Long Short-Term Memory Networks for Predicting Student Performance from Game Play

Jerrica Deloatch
College of Engineering and Technology
Department of Computer Science
East Carolina University
East 5th Street Greenville, NC 27858
Email: deloatchj22@students.ecu.edu

Christopher Johnson
College of Engineering and Technology
Department of Computer Science
East Carolina University
East 5th Street Greenville, NC 27858
Email: johnsonchr22@students.ecu.edu

Robert Johnson
College of Engineering and Technology
Department of Computer Science
East Carolina University
East 5th Street Greenville, NC 27858
Email: johnsonro18@students.ecu.edu

***Abstract*— The integration of game-based learning environments in education has produced vast datasets that present unique challenges and opportunities for data mining and predictive analytics. The primary problem addressed in this research is the limitation of traditional and linear models in capturing the nonlinear dynamics of time-series data inherent in student interactions within educational games. This study introduces a novel approach using Bidirectional Long Short-Term Memory (BLSTM) networks, which are well-suited for such data due to their ability to process sequences in both forward and backward directions, enhancing the understanding of temporal dependencies. Our solution involves two distinct implementations of BLSTM networks: one utilizing 100 timesteps to capture extensive temporal patterns, and another without timesteps to evaluate baseline performance. The models were rigorously tested and compared with both traditional models and other state-of-the-art neural network approaches. Results demonstrate that the BLSTM with 100 timesteps significantly outperforms other models, highlighting its effectiveness in capturing complex sequential patterns and predicting student performance with high accuracy. This research not only underscores the importance of advanced neural network architectures in educational data analytics but also sets a foundation for future developments in personalized educational tools and interventions. The potential for BLSTM networks to enhance predictive accuracy and provide insights into student learning behaviors offers a promising avenue for revolutionizing educational assessments and learning management systems.**

Keywords: Bidirectional Long Short-Term Memory (BLSTM), Educational Data Mining, Game-Based Learning, Time Series Analysis, Predictive Modeling, Neural Networks, Student Performance Prediction

I. Introduction

The integration of games into education has led to a new way of learning, known as game-based learning. This approach has transformed traditional education and created large sets of data. These data sets are mostly time series, a sequence of data points collected over time. They hold valuable information about how students interact with educational games. Research in educational data mining has demonstrated that students' choices and strategies substantially influence learning [1].

A. Background

The 'Student Performance Prediction Problem' has been a focus within the data analytics and mining community, with two primary groups of approaches identified. The first group relies on a traditional approach utilizing generalized linear models [2]. These types of traditional models such as logistic regression models and linear SVMs have a key flaw in that they are unable to accurately learn and predict nonlinear dynamics of a time-series. Significant improvements can be made using nonlinear models, yet their predefined nonlinear form renders them unable to assimilate new nonlinear dynamics found in the data [3].

The second group focuses on the hot topic of Neural Networks. During development, many of these new models were not without problems. The Multi-Layer Perceptron (MLP) network was one of the earliest models used in time-series forecasting. MLP networks have a serious disadvantage, however, due to the fact they cannot create a structural representation of dependent data points [3].

To address this issue Recurrent Neural Networks (RNNs) were devised. RNNs make use of feedback loops that feed into each subsequent layer, allowing each layer to learn any dependent relationship of the previous data point in a sequence [3]. However, this approach too had a major problem. During backpropagation, gradients would either vanish or explode. This issue severely hindered an RNN's ability to capture long-term dependencies, causing poor prediction performance [3].

B. Motivation and research contributions

Tackling this problem has led directly to the invention of Long Short-Term Memory (LSTM) networks. These networks nullify the adverse effects of gradients by employing a distinct backpropagation gradient calculation, enabling a much more accurate capture of long-term dependencies [3]. The significance of LSTM networks lies in their ability to understand and remember patterns over extended periods, a feature that could prove crucial for the analysis of game-based learning. However, it's noteworthy that the application of LSTMs in this context remains poorly researched, despite the acknowledgment that "BLSTM is regarded as one of the leading artificial recurrent neural networks (RNNs) widely used in classifications and predictions based on time series data" [4].

In response to the identified problem, we propose an innovative approach that utilizes Bidirectional Long Short-Term Memory (BLSTM) networks. Our method is specifically adapted for the analysis of game-based educational data, aiming to capitalize on the strengths of BLSTM in processing and comprehending complex sequential data patterns. This becomes particularly crucial in understanding student interactions within game-based learning scenarios. The specific contributions of our research can be outlined as follows:

1) We propose a new way to use LSTM networks for predicting student performance in educational games.
2) We explore the capabilities of BLSTM in managing complex, multivariate, and time-oriented data from educational games.
3) We assess the effectiveness of our model by comparing it with both traditional and state-of-the-art models.
4) We discuss what our findings mean for future educational technology and the development of personalized learning systems.

C. Paper organization

The rest of this paper is organized as follows. Section II reviews related work. Section III describes our approach using BD-LSTM networks. Section IV presents our model's evaluation. Section V concludes with discussions and directions for future research.

II. Related Work

A. Long Short-Term Memory Networks

Long Short-Term Memory (LSTM) networks have emerged as a cornerstone in the realm of sequence modeling, effectively capturing long-term dependencies and addressing the limitations of traditional Recurrent Neural Networks (RNNs). Foundational work has paved the way for advancements in fields ranging from speech recognition to time series prediction. Comprehensive examinations of LSTM's architecture highlight the critical role of learning rates, momentum, and activation functions on performance [5]. The significance of processing sequences in both temporal directions is underscored, showcasing LSTM's applicability in complex pattern recognition tasks [6]. The integration of convolutional structures through ConvLSTM demonstrates LSTM's versatility in capturing spatial-temporal patterns [7].

B. Long Short-Term Memory Networks in Education

In educational settings, LSTM networks offer insights into student behavior and performance prediction, leveraging time-series data to forecast outcomes in online courses and retention rates. Studies introduce models that outperform traditional methods in predicting student performance, emphasizing the importance of LSTM's ability to learn long-term dependencies [8][2]. Combining LSTM with Conditional Random Fields (CRF) illustrates further utility in predicting student retention, highlighting the necessity of feature extraction and model optimization for high precision and recall [4]. Another study applies LSTM networks to analyze data from Learning Management Systems (LMS), showing their utility in educational data mining [1].

C. Long Short-Term Memory Networks for Predicting Student Performance from Game Play

LSTM networks' integration into game-based learning environments offers a focused approach to analyzing and predicting student performance through gameplay. Utilizing LSTMs to trace learning trajectories within educational games demonstrates the potential of deep learning to decode complex interaction data [9]. Exploring the application of machine learning models, including LSTMs, to predict outcomes based on gameplay highlights the transformative impact of LSTMs in enhancing personalized learning experiences [10].

D. Gap Identification and Contributions

While existing research has extensively explored the capabilities of LSTM networks across various domains, a notable gap persists in the tailored application of these models to predict student performance from game-based learning data. This work aims to bridge this gap by introducing an innovative LSTM-based model specifically designed for analyzing the intricate patterns of student interactions within educational games. Our approach extends beyond the scope of traditional applications, offering a more nuanced understanding of learning processes and outcomes in digital game environments.

III. Proposed Method

In this section, we describe our proposed method for predicting student performance from gameplay data. Our approach leverages Bidirectional Long Short-Term Memory (BLSTM) networks, which are well-suited for sequential data processing.

A. Data Preprocessing and Feature Engineering

The initial step involves reading the raw dataset using Pandas with data types explicitly defined to optimize memory usage. After loading the gameplay events, times, and coordinates from the CSV files, we conduct feature engineering, which plays a pivotal role in our methodology.

We categorize features into categorical (CATS), numerical (NUMS), and event-related (EVENTS) types. Categorical attributes undergo a transformation to count the number of unique events within each game session. This dimensionality reduction technique helps capture the variety of interactions without overcomplicating the model with high cardinality data. For numerical attributes, we compute the mean and standard deviation per session to summarize the continuous data streams, capturing the central tendency and variability of gameplay metrics. Additionally, we aggregate event occurrences, converting categorical event types into a quantifiable format by summing their frequencies. This structured data is then merged and filled with placeholders for missing values, ensuring that the final input to our predictive model is complete and adequately represents the intricate gameplay dynamics.

Following preprocessing, the dataset undergoes meticulous segmentation based on level groupings. This step is essential to ensure that future data is not utilized for predictions. The gameplay data is divided into three checkpoint level segments: 0-4, 5-12, and 13-22, each corresponding to a distinct set of questions encountered by players. For each checkpoint, the data is systematically processed to capture the information available up to that point. Subsequently, separate datasets are created for each question set. These datasets are then individually saved as CSV files to facilitate the training of dedicated models for the 18 different questions per session. This partitioning strategy guarantees that the models are trained and validated on data that reflects the progression of the game, mirroring the real-world scenario where future gameplay cannot influence past performance.

B. Model Architecture

Our predictive model is structured around a Sequential architecture that integrates Bidirectional Long Short-Term Memory (BLSTM) layers. The bidirectional nature of the network allows it to learn dependencies from both past and future contexts, which is essential in time-series-like data where the sequence of events could influence the outcome. This architecture is enhanced with Dropout and Batch Normalization layers to mitigate the overfitting risk, ensuring generalization across unseen data. TimeDistributed Dense layers serve as the final part of our network, providing predictions at each time step. The model is compiled with binary cross-entropy loss, suitable for our binary classification task, and utilizes the Adam optimizer for its adaptive learning rate capabilities, which helps converge to an optimal set of weights efficiently.

C. Training and Validation

A K-Fold cross-validation approach is employed to assess the model's reliability. Each fold involves splitting the dataset into training and validation subsets, followed by the training of the model with early stopping based on validation accuracy to prevent overfitting. The process is repeated for all folds, ensuring that every data subset is used for both training and validation.

D. Implementation Details

We utilize TensorFlow and Keras as our primary frameworks, benefiting from their comprehensive suite of tools for deep learning. The code is designed to be modular and reproducible, with a clear separation between data preprocessing, model definition, training, and evaluation blocks.

Algorithm 1 Student Performance Prediction from Gameplay

Require: Raw dataset $\mathbf{D}$ with categorical and numerical features
Ensure: Predictive model $\mathbf{M}$ for student performance

$\mathbf{D}_{processed} \leftarrow$ Initialize empty data frame for processed features
$dataset \leftarrow$ read_csv("gameplay_data.csv", dtypes)
$labels \leftarrow$ read_csv("labels.csv")
for each session s in dataset **do**
  for each category c in CATS **do**
    $n\_unique \leftarrow$ count_unique(dataset[s][c])
    $\mathbf{D}_{processed}[s][c\_nunique] \leftarrow n\_unique$
  end for
  for each numerical feature n in NUMS **do**
    $mean, std \leftarrow$ compute_stats(dataset[s][n])
    $\mathbf{D}_{processed}[s][n\_mean] \leftarrow mean$
    $\mathbf{D}_{processed}[s][n\_std] \leftarrow std$
  end for
  for each event e in EVENTS **do**
    $sum\_events \leftarrow$ sum_occurrences(dataset[s][e])
    $\mathbf{D}_{processed}[s][e\_sum] \leftarrow sum\_events$
  end for
end for
$\mathbf{D}_{processed} \leftarrow$ fill_missing($\mathbf{D}_{processed}$, placeholder)
Define model $\mathbf{M}$ with bidirectional LSTM layers
Compile $\mathbf{M}$ with optimizer and loss function
for $i \leftarrow 1$ to number_of_folds **do**
  $train\_set, valid\_set \leftarrow$ k_fold_split($\mathbf{D}_{processed}, i$)
  $history \leftarrow \mathbf{M}$.fit($train\_set$)
  $evaluation \leftarrow \mathbf{M}$.evaluate($valid\_set$)
end for
$test\_performance \leftarrow \mathbf{M}$.evaluate(test_set)
return $\mathbf{M}, test\_performance$

IV. Experimental Results

The primary objective of our experiments is to demonstrate the effectiveness of the proposed BLSTM network model for predicting student performance from gameplay data.

A. Dataset Description

The Field Day Lab at the University of Wisconsin-Madison, in collaboration with Vanderbilt University and The Learning Agency Lab, has released a comprehensive dataset for the "Predict Student Performance from Game Play" competition. This dataset contains detailed logs from one of the largest collections of game interactions available publicly. Unlike other datasets which may lack detailed temporal data, this dataset meticulously records every action within the gaming sessions, cataloged by format $\langle$session_id$\rangle$_$\langle$question #$\rangle$ for precise tracking of student responses over time. The dataset covers a range of game levels, specifically levels 0-4, 5-12, and 13-22, which are designed to progressively challenge students and assess their learning outcomes in real time. Comprehensive and open-source, this dataset is a crucial resource for developing models that predict student performance, thereby enhancing the effectiveness of educational games across diverse learning environments. Further details and download options for this dataset are available on the official competition webpage at: https://kaggle.com/competitions/predict-student-performance-from-game-play

B. Evaluation Metrics

Model performance is evaluated using several standard metrics, namely Precision, Recall, and F1 scores:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

where TP (true positive) is the number of correctly identified positive cases. This means that the model correctly predicts the positive class. FP (false positive) is the number of incorrectly identified positive cases. This occurs when the model incorrectly predicts the positive class for an observation that is actually negative. TN (true negative) is the number of correctly identified negative cases, indicating that the model correctly predicts the negative class. FN (false negative) is the number of incorrectly identified negative cases, which happens when the model incorrectly predicts the negative class for an observation that is actually positive. Note that the results observed from these formulas are weighted, as they take into account the class distribution within the dataset, giving more influence to classes with a higher number of instances. There is also a macro average method.

The macro average method calculates individual metrics for each class without considering class imbalance and then computes the average of these metrics to provide a balanced measure of overall performance across all classes.

$$\text{Macro F1 Score} = \frac{1}{N} \sum_{i=1}^{N} \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$$

$$\text{Macro Recall} = \frac{1}{N} \sum_{i=1}^{N} \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i}$$

$$\text{Macro Precision} = \frac{1}{N} \sum_{i=1}^{N} \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i}$$

Though we observe the various metrics, the focus is placed on the F1 score, due to its balance between precision and recall, which is crucial in educational settings where both false positives and negatives carry significant weight. Additional metrics like test loss and accuracy are also calculated to provide a comprehensive overview of the model's performance.

C. Comparison Models

In order to assess the efficacy of our proposed framework for predicting student performance, we compared it against five contemporary models. These models represent a diverse set of approaches that range from machine learning algorithms to deep learning architectures designed for time-series and sequence data processing.

1) BLSTM* (our model): This model is structured around a Bidirectional Long Short-Term Memory (BLSTM) network, a variant of traditional LSTMs that process data in both forward and backward directions to capture temporal dependencies more effectively. Designed with a 100-time-step sequence input, the model is capable of understanding long-range dependencies in gameplay data. The inclusion of K-Fold cross-validation aims to provide a more generalized model by ensuring that it is trained and validated across different subsets of the data, mitigating the risk of overfitting to a particular data partition.
2) BLSTM** (our model): Constructed using a BLSTM network, this model processes the sequential gameplay data without the segmentation of K-Fold cross-validation, allowing it to train on the entirety of the available dataset in a singular comprehensive pass. The architecture is designed to recognize patterns and make predictions without the defined time-step constraint, offering a different approach to model temporal sequences that may vary in length, which is particularly common in gameplay data where actions and interactions do not conform to a fixed interval.
3) ANN-LSTM [8]: This model employs Long Short-Term Memory networks to analyze gameplay data and predict student performance metrics. The ANN-LSTM architecture includes layers specifically designed to process sequential data, addressing the challenges of time-dependent patterns in student interactions.
4) TensorFlow Decision Forests [11]: Utilizes TensorFlow's implementation of decision forests to predict student performance. This model combines decision trees with ensemble learning methods, providing a scalable solution that is applied to gameplay data.
5) Random Forest and CatBoost [12]: Combines Random Forest and CatBoost machine learning algorithms. This ensemble method aims to improve prediction accuracy by aggregating the results of multiple decision trees with a gradient-boosting framework.
6) XGBoost with Selected Features [13]: An implementation of the XGBoost algorithm that features a selection of input features tailored to model student performance. This approach emphasizes the importance of feature selection in model accuracy.
7) Convolution Neural Networks Model [14]: Applies Convolutional Neural Networks to process sequential gameplay data. The model's structure allows it to capture spatial and temporal patterns, facilitating the analysis

and prediction of student performance.

D. Results

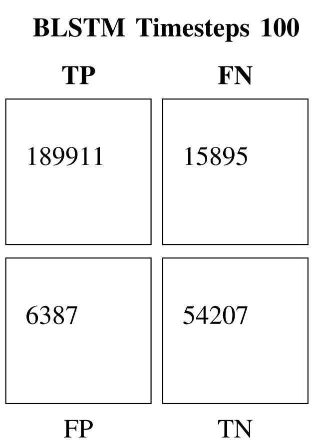
The performance of our models, BLSTM with timesteps and BLSTM without timesteps is compared with established methods to provide insights into their effectiveness in predicting student performance. The BLSTM with timesteps demonstrates a high F1 Score of 0.89, which illustrates a strong alignment between precision and recall, indicative of reliable performance across both classes. Its accuracy of 0.92 further confirms its consistency in correctly predicting student outcomes.

TABLE I
BLSTM TIMESTEPS 100

	Precision	Recall	F1-score	Support
0.0	0.97	0.92	0.94	205806
1.0	0.77	0.89	0.83	60594
Accuracy			0.92	266400
Macro Avg	0.87	0.91	0.89	266400
Weighted Avg	0.92	0.92	0.92	266400

A confusion matrix is a table often used to describe the performance of a classification model on a set of test data for which the true values are known. It typically consists of two dimensions: "Actual" and "Predicted," each split into the number of categories corresponding to the classification tasks, usually two in binary classification tasks.

The confusion matrix provides a visual representation of the classifier's performance. In the matrix, the True Positives and True Negatives represent the number of correct predictions for the positive and negative classes, respectively. Conversely, False Positives and False Negatives reflect the instances where the model has erred in predicting the positive and negative outcomes, respectively.



The BLSTM without timesteps, while not utilizing the K-Fold cross-validation or timesteps, achieves a modest F1 Score of 0.63. This might suggest that while the model is capable, it may benefit from the timestep's ability to capture the sequential dependencies within the data more effectively, which can significantly enhance the accuracy by providing a context for each prediction based on historical information.

TABLE II
BLSTM NONE TIMESTEPS

Class	Precision	Recall	F1-score	Support
0.0	0.58	0.34	0.43	24654
1.0	0.77	0.90	0.83	60594
Accuracy			0.74	85248
Macro Avg	0.67	0.62	0.63	85248
Weighted Avg	0.71	0.74	0.71	85248

In comparison to the other models, such as the ANN-LSTM [8] and the Convolution Neural Networks Model [14], our BLSTM with 100 timesteps model exhibits superior performance metrics, indicating a higher capacity for capturing the complex patterns inherent in the data. The precision metric being closer to 1 for the BLSTM with timesteps model implies that when it predicts a student's performance, it is correct more often than not, making it a reliable tool for educational settings where intervention decisions are critical.

TABLE III
MACRO MODEL COMPARISON

Method	F1 Score	Accuracy	Recall	Precision
BLSTM w/o Time Steps	0.63	0.74	0.62	0.67
BLSTM w/ Time Steps	0.89	0.92	0.91	0.87
ANN-LSTM [8]	0.66	0.72	0.56	0.82
RF & CatBoost [12]	0.68	0.73	0.68	0.67
XGB w/ Sel Features [13]	0.69	0.74	0.69	0.69
GBT Model [11]	0.63	0.76	0.62	0.71
CNN Model [14]	0.57	0.74	0.58	0.69

In contrast, the CNN Model, despite its architectural advantages in feature extraction, yields a lower F1 Score of 0.57, which may point to a need for more extensive training data or refinement of its architecture to handle the specific nuances of gameplay data in educational contexts.

The Gradient Boosted Trees Model, represented by the TensorFlow Decision Forests [11], Random Forest, and CatBoost [12], along with the XGBoost with Selected Features [13], all demonstrate competitive but not leading results. This indicates that while ensemble methods are strong contenders, the recurrent nature of BLSTM and the capability to understand both past and future context in data sequences offers a distinct advantage for this type of time-series prediction task.

Our results emphasize the significance of capturing temporal dynamics in predicting student performance from gameplay data, as exemplified by the BLSTM model with timesteps. This approach could facilitate more precise and timely educational interventions.

V. DISCUSSION

Our examination of BLSTM models, differentiated by their use of 100 timesteps and no timesteps, raises several points of interest. The model utilizing 100 timesteps has shown an improvement in accuracy, highlighting the value of capturing extended temporal patterns within the gameplay data. This capacity to leverage longer sequences assists the model in contextualizing student actions over time, providing it with a comprehensive perspective for prediction.

The advance in accuracy can be attributed to the model's enhanced capability to integrate information across a wider temporal window. This broader view allows for the identification of long-term trends and dependencies that are not immediately apparent with shorter sequences or no timesteps at all. It is especially relevant in educational data, where student behaviors and performance indicators unfold over a period and are not strictly event-driven.

Addressing the risk of data leakage, which is a significant concern when time-based features are used, our methodical approach to data handling has been crucial. By systematically separating the data for each question into individual CSV files and enforcing strict isolation between training, validation, and test sets, we have ensured the integrity of our evaluation process. This diligent partitioning reinforces our confidence that the enhanced performance is not due to data leakage but is an authentic reflection of the model's predictive strength.

VI. Conclusion

This study has put forth a comprehensive analysis and application of Bidirectional Long Short-Term Memory (BLSTM) networks for the prediction of student performance in game-based learning environments. Our investigations have led to the development of two distinct BLSTM models—one utilizing a 100 time-step sequence input, and another processing data without this temporal segmentation.

The BLSTM model with time steps has demonstrated a noteworthy increase in predictive accuracy, substantiating our hypothesis that extended temporal patterns play a significant role in understanding and forecasting student interactions and performance in educational games. The use of multiple time steps has been essential in understanding the complex patterns of sequence dependencies, highlighting the detailed nature of learning behaviors as they develop over time.

On the other hand, the BLSTM model without time steps, while showing competence, has highlighted the potential for enhanced performance with the adoption of temporal sequences in training. This model's more modest performance emphasizes the complexity of the task at hand and the necessity for sophisticated modeling techniques that can capture the evolving nature of educational data.

Our methodological framework has successfully addressed initial concerns regarding data leakage. Through the meticulous separation of data and a thorough training-validation-test split, we have established a solid protocol that upholds the integrity of our predictive modeling process.

The comparative analysis with other contemporary models has provided valuable insights into the strengths of our proposed methods. While other models have their merits and applications, our BLSTM models' ability to process both the past and future context has showcased a clear advantage in this domain.

This research marks a substantial advance in educational technology, especially in analyzing game-based learning. Our results highlight the critical role of time in making predictions and provide a practical way forward in developing personalized learning aids. As game-based learning expands, the techniques we've described here are set to lead to better and more captivating educational experiences.

In conclusion, the exploration of BLSTM networks in this context has created opportunities for deeper understanding and improvement of student learning outcomes through analytical modeling. It stands as a testament to the potential of deep learning to revolutionize educational assessment and intervention strategies, paving the way for future innovations in the field.

VII. Future Work

This research project, conducted as part of a data mining course over a single semester, has provided valuable insights into the application of BLSTM networks for predicting student performance in game-based learning. Given the limited timeframe, it was not feasible to explore every aspect of this promising area as thoroughly as we would have liked.

Moving forward, we aim to extend our investigation in several key areas. First, we plan to refine the BLSTM model by experimenting with various configurations of time steps and network parameters to optimize performance further. Additionally, integrating more diverse and comprehensive datasets could enhance the generalizability and reliability of our findings.

Another exciting area for future research involves exploring the potential of hybrid models that combine BLSTM with other advanced machine learning techniques, such as reinforcement learning or graph neural networks. These models could potentially capture even more complex patterns and interactions in the data, offering better insights into student learning processes.

Ultimately, the continuation of this work will focus on closing the gap between theoretical research and practical application, ensuring that our findings contribute effectively to the field of educational technology.

References

[1] Fu Chen and Ying Cui. Utilizing Student Time Series Behaviour in Learning Management Systems for Early Prediction of Course Performance. *Journal of Learning Analytics*, 7(2):1–17, September 2020.

[2] Byung-Hak Kim, Ethan Vizitei, and Varun Ganapathi. GritNet: Student Performance Prediction with Deep Learning, April 2018. arXiv:1804.07405 [cs, stat].

[3] Cathal John Murray, Naomi Du Bois, Lynsey Hollywood, and Damien Coyle. State-of-The-Art Deep Learning Models are Superior for Time Series Forecasting and are Applied Optimally with Iterative Prediction Methods, February 2023.

[4] Diaa Uliyan, Abdulaziz Salamah Aljaloud, Adel Alkhalil, Hanan Salem Al Amer, Magdy Abd Elrhman Abdallah Mohamed, and Azizah Fhad Mohammed Alogali. Deep Learning Model to Predict Students Retention Using BLSTM and CRF. *IEEE Access*, 9:135550–135558, 2021. Conference Name: IEEE Access.

[5] Thomas M. Breuel. Benchmarking of LSTM Networks, August 2015. arXiv:1508.02774 [cs].

[6] Alex Graves, Santiago Fernández, and Jürgen Schmidhuber. Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Dough Tygar, Moshe Y. Vardi, Gerhard Weikum, Włodzisław Duch, Janusz Kacprzyk, Erkki Oja, and Sławomir Zadrożny, editors, *Artificial Neural Networks: Formal Models and Their Applications – ICANN 2005*, volume 3697, pages 799–804. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005. Series Title: Lecture Notes in Computer Science.

[7] Xingjian SHI, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun WOO. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[8] Fatima Ahmed Al-azazi and Mossa Ghurab. ANN-LSTM: A deep learning model for early student performance prediction in MOOC. *Heliyon*, 9(4):e15382, April 2023.

[9] Danial Hooshyar, Yueh-Min Huang, and Yeongwook Yang. GameDKT: Deep knowledge tracing in educational games. *Expert Systems with Applications*, 196:116670, June 2022.

[10] Hyunbae Jeon, Harry He, Anthony Wang, and Susanna Spooner. Modeling Student Performance in Game-Based Learning Environments, September 2023. Publication Title: arXiv e-prints ADS Bibcode: 2023arXiv230913429J.

[11] Gusthema. Student performance w/ tensorflow decision forests. Kaggle Code, 2023. Accessed: 2024-04-21.

[12] javihm77. Random forest lb & catboost gameplay chapu. Kaggle Code, 2024. Accessed: 2024-04-21.

[13] Aravind Aspillai. Predict student performance xgb fold sel feats. Kaggle Code, 2024. Accessed: 2024-04-21.

[14] Guru Patil. Research lstm and cnn. Kaggle Notebook, 2023. Accessed: 2024-04-21.