ASEC 2024 Conference
The 5th International Electronic Conference on Applied Sciences
04–06 December 2024 | Online
MDPI

Detection of Students' Emotions in an Online Learning Environment Using CNN-LSTM Model

Bilkisu Muhammad Bashir*, Dr. Hadiza Ali Umar
Department Of Computer Science, Bayero University, kano

INTRODUCTION & AIM

Emotion recognition is crucial for understanding human interactions and enhancing communication. In online learning, recognizing emotions like interest, boredom, and confusion is essential for improving student engagement and efficiency. Traditional facial emotion recognition (FER) systems, which focus on basic emotions like happiness and sadness, fail to capture learning-specific emotions. Additionally, convolutional neural networks (CNNs) struggle to track temporal changes in emotional expressions.This study addresses these challenges by developing a hybrid CNN-LSTM model that combines the spatial feature extraction of CNNs with the temporal dynamics captured by Long Short-Term Memory (LSTM) networks. The proposed model aims to provide real-time, accurate detection of learning-specific emotions, enabling personalized and adaptive online education.

METHOD

This study utilizes a hybrid CNN-LSTM approach tailored for emotion recognition in online learning. The methodology involves the use of three datasets: FER2013, containing 28,709 training and 3,589 validation images across seven emotion categories; CK+48, which includes 750 images labeled with five basic emotions; and JAFFE, featuring 213 grayscale images with seven emotion categories. During preprocessing, images were resized to 48×48 pixels and normalized, and Action Units (AUs) were identified to filter images relevant to the target emotions of interest.

The model architecture integrates CNN layers to extract spatial features from facial images and LSTM layers to capture temporal dependencies in the sequence of facial expressions. The final output layer classifies emotions into three categories: interest, boredom, and confusion. The model was trained on filtered datasets with training-validation splits, and its performance was evaluated using metrics such as accuracy, precision, recall, and F1-score.

RESULTS & DISCUSSION

Table 4.3: Performance Matric Comparison

Algorithm	Dataset	Accuracy	precision	Recall	F1-score
CNN model	CK+	0.96	0.94	0.96	0.95
CNN-LSTM(Proposed Model)	**CK+**	**0.99**	**0.99**	**0.97**	**0.98**
CNN model	FER 2013	0.97%	0.94%	0.96%	0.95%
CNN-LSTM(Proposed Model)	**FER 2013**	**0.98**	**0.97**	**0.98**	**0.98**

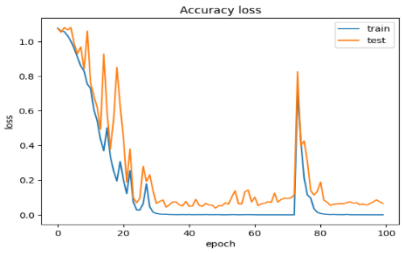

Figure 4.3: Training and validation of loss function for FER 2013 dataset.

The CNN-LSTM model demonstrated exceptional performance with an accuracy of 98%, precision of 97%, recall of 98%, and an F1-score of 98%, outperforming standalone CNN models by effectively combining spatial and temporal feature analysis. Its effectiveness was validated across multiple datasets, ensuring adaptability to diverse learning scenarios. Compared to traditional methods, the hybrid model excelled in detecting learning-specific emotions, particularly in contexts with significant temporal dynamics. These results highlight the model's potential for real-time application in online learning, enabling personalized teaching strategies and enhanced student engagement.

CONCLUSION

The study successfully developed a CNN-LSTM-based model for detecting learning-specific emotions in online education environments. By combining CNN's spatial feature extraction and LSTM's temporal analysis capabilities, the model effectively recognizes complex emotional states like interest, boredom, and confusion. The proposed approach addresses limitations of traditional FER systems and contributes significantly to the field of affective computing in education.

FUTURE WORK / REFERENCES

Future research can explore:

1. Expanding the dataset to include more diverse demographics and real-world data.
2. Incorporating multimodal inputs, such as audio cues or physiological signals, for comprehensive emotion recognition.
3. Enhancing the model's real-time processing capabilities for live online educational scenarios.
4. Addressing ethical considerations, such as privacy concerns, for wider adoption in educational platforms.

REFERENCES

Abdulhussein, W. R., El Abbadi, N. K., & Gaber, A. M. (2021). Hybrid Deep Neural Network for Facial Expressions Recognition. *Indonesian Journal of Electrical Engineering and Informatics*, *9*(4), 993–1007. https://doi.org/10.52549/ijeei.v9i4.3425

Abdullah, S. M. S. A., Ameen, S. Y. A., M. Sadeeq, M. A., & Zeebaree, S. (2021). Multimodal Emotion Recognition using Deep Learning. *Journal of Applied Science and Technology*