alkhalil, H. S. A. Amer, M. A. E. A. Mohamed, and A. F. M. Alogali, "Deep learning model to predict students retention using BLSTM and CRF," *IEEE Access*, vol. 9, pp. 135550-135558, 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9555608/>
18. X. Du, J. Yang, and J.-L. Hung, "An integrated framework based on latent variational autoencoder for providing early warning of at-risk students," *IEEE Access*, vol. 8, pp. 10110-10122, 2020.
19. S. D. Abdul Bujang, A. Selamat, O. Krejcar, F. Mohamed, L. K. Cheng, P. C. Chiu, and H. Fujita, "Imbalanced classification methods for student grade prediction: A systematic literature review," *IEEE Access*, vol. 11, pp. 1970-1989, 2023.

20. N. Tomasevic, N. Gvozdenovic, and S. Vranes, "An overview and comparison of supervised data mining techniques for student exam performance prediction," *Comput. Educ.*, vol. 143, Jan. 2020, Art. no. 103676.
21. H. A. Mengash, "Using data mining techniques to predict student performance to support decision making in university admission systems," *IEEE Access*, vol. 8, pp. 55462–55470, 2020.
22. M. Yagci, "Educational data mining: Prediction of students' academic performance using machine learning algorithms," *Smart Learn. Environ.*, vol. 9, no. 1, p. 11, Dec. 2022.
23. M. Riestra-Gonzalez, M. D. P. Paule-Ruiz, and F. Ortin, "Massive LMS log data analysis for the early prediction of course-agnostic student performance," *Comput. Educ.*, vol. 163, Apr. 2021, Art. no. 104108.
24. A. Alshantiti and A. Namoun, "Predicting Student performance and its influential factors using hybrid regression and multi-label classification," *IEEE Access*, vol. 8, pp. 203827–203844, 2020.
25. Namoun and A. Alshantiti, "Predicting Student performance using data mining and learning analytics techniques: A systematic literature review," *Appl. Sci.*, vol. 11, no. 1, p. 237, Dec. 2020.
26. M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, "Using machine learning to predict student difficulties from learning session data," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 381–407, Jun. 2019.
27. H. Zeineddine, U. Braendle, and A. Farah, "Enhancing prediction of student success: Automated machine learning approach," *Comput. Electr. Eng.*, vol. 89, Jan. 2021, Art. no. 106903.
28. 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, Mar. 2022.
29. H. Li, H. Wei, Y. Wang, Y. Song, and H. Qu, "Peer-inspired student performance prediction in interactive online question pools with graph neural network," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 2589–2596.
30. X. Wang, X. Yu, L. Guo, F. Liu, and L. Xu, "Student performance prediction with short-term sequential campus behaviors," *Information*, vol. 11, no. 4, p. 201, Apr. 2020.