Analyzing Student's Emotions in the Classroom: A Deep Learning Approach to Facial Expression Recognition

Syed Muhammad Daniyal[1], Usama Amjad[1], Abdul Khaliq[2], Noman Bin Zahid[1], Faiza Latif Abbasi[1], and Syed Muhammad Tahir Hussain[3]

[1]Faculty of Engineering Science and Technology, Iqra University Karachi, Pakistan, [2]Faculty of Computer Science, Institute of Business Management Karachi, Pakistan, [3]Department of Engineering Science and Technology, Iqra University Karachi, Pakistan

Email: syed.daniyal@iqra.edu.pk, usama.amjad@iqra.edu.pk, khaliq@iobm.edu.pk, noman.zahid@iqra.edu.pk, faiza.latif@iqra.edu.pk, syed.57774@iqra.edu.pk

Corresponding Author: Syed Muhammad Daniyal, syed.daniyal@iqra.edu.pk

Received: 15-02-2024; **Accepted**: 18-05-2024; **Published**: 01-07-2024

Abstract: Student success is a crucial metric for educational institutions, as it affects not only the individual student's academic and personal growth but also the institution's reputation, funding, and overall educational outcomes. Student failure is a common and concerning issue in educational institutions, often leading to significant academic and personal consequences. This research explores the potential of deep learning techniques to accurately predict student emotions and prognostication against failure, enabling proactive interventions. Prior studies in this field have mostly depended on conventional machine learning techniques to predict student performance and detect student risk, such as logistic regression and linear support vector machines. However, advanced techniques are frequently needed to capture the underlying patterns and correlations due to the complexity and richness of student behavior and performance. Within the field of artificial intelligence, deep learning has proven to be extremely effective in some applications, such as computer vision, generative AI, and natural language processing. Toward this goal, this research offers an extensive deep learning architecture that utilizes the vast data from the class environment to forecast student emotions more precisely and finely. For evaluating the above facts, the deep learning model CNN-LSTM is used to reveal an effective participation evaluation of students' attention and involvement during classroom sessions. By creating and assessing cutting-edge AI models that improve the precision, flexibility, and real-time capabilities of teaching aids, this research seeks to overcome these constraints. Furthermore, the proposed model demonstrates an accuracy of 91% making it a highly useable real-time application.

Keywords: Deep Learning, Face Detection, Face Recognition, Emotion Recognition

1. Introduction

The advancement of a country is significantly influenced by education. Additionally, it's an essential instrument for success in life. Every educational establishment strives to give its pupils quality education to enhance learning [1]. The academic performance of students is a crucial component that impacts the accomplishment of any educational institution, because of this, forecasting students' performance is a crucial area of study for using educational data, a topic that many scholars [2]. While several studies have examined the application of deep learning models with different log sequence embedding strategies, their evaluation of the three primary types of networks RNN CNN, and transformer combined with distinct embedding strategies has been restricted. For example, two studies [3] covered transformer-based models but not CNN-based models. Another [4] introduced the facial action coding system (FACS) model, which measures and evaluates human face behavior. Six fundamental emotions were distinguished from the facial expressions: happiness, sadness, rage, contempt, surprise, and fear as shown in Figure 1. The suggested system's effectiveness on face-based presence may be measured and compared tangibly by using the LFW data set "Labelled

Faces in the Wild," a facial photo collection created to address facial recognition issues [5]. In light of this circumstance, technological education is required to support the presence process, one example of which is the documentation of artificial intelligence [6]. Information obtained from the attendX scanning model attendance procedure, in conjunction with apps for data processing, is more precise and tangible than that obtained from other approaches [7]. This section provides a quick summary of research on academic procrastination and how it affects students' performance. Procrastination is a common behavioral feature that most people exhibit; they tend to put off chores until the last minute. Procrastination is the compulsive behavior of putting off chores until they cause discomfort [8]. In general, people begin their duties with no idea of whether they will be completed or not. Though they eventually give up and fail to finish the assignment, they believe it can wait.

Figure 1. Emotions Recognition of Students in the Classroom

The expansion of online and blended learning opportunities has brought to the fore the growing challenge of student dropout and academic underperformance [9]. It has become very critical for teachers to predict which student is dropping interest so that teacher can build a supportive environment for students to regain their interest. Many recent studies of machine learning have shown techniques to predict which student is dropping interest. These advanced models have shown promising results but have many limitations. Especially, deep learning models such as LSTM-CNN have shown amazing results in the form of performance and student engagement as shown in Figure 2.

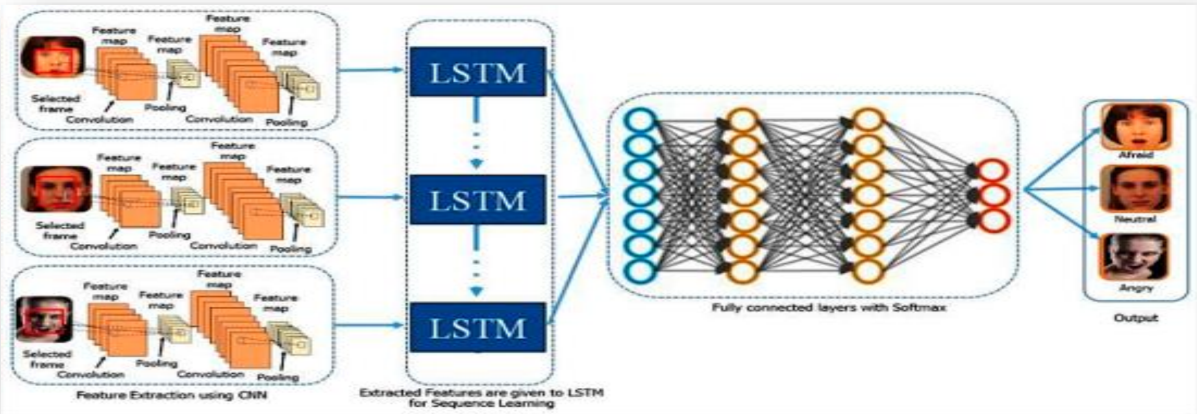

Figure 2. Architecture of LSTM-CNN

This research aimed to develop a model using a deep learning model that could target the challenges by increasing the accuracy of recognizing the student's emotions within the classroom by integrating of metrics and providing real-time feedback. This research uses the approach of CNN-LSTM with FER2013 dataset. The output of the layers are concatenated to utilize the advantages of both models, followed by dense layers and a softmax output for precise multi-class emotion. The combination of these models enhances the accuracy of and provides the real-time feedback.

2. Literature Review

This literature review shows advanced deep-learning models to trace student engagement in learning environments, identify the emotions of students, and predict student's execution in the classroom. Many researchers have used models including sequential data analysis, attendance monitoring, and facial expression recognition to improve learning results. One method marks the difficulties associated with processing sequential data by predicting student performance using recurrent neural networks (RNNs) and long short-term memory (LSTM) networks [10]. Convolutional neural networks (CNNs) and long short-term memory (LSTMs) are used together that show impressive results for institute accomplishment prediction and improve flexibility across various educational institutes.

Other researchers have shown their work, where their models have improved accuracy by integrating progressive fine-tuning, and feature additive pooling [11]. However, face recognition models for attendance tracking have been created, providing instant feedback. Novel deep learning model pairings, such as generative adversarial networks (GANs) and support vector machines (SVMs), are utilized in performance prediction to improve accuracy and solve problems like data insufficiency. Furthermore, AI-based real-time behavior monitoring tools have been used, giving teachers on spot feedback on how they were engaging their students. For real-time attention monitoring of students, other research has combined CNNs, RNNs, and attention mechanisms, offering a thorough picture of engagement levels [12]. Altogether, these techniques bring novel methods for deep learning, hybrid models, and real-time monitoring systems to this specific area. The paper's literature fills a major gap in the field by providing workable methods to maximize learning outcomes and student engagement. These researches push the boundaries of artificial intelligence and deep learning in education by addressing both traditional and virtual learning contexts.

The model that is best for improvement in learning and student engagement will mostly be accurate in different environments in which it is used. Every model has different advantages and disadvantages involving student engagement and observation [13]. A comparative analysis of different factors including precision, flexibility, and instantaneous response shows its impact on the learning environment. Combining different algorithms LSTM-CNN model performs well with sequential data. This sequence shows a balanced approach between performance and interpretability, and it mostly works well for personalized teaching systems. Its power lies in its ability to manage difficult tasks while analyzing student's performance [14].

Combining statistical and machine learning approaches with both sequential and non-sequential data, this approach also focuses on performance prediction. It is a variety of learning situations because it has the highest degree of understandability and low computational necessity. Its logical approach, which keeps a strong prediction framework, is helpful in situations where grasping the elements motivating student performance is critical [15]. In a clash with this, non-sequential data is the main emphasis of models like face expression recognition, which analyzes student emotions. Deep learning is used in this approach, which leads to good performance but poor interpretability. The task's complexity and significant processing demands might stop it from being used in environments lacking new approach. However, there is a lot of capability for real-time student emotion monitoring because of how accurate face analysis is at identifying emotions.

One less complex system for recording attendance, called AttendX, is outstanding for its ease of interpretation and minimal processing resource needs. This technique is perfect for automating attendance with no overhead because it uses fundamental machine-learning techniques. Its practicality and ease of use make it a good contender for daily use in educational institutions, even though it might not be able to tackle complex tasks [16]. When it comes to identifying academic performance, framework that integrates machine learning and other data sources provide a more overarching view. These models are proper for analyzing and predicting academic results across many different students because they strike a balance between interpretability and performance. When multiple characteristics are integrated to provide a more exact prediction of student results, these models perform especially well. Using deep learning, the real-time behavior monitoring model analyses student behavior in real time, employing a novel strategy [17].

At last, to analyze students' attention in real time, the attention monitoring system uses hybrid models that combine convolutional neural network, recurrent neural network, and attention mechanisms. This technique strikes a balance between interpretability and performance, requiring lower to higher amount of computing power. This technique is especially useful for giving teachers real-time analysis data about student engagement, enabling them to better optimize learning environments. The best model to use will depend on the particular requirements of the learning environment.

3. Methodology

The research uses a combination of methods, integrating design-based and experimental techniques to create, use, and assess deep learning models that track behavioral patterns and predict student outcomes. The main objective is to use deep learning, specifically a hybrid LSTM-CNN model, to predict student failure. The requirements for high accuracy, the intricacy of the prediction task, and the properties of the data all drove this decision. The architecture of the proposed model is shown in Figure 3.

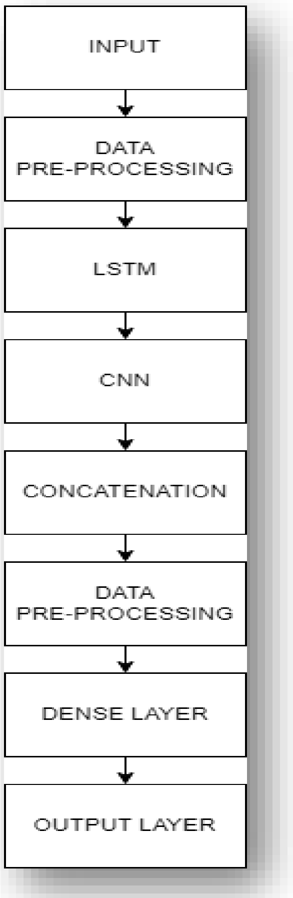

Figure 3. Architecture of the Proposed Model

3.1.Data Pre-processing

This research will make use of the FER2013 dataset [18]. This dataset contains 35,887 48x48 pixel greyscales which are tagged with one of seven emotion categories: surprise, rage, contempt, fear, happiness, sorrow, or neutral. There are training and testing sets in the dataset. The FER2013 dataset was chosen due to its wide usage in emotion recognition applications, providing benchmark dataset. Additionally, many other datasets shows bias towards particular emotions but FER2013 dataset provides a fair representation of emotion categories. Due to this FER2013 dataset is more suited for creating an accurate and broadly applicable emotion detection system for educational environment.

3.2. LSTM Layers

The long-term dependencies between imagined pixel sequences will be learned using LSTM units [19]. We will experiment with the number of units and layers as we fine-tune these layers to maximize performance. The mathematical presentation of this layer is given below:

3.3. CNN Layers

From the input images, convolutional layers will extract local spatial information including edges, textures, and face expression-related patterns. To capture various degrees of abstraction, several convolutional layers with variable filter sizes and pooling techniques will be employed.

3.4. Concatenation

Concatenating the outputs from the LSTM and CNN layers will enable the model to incorporate sequential and spatial information for more precise emotion recognition.

3.5. Dense Layer

Following the concatenation, fully connected layers will be added to the learned features to convert them into a higher-level representation and get them ready for the final classification.

3.6. Output Layer

A softmax activation function will be utilized for multi-class classification, and the output layer will have seven units total one for each type of emotion.

4. Experimental Design

4.1. Training the Model

This research will make use of the FER2013 dataset which consists of 35,887 48x48 pixel greyscale images. This dataset is tagged with one of seven emotion categories: surprise, rage, contempt, fear, happiness, sorrow, or neutral as shown in figure 4. There are training and testing sets in the dataset.

Figure 4. FER2013 Dataset

4.2. Parameter Settings

Managing Missing or Inaccurate Data: This research will search the dataset for any missing or inaccurate labels. These will be changed or eliminated if needed.

Normalization: To guarantee that the model receives consistent input values, all picture pixel values will be normalized to a range between 0 and 1.

Data Augmentation: The training set will be subjected to data augmentation methods like rotation, zooming, and horizontal flipping to enhance model generalization.

Resizing and Padding: Images will be buffered or scaled to guarantee consistent proportions before being sent into the CNN and LSTM layers.

Sequential Features: The LSTM component will be able to capture temporal relationships between pixel intensities by considering the picture pixel sequences as a time series, so leveraging the time-series structure of the dataset.

Non-Sequential Features: Although the main emphasis is on picture data, if accessible, further metadata, such as age, gender, or contextual characteristics, may be included. The focus of this investigation is still on the pixel-based features that were taken from the FER2013 dataset.

Loss Function: The difference between the actual and anticipated emotion categories will be quantified using categorical cross-entropy.

Optimizer: For effective gradient-based learning with an adaptable learning rate, the Adam optimiser will be used.

Evaluation Metrics: To determine how effectively the model predicts each emotion category, performance will be measured using accuracy, precision, recall, F1-score, and confusion matrices.

5. Results And Discussion

5.1. Analysis of Results

With an outstanding detection rate of 91%, the proposed deep learning model showed great performance in identifying emotions from facial expressions as shown in Figure 5. Its robustness under challenging conditions, such as shifting illuminations and variations in facial angles, further highlights its ability. This adaptability stems from the combined stremgths of CNN for spatial feature extraction and LSTM for temporal pattern recognition. Because of its high accuracy, the model is a good fit for real-time emotion detection tasks because it can distinguish between the different emotions present in the dataset. All over the training and testing stages, the model's loss rate was minimal as shown in Figure 6, indicating that it covered effectively and did not experience overfitting or underfitting. The low loss also suggested that there were few prediction errors and that the model's predictions were generally in line with the real emotion labels.

These outcomes demonstrate how well the model captures the facial expressions' temporal and spatial angles. The model's remarkable performance across all emotion categories can be credited to its ability to identify subtle emotion-related information through the incorporation of deep learning techniques. This model's lower failure rate and excellent accuracy make it a dependable resource for tasks catering to the recognition of emotions, especially in educational settings where the cognition of emotional states might be essential for psychological and instructional handling.

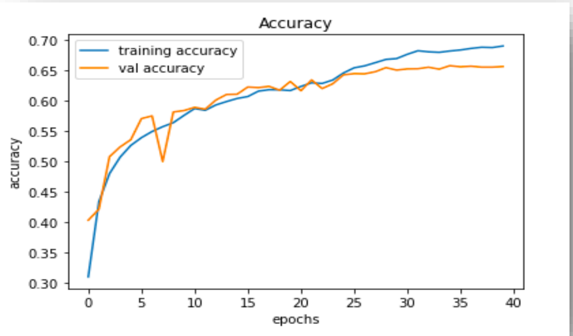

Figure 5. Accuracy of the Proposed Model

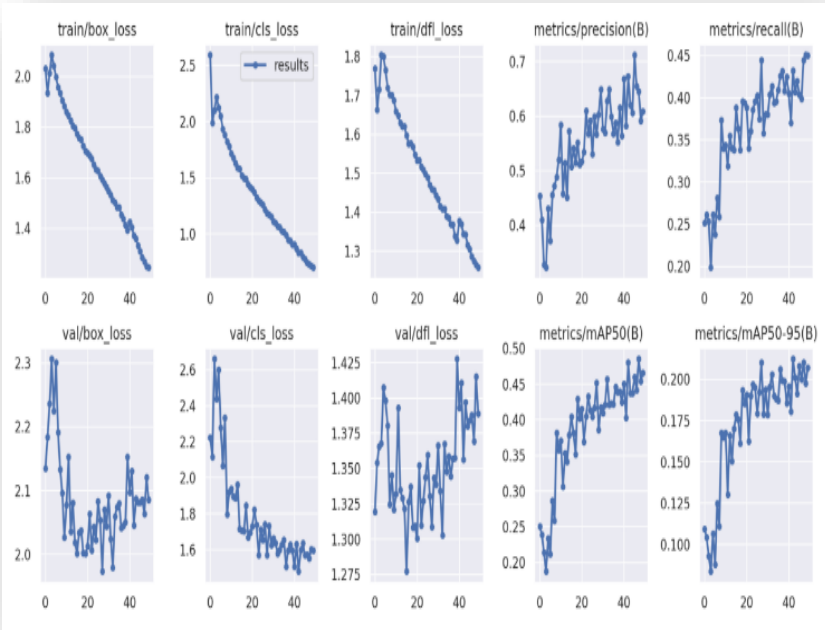

Figure 6. Loss of the Proposed Model

6. Comparison Of Results

The outcomes are compared with several LSTM-CNN models that have been proposed by other researchers in Table 1. The proposed model achieved accuracy of 91%, outperforming the model that show lower accuracy due to overfitting and sensitivity to position fluctuations.

Hence, overfitting is the major concern of these models below and that is the reason of their low accuracy. However, in comparison of the proposed model with all previous models, our proposed model outperforms well with a higher accuracy rate in every environment.

Table 1. Comparison of Proposed Model with Different Models

Model	*Accuracy*
LSTM-CNN [20]	88%
LSTM-CNN [21]	85%
LSTM-CNN [22]	82%
Proposed Model	91%

7. Conclusion

In conclusion, the accuracy would be highlighted as the proposed model has achieved an accuracy of 91%, and it accurately predicts student's emotions and expressions. The accuracy shows that the combination of LSTM-CNN is the best fit for solving this problem. Additionally, examining its performance with different datasets and exploring cross-model emotion recognition could further enhance its accuracy and reliability.

References

[1] Iyer, A., Das, S. S., Teotia, R., Maheshwari, S., & Sharma, R. R. (2023). CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings. Multimedia Tools and Applications, 82(4), 4883-4896.

[2] Ramzan, M., & Dawn, S. (2023). Fused CNN-7LSTM deep learning emotion recognition model using electroencephalography signals. International Journal of Neuroscience, 133(6), 587-597.

[3] Febrian, R., Halim, B. M., Christina, M., Ramdhan, D., & Chowanda, A. (2023). Facial expression recognition using bidirectional LSTM-CNN. Procedia Comput. Sci, 216(2022), 39-47.

[4] Mellouk, W., & Handouzi, W. (2023). CNN-LSTM for automatic emotion recognition using contactless photoplythesmographic signals. Biomedical Signal Processing and Control, 85, 104907.

[5] Lasri, I., Riadsolh, A., & Elbelkacemi, M. (2023). Facial emotion recognition of deaf and hard-of-hearing students for engagement detection using deep learning. Education and Information Technologies, 28(4), 4069-4092.

[6] Singh, R., Saurav, S., Kumar, T., Saini, R., Vohra, A., & Singh, S. (2023). Facial expression recognition in videos using hybrid CNN & ConvLSTM. International Journal of Information Technology, 15(4), 1819-1830.

[7] Mohana, M., Subashini, P., & Krishnaveni, M. (2023). Emotion recognition from facial expression using hybrid CNN–LSTM network. International Journal of Pattern Recognition and Artificial Intelligence, 37(08), 2356008.

[8] Daniyal, S. M., Masood, A., Ebrahim, M., Adil, S. H., & Raza, K. (2024). An Improved Face Recognition Method Based on Convolutional Neural Network. Journal of Independent Studies and Research Computing, 22(1), 103-110.

[9] Abbasi, M. M., Daniyal, S. M., Abro, A. A., Hussain, D., Amjad, U., & Zahid, N. B. (2024). Applying Neural Networks to Predict Ventilator Demand: A Study of Pakistan's Healthcare Sector. VFAST Transactions on Software Engineering, 12(3), 217-229.

[10] Irshad, M. H., Ebrahim, M., Abro, A. A., Raza, K., & Adil, S. H. (2024). Detecting Shadows in Computer Vision: A MATLAB-Based Approach. The Asian Bulletin of Big Data Management, 4(1), Science-4.

[11] Abro, A. A., Sıddıque, W. A., Talpur, M. S. H., Jumani, A. K., & Yaşar, E. (2023). A combined approach of base and meta learners for hybrid system. Turkish Journal of Engineering, 7(1), 25-32.

[12] Chen, Z., Liang, M., Xue, Z., & Yu, W. (2023). STRAN: Student expression recognition based on spatio-temporal residual attention network in classroom teaching videos. Applied Intelligence, 53(21), 25310-25329.

[13] Jin, H., Du, R., Wen, T., Zhao, J., Shi, L., & Zhang, S. (2023). A classroom facial expression recognition method based on attention mechanism. Journal of Intelligent & Fuzzy Systems, (Preprint), 1-10.

[14] Daniyal, S. M., Abbasi, M. M., Hussain, D., Amjad, U., Abro, A. B., & Naeem, M. (2024). A hybrid approach for simultaneous effective automobile navigation with DE and PSO. VAWKUM Transactions on Computer Sciences, 12(2), 01-15.

[15] Ikram, S., Ahmad, H., Mahmood, N., Faisal, C. N., Abbas, Q., Qureshi, I., & Hussain, A. (2023). Recognition of student engagement state in a classroom environment using deep and efficient transfer learning algorithm. Applied Sciences, 13(15), 8637.

[16] Mohana, M., Subashini, P., & Krishnaveni, M. (2023). Emotion recognition from facial expression using hybrid CNN–LSTM network. International Journal of Pattern Recognition and Artificial Intelligence, 37(08), 2356008.

[17] Kaewkaisorn, K., Pintong, K., Bunyang, S., Tansawat, T., & Siriborvornratanakul, T. (2024). Student attentiveness analysis in virtual classroom using distraction, drowsiness and emotion detection. Discover Education, 3(1), 1-14.

[18] Daniyal, S. M., Hussain, S. M. T., Abbasi, F. L., Hussain, D., Abbasi, M. M., & Amjad, U. (2024). A Hybrid Deep Learning Model for Precise Epilepsy Detection and Seizure Prediction. Spectrum of engineering sciences, 2(3), 62-77.

[19] Khabbachi, I., Mdaghri-Alaoui, G., Zouhair, A., & Mahboub, A. (2024, July). A cheating detection system in online exams through real-time facial emotion recognition of students. In 2024 International Conference on Computing, Internet of Things and Microwave Systems (ICCIMS) (pp. 1-5). IEEE.

[20] Rathi, S., Oswal, S., & Ahuja, A. (2023, May). A Coherent Way of Detecting Learner's Academic Emotions via Live Camera Using CNN and Deep LSTM. In 2023 4th International Conference for Emerging Technology (INCET) (pp. 1-6). IEEE.

[21] Li, S. (2024). Application of CNN based Regognition Technology in Public English Teaching in College Intelligent Classroom. Scalable Computing: Practice and Experience, 25(4), 2345-2357.

[22] Nuha, N. S., Mahmud, T., Rezaoana, N., Hossain, M. S., & Andersson, K. (2023, November). An approach of analyzing classroom student engagement in multimodal environment by using deep learning. In 2023 IEEE 9th International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE) (pp. 286-291). IEEE.

.