2025 Third International Conference on Industry 4.0 Technology (I4Tech)
Pune, India. Sep 18-20, 2025

Recognizing Teacher and Student Activities in Higher Education Using a Hybrid CNN-LSTM Approach

M. V. Krishna,
Assistant Professor, Department of Information Technology, Prasad V Potluri Siddhartha Institute of Technology,
Kanuru, Vijayawada, India,
imvkrishna@pvpsiddhartha.ac.in

Dr. K. Nanthitha,
Department of English, St.Joseph's College of Engineering,
Chennai, Tamil Nadu,
nanthithaenglish@gmail.com

Mustafayeva Dunya Azeri,
Deputy Director of UNEC Business School, Dunya, Azerbaijan State University of Economics (UNEC)
Dunya,
mustafayeva@unec.edu.az

Serikkhan Zhuzeyev,
Humanitarian and Pedagogical Institute, Korkyt Ata Kyzylorda University,
Kyzylorda, Kazakhstan,
serikkhan.zhuzeyev@mymail.academy

Dr. Raj Kumar Gupta,
Senior Assistant Professor, Department of Physics, Sardar Vallabhbhai Patel College,
Bhabhua (VKSU, Ara, Bihar), India,
rajkgsw66@gmail.com

Rudresh S,
Department of Management, St. Claret College Autonomous,
Bangalore, Karnataka, India,
rudip.b3@gmail.com

2025 Third International Conference on Industry 4.0 Technology (I4Tech) | 979-8-3315-2511-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/I4Tech64670.2025.11277856

***Abstract* – Adapting to new academic and social settings is a major problem for first-year students in university. This problem has become increasingly more pressing due to the increasing complexity of societal, spiritual, and economic shifts. When it comes to helping students adjust to college life and thrive academically and personally, no one does a better job than universities and colleges. Data preparation, PCA, and model training make up the three stages of this study's methodology. The feature selection process used min-max normalisation, and principal component analysis was used to lower the dimensionality of the input and output indicators. The suggested architecture combines CNNs for hierarchical feature extraction and LSTMs for collecting long-term dependencies in temporal data; both networks work in tandem to achieve this goal. At 203.64% MAE, 241.22% RMSE, and 1.78% MAPE, the CNN-LSTM model fared better than the alternatives. This research proves that the model is useful for assessing college students' actions. Finally, the study highlights how important it is to use advanced machine learning techniques to help students adjust and how teachers may play a bigger role in creating meaningful learning experiences.**

Keywords—Back Propagation Through Time (BPTT), Principal Component Analysis (PCA), Feedforward Neural Network (FFNN), Multi-Layer Neural Network (MLNN).

I. INTRODUCTION

The material on student-teacher in higher education is enough to leave anyone is wildered. Researchers asserting to be examining the same subject frequently employ distinct variables and methodologies, and the issues under inquiry are highly diverse. In comparison a typical engaged student doe's care about their grades rarely shows up to class skips out on extracurricular and has minimal contact with teachers and students. In his groundbreaking work on higher education argued that teachers' examples of immediacy contribute to student conduct the quality of a higher education student's course experience was predicted by the immediacy of their instructors. There is some evidence that the level of student-teacher contact in university classrooms affects the number of absence [1]. To compare the beliefs of the teachers and students the coded data was used in conjunction with a Degree of Similarity Scale. In order to determine how similar the instructors' and students' perspectives were researchers used a scale that was created specifically for this study to assess the quantity and intensity of educational ideas held by both groups. Teachers and students can be systematically ranked according to how similar their beliefs are using the Degree of Similarity Scale [2]. The scale ranges from very low to very low. After identifying the participants' educational views and determining the degree to which teachers' and students' beliefs were similar the study's quantitative data were scrutinized to confirm and strengthen the conclusions reached from the qualitative data analysis.

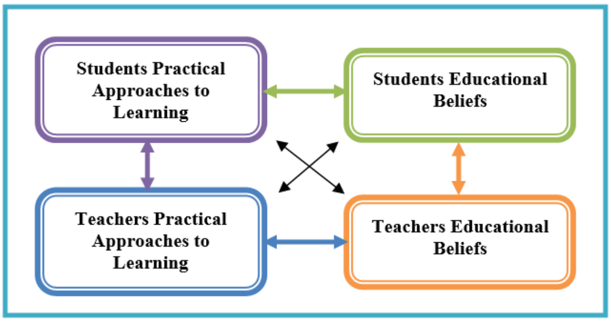

Fig. 1. Tearchers and Students Network for Educational Practies amd Beliefs

The results show that educational beliefs are connected to the actual ways in which both students and teachers approach learning and teaching, and the other way around. These results also show how important it is to think about these connections while making lessons [3]. Figure 1 depicts this intricate web of ideas and behaviours held by both instructors and pupils. They tackle the topic of higher education instruction from the perspectives of both instructors and students. The fact that many different types of learning outcomes were taken into account helped with this [4]. Students' motivation and attitude towards the subject remain uncertain even though student-activating instruction in a project-based learning setting seems to greatly enhance the acquisition of practical skills. When it comes to tests of new information in particular the allow us to make any definitive conclusions about the connection between active learning and better comprehension. The importance of these interactions in college for both students and instructors has been highlighted by findings that link student-teacher connections to learning classroom management and student absenteeism. The need to reevaluate the conventional wisdom about student-teacher relationships which have been defined in terms of proximity or immediacy and evaluated in [5].

The subsequent structure of the paper is outlined as follows. Section 2 addressed pertinent research within our field of interest and included a summary of the higher education system. Section 3 presented the CNN-LSTM framework, which will execute the proposed technique. Section 4 presents the experimental data and a comparison of the approaches. Our conclusions and recommendations for future research are presented in section 5, the final section of this report.

II. Literature Survey

The phenomenon of student-teacher relationships in university environments has often been analyzed through the concept of teacher immediacy defined as the extent to which the teacher signals of warmth friendliness and affection. In order to maximize the model's congruence with the input data DL approaches use multi-neuron architectures to carry out learning tasks with neurons connected to the data via a loss function that allows for weight updates [6]. In RL agents interact with their environment in a trial-and-error fashion to maximize cumulative rewards all without the need for labeled data. This framework allows for experience-based autonomous learning. Methods like policy search and approximating value functions are the backbone of RL. In order to better manage and develop data two tools that analyze and interpret data were created: AI and ML. Many industries including education have recognized the trend of integrating these processes into company as having the potential to revolutionize their operations [7]. Higher education institutions and online education can thus greatly benefit from the application of AI and ML. Explore the role DL as a mediator between student-centered teaching methods and the enhancement of students' theoretical and practical skills. It developed a competency-based assessment instrument for classrooms by DL combining exploratory confirmatory and reliability analyses [8]. Students' use of DL techniques and self-reported ability improvement were both positively predicted by student-centered teaching in large classes and DL mediated the relationship. With the development of ML methods some studies have evaluated learning system usage prediction models using a hybrid strategy that combines ML with more conventional SEM. Based on students' real-world tasks in information management one study by used their self-reported data to predict their behavior in regard to educational mobile cloud computing [9].

Through the integration of ML methods with a conventional SEM. Innovations in data science as a discipline, big data for large-scale analysis, and state-of-the-art methods and technologies such as AI, NN, DL and ML have paved the way for new approaches to quantitative knowledge production and decision-making complementing traditional statistical methods [10]. The goal of this study is to take a look at how different assessment and evaluation methods that leverage AI can improve educational outcomes. A major goal of this research was to compile a list of the most widely utilized AI and ML algorithms for enhancing students' academic performance. It compared FCN to its rivals using ANN, XG Boost, SVM, RF and DT as performance indicators. The FCN consistently produced better performance [11]. The results of the comparative analysis study can help school administrator's principals and teachers better understand how to incorporate data into their practice develop less DL assessments that are still valid reliable and constructive and discover new ways to use assessment and feedback to Adaboost students' learning [12].

This research introduces an innovative DL architecture that integrates LSTM networks with CNNs to address the issues associated with Teachers and Students Activities. This approach applies a CNN preprocessing step to convert raw data into multidimensional inputs in an effort to improve LSTM's prediction performance. Datasets pertaining to students' academic performance proved that this hybrid model was effective, showing that it outperformed conventional prediction approaches and independent CNN and LSTM models.

III. Proposed System

According to studies in education, creating a positive classroom climate for learning depends on students and teachers developing a strong relationship. Educators and politicians must scrutinise college classroom methods and rules regarding teaching and learning in university environments, considering the disparities in student outcomes. Colleges draw a varied student population and yield varying outcomes, underscoring the necessity for more dynamic pedagogical approaches in higher education.

The collection has 480 student records and sixteen attributes. The features are categorised into three primary groups: Gender and nationality exemplify demographic characteristics. Characteristics of the learner's educational background, encompassing the present level of education, grade, and sector. Behavioural characteristics including hand-

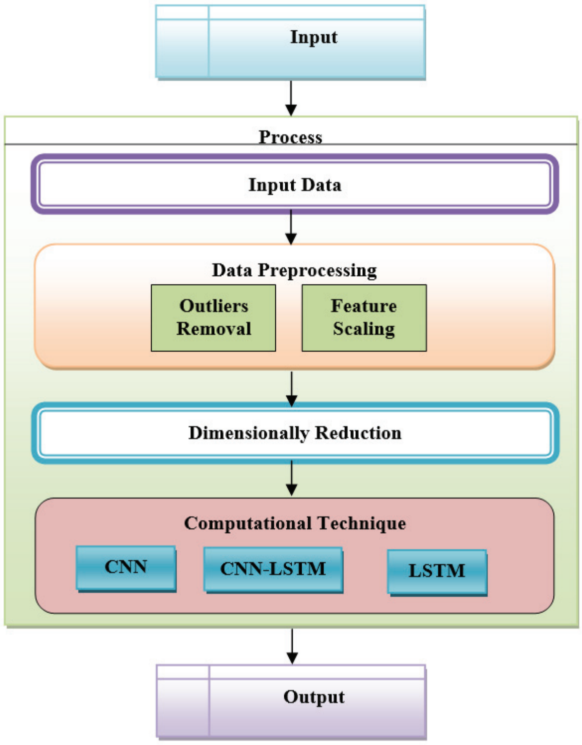

Fig. 2. Proposed Model Research Framework

raising in class, accessing resources, parental survey responses, and overall school satisfaction. The dataset was compiled during two academic terms, comprising 245 records from students in the first term and 235 records from the second. The dataset contains information on students' school attendance, categorised into two groups based on the number of days absent [13]. Of the learners, 289 have fewer than seven days of absence, whereas 191 have more than seven days of absence.

Figure 2 shows the suggested system's research framework. Information Gathering, Data Pretreatment, Computing Method, and Dimensionality Reduction [14].

A. Data Preprocessing:

Before training our DL models, which include prediction and classification models, this section details the data pretreatment methods that were carried out. There were three stages to the preprocessing step: A) Eliminating extreme values, and B) Scaling aspects that are thought to be indicative of the quality of the education that is being delivered. Python, gnuplot, and Microsoft Excel were utilised to do the preprocessing [15].

Most people could not have predicted the presence of outliers. If one item stands out from the others by a large margin, it call it an outlier. To find the really significant data points that should be removed from the dataset, outlier analysis is a necessary. The following sections detail our data pretreatment steps, which included analysing and removing outlier data.

1) Outliers Removal:

The lack of documentation for a formal absence from school means that kids are sometimes taking short breaks or completely dropping out. As a result, they narrowed our focus to students whose TTD was less than ten years and who had not skipped more than two consecutive semesters. About 2% of the whole dataset was comprised of the eliminated outliers.

2) Scaling the Feature:

range normalisation through feature scaling. Data normalisation is a prevalent operation in data processing, usually conducted during the data pretreatment phase. Features were rescaled to the [0, 1] interval as a result of using the min-max normalisation strategy. Equation (1) shows the formula that was used to normalise and scale all data values:

$$V' = \frac{V - min(V)}{max(V) - min(V)} \tag{1}$$

The normalised value is denoted by V', and V represents the original data value.

B. PCA:

The inputs and outputs of higher education are so diverse that conventional business analysis tools, including production functions and ROI, are inadequate for studying input-output relationships. In order to demonstrate the association among the input-output aspects, this research will utilise multivariate analysis, namely PCA [16]. The 31 federally-administered autonomous regions, provinces, and municipalities can have their efficiency input-output compared and studied using PCA, which reduces multidimensional variables to a small set of synthetic variables. There are four steps to extract the primary component:

1) The Original Variable Standardize:

Equation (2) can be used to construct a data matrix V, assuming there are s samples and r indicators:

$$V = \begin{bmatrix} V_{11} & V_{12} & \dots & V_{1r} \\ V_{21} & V_{21} & \dots & V_{2r} \\ V_{s1} & V_{s2} & \dots & V_{sr} \end{bmatrix} \tag{2}$$

To remove the negative effects of having different dimensions and orders of magnitude for each variable, all of the variables will be standardised. Equation (3)provides the formula for one method called Z-score that can be used for data standardisation:

$$y_{ml} = \left(V_{ml} - \overline{V}_l\right)/N_l \tag{3}$$

Following the transformation, $\overline{V}_l = \sum_{m=1}^{s} V_{ml}/s$the mean becomes zero and the variance becomes one. $N_l^2 = \sum_{m=1}^{s}\left(V_{ml} - \overline{V}_l\right)^2/(s-1)\ m = 1,2,..s\ , l = 1,2,..o$ .Statistical Package for the Social Sciences allows for the automatic completion of the transformation.

2) The Covariance Matrix of Indicators and Data Evaluate:

The covariance matrices for data and indicators can be computed using the subsequent formula. In equation (4), the covariance matrix $P = (p_{lo})_{r\times r}$.

$$p_{lo} = \frac{1}{s-1}\sum_{m=1}^{s} y_{ml} \times y_{mo} \tag{4}$$

In the formula, p_{lo} denotes the coefficient correlation between indicator l and indicator o, where $m = 1,2,..s\ , l = 1,2,..o, thus\ p_{mm} = 1, p_{lo} = p_{ol}$.

C. Model Training:

1) LSTM based RNN:

A subset of RNNs known as LSTM models incorporate connections feedback into every neurone. Because RNNs take into account both the present neuron's weights and inputs as well as those of neurones that came before them, they are naturally well-suited to processing sequential data. However, LSTMs were developed because RNNs encounter major problems, like gradient explosion and vanishing issues, when processing long-term data that is sequential [17].

By including techniques to selectively preserve or reject information, LSTMs are supposed to reduce the problem of gradients that vanish encountered by RNNs. The following components are used to characterise the computation forward process in an LSTM.

$$A_q = h_g.A_{q-1} + m_q.\tilde{A}_q \tag{5}$$

$$K_q = \rho\left(U_{kq}.f_{q-1} - U_{kv}.v_q + d_k\right) \tag{6}$$

$$f_q = K_q.\tan f\left(A_q\right) \tag{7}$$

The current state cell value, the prior state cell value, and the update to the cell current state are denoted as A_q, A_{q-1} *and* $\tilde{A}_q$, respectively. h_q, m_q, and k_q stand for the input gate, forget gate, and output gate, respectively. Equations (5) to (7) are used to generate the output value f_q, with the necessary parameter settings, using the values of $\tilde{A}_q$ and A_q. The BPTT technique updates all weight matrices depending on the discrepancy between the output values and the real values.

2) Temporal CNN:

Among the many types of deep learning neural networks, CNNs are now seeing the most widespread use in computer vision, particularly for teacher and student's activities and classification tasks. Despite the large volume of raw data samples, CNN are able to quickly extract a valuable subset. After MLNN, CNNs evolved into FFNN. Parameter sharing and sparse interactions are two features that set CNN apart from more conventional MLNNs.

Each output neurone has the ability to interact with every input neurone in traditional MLNNs since they use a fully linked technique to create neural networks between the input and output layers. The weight matrix would consist of $i \times s$ parameters if there are i input neurones and s output neurones.CNNs use $o \times o$ kernels convolutional to drastically decrease the weight matrix's parameter count. Two attributes of CNNs enhance the efficiency of parameter optimisation during training; CNNs are capable of building deeper neural networks with additional hidden layers at equivalent computational expense.

To analyse univariate data, Temporal CNN use specific 1D convolutions. In contrast to conventional CNNs, which utilise convolutional kernels of size $o \times o$, temporal CNNs utilise $o \times 1$ kernels. A dataset with i -dimensional characteristics can be extracted from the initial univariate dataset via temporal convolution techniques. Consequently, temporal CNNs transform time series data from a univariate dataset into a multi-dimensional feature-extracted dataset using 1D convolution; consequently, LSTM prediction is more appropriately aligned with the enhanced multi-dimensional feature data.

3) The Framework Prediction for CNN-LSTM:

The research presents a DNN that combines CNN and LSTM models to address dependency sequence and univariate data. Through the application of convolution, the CNN transforms input data from a univariate format into a multidimensional feature format, thereby extracting essential information during the preprocessing stage. During the subsequent phase, the restructured feature data are fed into LSTM units for forecasting.

Figure 3 shows the architecture of the CNN-LSTM hybrid neural network model that is suggested in this article [18].

The input dataset undergoes preprocessing via a convolutional neural network comprising two hidden layers. Significantly, when the quantity of hidden layers surpasses five, conventional temporal CNNs generally incorporate pooling methods to mitigate overfitting. This article excludes pooling processes to optimise the retention of extracted feature information.

In order to train and predict diseases, an LSTM neural network is created once the input data has been preprocessed. The LSTM neural network is augmented with a layer dropout to mitigate overfitting. To optimise the weights of all LSTM

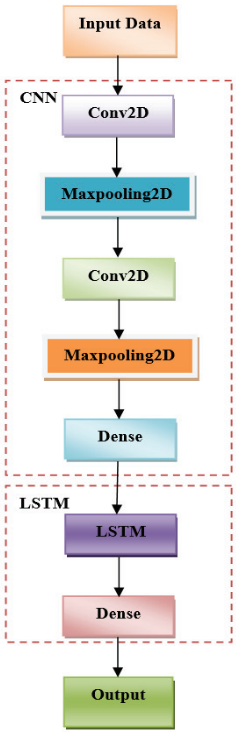

Fig. 3. The Proposed CNN-LSTM Model Structure

units, the loss, defined as the discrepancy between the predicted and actual output values, is employed. The optimisation procedure employs the RMSprop gradient descent technique, frequently utilised for weight optimisation in DNNs.

IV. Result and Discussion

This study delineates some classroom innovations implemented by diverse college educators within the setting. According to Weimer's working theory, university and college classes are highly instructor-centered, which hinders students' ability to become independent, successful learners. She lays out the research and gives examples of several ways to build student-centered classrooms for each area. An expanding number of schools and institutions are providing online resources to assist educators in promoting student-centered learning.

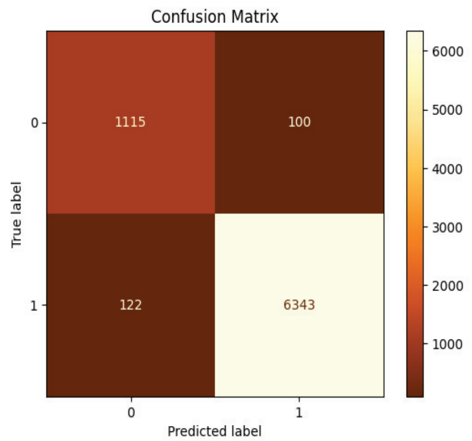

Fig. 4. Confusion Matrix

Figure 4 provides a clear and succinct presentation of the model's performance on the train set. In Figure 4, they can see the total number of correct, incorrect, false positive, and false negative predictions for each group.

TABLE I. PERFORMANCE COMPARISON (%)

Model	Variance	R^2	MAE
CNN	0.678	0.513	2.876
CNN-LSTM	**0.618**	**0.567**	**1.114**
LSTM	0.714	0.471	2.543

Table 1 shows the results of comparing the performance of different models that attempt to teacher and student Activity data.

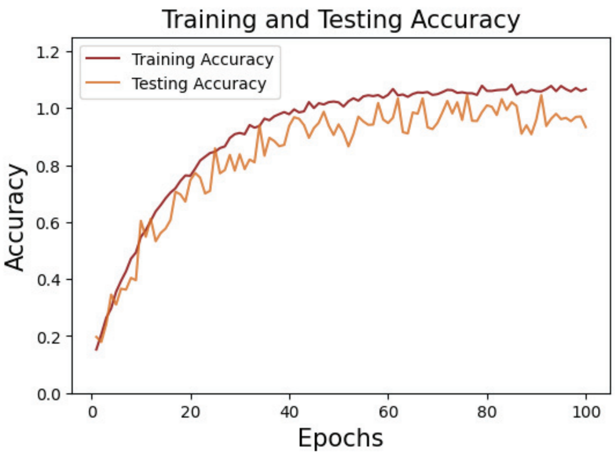

Fig. 5. Training and Testing Accuracy

As shown in Figure 5, our CNN-LSTM approach achieves satisfactory accuracy. With 100 epochs under our belts, they can confidently say that our training data yields an approximate accuracy of 97.21%.

TABLE II. THE MODEL PREDICTION VALUE(%)

Model	Prediction Value		
	MAE	RMSE	MAPE
CNN	227.32	284.63	2.56
CNN-LSTM	**203.64**	**241.22**	**1.78**
LSTM	348.50	312.18	3.64

Each performance parameter for value prediction has an average result, which are shown in Table 2. When it comes to value prediction, CNN-LSTM clearly stands out.

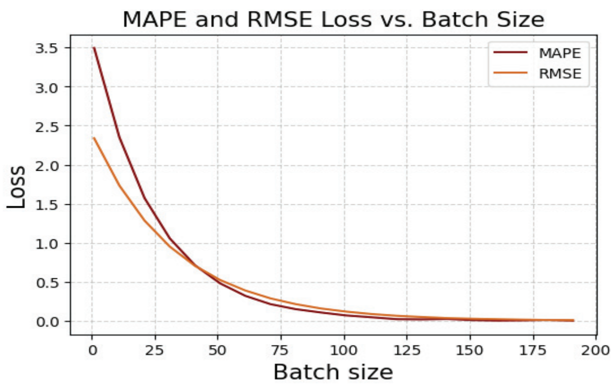

Fig. 6. MAPE and RMSE Loss Vs Batch Size

As demonstrated in Figure 6, this can be accomplished by taking the prediction loss into account. The prediction Loss drops as batch size rises, as is clearly seen.

TABLE III. MODEL TRAINING TIME ANALYSIS

Model	Spend Time	Time
CNN	174.32	151
CNN-LSTM	**118.09**	**115**
LSTM	151.41	134

The CNN-LSTM model requires 118.09s, as shown in Table 3. After being finished 115 times, it converged. In terms of convergence characteristics and training duration, the CNN-Bi-LSTM model performed admirably.

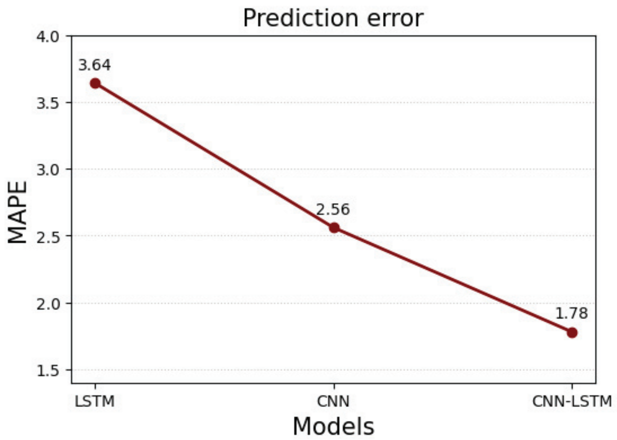

Fig. 7. MAPE for the Proposed Model

Figure 7 displays the results of our comparisons with different models and the CNN-LSTM model: According to MAPE 1.78, our CNN-LSTM method outperforms the other methods in terms of prediction error.

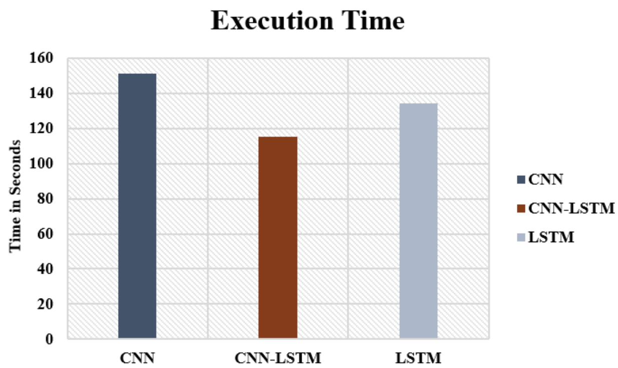

Fig. 8. Execution Time

Figure 8 shows a comparison of the CNN and LSTM classifiers' execution times. Compared to CNN, LSTM classifier, CNN-LSTM classifier has a shorter execution time, according to the analysis.

V. CONCLUSION

Everyone in the classroom, from professors to students, has their own set of views on a wide variety of issues, including but not limited to politics, health, religion, and education. Research on educational beliefs has focused on how they play out in real-world classrooms. Numerous studies in the field of education have examined the dissemination and implementation of such views in recent years, with a number of these studies focussing on instructors' beliefs rather than students. Min-max normalisation is an effective technique for preparing data in feature selection. A dimensionality reduction method known as PCA was employed to decrease the quantity of input and output variables. The integration of CNNs and LSTMs enables the system to leverage the capabilities of both networks: CNNs excel at extracting hierarchical feature representations from intricate data, while LSTMs are adept at capturing long-term dependencies in temporal information. In comparison to alternative models, the CNN-LSTM demonstrates exceptional performance with a MAE of 203.64%, a RMSE of 241.22%, and a MAPE of 1.78%.

REFERENCES

[1] E. Fischer and M. Hänze, “Back from ‘guide on the side’ to ‘sage on the stage’? Effects of teacher-guided and student-activating teaching methods on student learning in higher education,” *Int. J. Educ. Res.*, vol. 95, no. September 2018, pp. 26–35, 2019, doi: 10.1016/j.ijer.2019.03.001.

[2] R. Tormey, “Rethinking student-teacher relationships in higher education: a multidimensional approach,” *High. Educ.*, vol. 82, no. 5, pp. 993–1011, Nov. 2021, doi: 10.1007/s10734-021-00711-w.

[3] M. Northcote, “Educational beliefs of higher education teachers and students: Implications for teacher education,” *Aust. J. Teach. Educ.*, vol. 34, no. 3, pp. 69–81, 2009, doi: 10.14221/ajte.2009v34n3.3.

[4] M. Kuleto *et al.*, “Explorando oportunidades y desafíos de la inteligencia artificial y el aprendizaje automático en instituciones de educación superior,” *Sustain.*, vol. 13, no. 18, pp. 1–16, 2021, [Online]. Available: https://www.mdpi.com/2071-1050/13/18/10424

[5] S. Wang and D. Zhang, “Student-centred teaching, deep learning and self-reported ability improvement in higher education: Evidence from Mainland China,” *Innov. Educ. Teach. Int.*, vol. 56, no. 5, pp. 581–593, Sep. 2019, doi: 10.1080/14703297.2018.1490662.

[6] A. Kumar, S. R. Sangwan, and A. Nayyar, *Multimedia Big Data Computing for IoT Applications*, vol. 163, no. July 2019. 2020. [Online]. Available: http://link.springer.com/10.1007/978-981-13-8759-3

[7] G. Radić, “E-learning Experiences of higher education institutions in the Republic of Serbia during COVID-19, content analysis and case study ITS Belgrade. Sustainable Economic Development and Advancing Education Excellence in the era of Global Pandemic.",” *Proc. 36th Int. Bus. Inf. Manag.*, no. February, 2021.

[8] P. Shah *et al.*, “Artificial intelligence and machine learning in clinical development: a translational perspective,” *npj Digit. Med.*, vol. 2, no. 1, p. 69, Jul. 2019, doi: 10.1038/s41746-019-0148-3.

[9] O. Awogbemi, D. V. Von Kallon, and K. S. Kumar, “Contributions of artificial intelligence and digitization in achieving clean and affordable energy,” *Intell. Syst. with Appl.*, vol. 22, no. April, p. 200389, Jun. 2024, doi: 10.1016/j.iswa.2024.200389.

[10] M. Maleki, P. K. H. Anvari, Enayat, Z. Noorimotlagh, and S. A. Mirzaee, “Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ’ s public news and information.,” *Psychiatry Res.*, vol. 14(4), no. January, p. 293, 2020.

[11] S. Pokhrel and R. Chhetri, “A Literature Review on Impact of COVID-19 Pandemic on Teaching and Learning,” *High. Educ. Futur.*, vol. 8, no. 1, pp. 133–141, Jan. 2021, doi: 10.1177/2347631120983481.

[12] A. Badia, C. Garcia, and J. Meneses, “Approaches to teaching online: Exploring factors influencing teachers in a fully online university,” *Br. J. Educ. Technol.*, vol. 48, no. 6, pp. 1193–1207, Nov. 2017, doi: 10.1111/bjet.12475.

[13] I. Aljarah, “Students’ Academic Performance Dataset,” Kaggle. Accessed: Mar. 07, 2025. [Online]. Available: https://www.kaggle.com/datasets/aljarah/xAPI-Edu-Data

[14] V. Hegde; P. P. Prageeth, “Higher Education Student Dropout Prediction and Analysis through Educational Data Mining - ICISC 2018 : proceedings of the 2nd International Conference on Inventive Systems and Control (ICISC 2018) : 19-20 January 2018,” *2018 2nd Int. Conf. Inven. Syst. Control*, no. Icisc, pp. 694–699, 2018.

[15] O. Iatrellis, I. Savvas, P. Fitsilis, and V. C. Gerogiannis, “A two-phase machine learning approach for predicting student outcomes,” *Educ. Inf. Technol.*, vol. 26, no. 1, pp. 69–88, 2021, doi: 10.1007/s10639-020-10260-x.

[16] L. Chen and G. Shen, “Research of higher education input-output efficiency based on principal component analysis,” *2011 Int. Conf. Manag. Sci. Ind. Eng. MSIE 2011*, pp. 174–177, 2011, doi: 10.1109/MSIE.2011.5707690.

[17] J. Tian, A. Xiang, Y. Feng, Q. Yang, and H. Liu, “Enhancing Disease Prediction with a Hybrid CNN-LSTM Framework in EHRs,” *J. Theory Pract. Eng. Sci.*, vol. 4, no. 02, pp. 8–14, 2024, doi: 10.53469/jtpes.2024.04(02).02.

[18] Y. Li and W. Dai, “Bitcoin price forecasting method based on CNN-LSTM hybrid neural network model,” *J. Eng.*, vol. 2020, no. 13, pp. 344–347, 2020, doi: 10.1049/joe.2019.1203.