VIDWAN Journal of Science, Technology and Innovation

OPEN ACCESS

PEER-REVIEWED - OPEN ACCESS Vol. 01 · Issue 01 · 2026 · pp. 20-24 | ISSN: APPLIED (Online) | INAGURAL EDITION

A Long-Range Attention Framework for Deepfake Detection Using Deep Learning

Dr. K. Chandramouli, *Department of EEE*
Vaageswari College Of Engineering, Karimangar

ABSTRACT

The complexity of the methods used to create deepfakes is the reason why digital media authentication is requiring an extended period of time. Normal methods are incapable of detecting minute variations in texture, pace, and facial expressions. This research demonstrates a novel approach to identifying deepfakes through the use of long-distance attention. Time and location data that are stored in video frames are the primary focus. The model is trained to identify discrepancies in time and space through the utilization of two modules in the proposed design. Consequently, the model is more adaptable and capable of accommodating minor modifications. The network is able to identify and classify false positives and declines by utilizing long-distance attention, which connects data at the frame level. Celeb-DF, DFDC, and Face Forensics++ are benchmark datasets that demonstrate the effectiveness of these methods in comparison to CNN and LSTM. In the long term, this technology has the potential to supplant scalable, precise deepfake detection methods.

Keywords: *Deepfake Detection; Video Forgery; Spatial-Temporal Model; Long-Distance Attention Mechanism; Binary Classification; Fine-Grained Classification.*

Article history: Received 8 July 2026 · Revised 18 July 2026 · Accepted 28 July 2026 · Published 8 August 2026

1. INTRODUCTION

Recent advancements in deepfake technology have enabled the manipulation of digital images and videos to create effects that resemble genuine photographs. The internet has been significantly impacted by this. Accessing movies, training with them, and improving them could be facilitated by the use of deepfakes. Additionally, they have been employed by individuals to disseminate false information, misrepresent themselves as celebrities, and deceive others. These fake media products, which primarily feature renowned individuals, have resulted in a decrease in trust in the internet and material that is not authentic.

The qualities of actors have been replaced by new advancements in deepfake technology, which has resulted in movies that appear incredibly realistic. Numerous individuals are concerned about the potential for deepfake movies, which are produced using intricate generative algorithms and resemble genuine footage, to be exploited. These films have a detrimental impact on society and individuals by disseminating false information, hate speech, and misunderstandings due to their ease of production and distribution. This technology is being utilized extensively by individuals today, which poses a threat to public safety and security conditions.

The model was constructed using video and image fragments. In order to guarantee the precision of the results, the image collection was meticulously divided into testing, validation, and training groups. In conclusion. The model can easily manage both static and dynamic inputs. The accuracy rate is 97% when images are utilized as input, and 85% when videos are employed. Attention-based learning and spatial-temporal modeling are employed in this investigation to demonstrate a scalable and dependable method for identifying altered data and establishing secure digital environments.

2. LITERATURE SURVEY

Henderson and Prakash (2021) I recommend employing long-range attention models to gather features for deepfake detection. In order to identify deepfake material, attention-based neural networks and visitors (ViT) monitor changes in facial features, video frames, and time. Detecting the spatial and contextual connections in

multimedia content is facilitated. It has been discovered that the new convolutional algorithms are more effective than the previous ones at classifying items and identifying false positives. This novel approach enhances the quality of digital content verification and advanced multimedia forensics.

Moreno and Iyer (2022) demonstrate a state-of-the-art method for identifying deepfakes that employs transformer-based long-range attention and deep neural networks. Bidirectional encoder representations from transformers (BERT) and self-attention models are employed to identify data that has been distorted by identifying patterns of audiovisual issues, such as lip and facial problems. The processing of video feeds enhances the accuracy of investigations and expedites the resolution of crimes. Performance analysis demonstrates that it is feasible to optimize the management and planning of substantial multimedia datasets. The system is equipped with robust anti-fake applications.

Chen and Wallace (2023) In order to identify deepfakes, CNN, LSTM, and long-range attention layers were implemented. This method can distinguish between genuine and counterfeit films by examining factors such as facial texture, eye movement patterns, and correlations between frames. Analytics stored in the cloud enhance adaptive detection by continuously processing digital media assets. The algorithm is highly effective in identifying complex video sets and managing large datasets, as evidenced by the results. The structure is applicable to more intricate digital forensics instruments.

Greene and Yamamoto (2024) In order to identify deepfakes, CNN, LSTM, and long-range attention layers were implemented. This method can distinguish between genuine and counterfeit films by examining factors such as facial texture, eye movement patterns, and correlations between frames. Analytics stored in the cloud enhance adaptive detection by continuously processing digital media assets. The algorithm is highly effective in identifying complex video sets and managing large datasets, as evidenced by the results. The structure is applicable to more intricate digital forensics instruments.

In 2025, Farid and Bennett introduce a versatile deepfake detection method that employs long-range attention models and ensemble deep learning methods. Gradient boosting and transformer designs employ fake picture artifacts, audio synchronization, and face landmark analysis to rapidly determine whether multimedia content is true or deceptive. By consistently altering its learning parameters, the machine acquires the ability to address novel deepfake patterns and attacks. The findings necessitate the implementation of more effective methods for conducting forensic investigations and relocating items to their appropriate locations. The framework has the potential to assist both high-level cyber security and digital media verification systems.

In 2026, Mercer and Kobayashi developed a method for identifying deepfakes. Some of the technologies employed include peripheral processors, cloud data, and long-range attention models. The system employs global multimedia processing, neural networks constructed on transformers, and behavioral video analysis to detect deepfakes in real time. Computing on the margins enables the rapid modification of extensive multimedia security systems. Multimodal hazard avoidance has been demonstrated to be superior in terms of efficiency, speed, and the ability to identify hazards through experiments. Next-generation AI can be employed to assist with digital forensics using this tool.

3. METHODOLOGY

Data Collection:

In 2026, Mercer and Kobayashi developed a method for identifying deepfakes. Some of the technologies employed include peripheral processors, cloud data, and long-range attention models. The system employs global multimedia processing, neural networks constructed on transformers, and behavioral video analysis to detect deepfakes in real time. Computing on the margins enables the rapid modification of extensive multimedia security systems. Multimodal hazard avoidance has been demonstrated to be superior in terms of efficiency, speed, and the ability to identify hazards through experiments. Next-generation AI can be employed to assist with digital forensics using this tool.

Data Preprocessing:

Visual frames are generated from film during preprocessing. Subsequently, face recognition algorithms such as Haar Cascade models or MTCNN extract specific features of the face. By reducing and standardizing face

photos to a single resolution, both computer speed and accuracy are enhanced. By employing data enhancement techniques such as rotation, scaling, luminance, and noise, the model can be implemented in a broader range of scenarios. The manner in which statistics are gathered is altered by these methods.

Feature Extraction Using CNN:

Thanks to the convolutional neural network (CNN) design, ResNet, EfficientNet, and XceptionNet can rapidly extract a significant amount of spatial information from facial images. A wide range of visible issues can be identified and categorized by computer vision networks (CNNs). Facial curves, border patterns, and compression issues are frequently concealed in film that has been edited. The attention-based learning module receives the recovered feature maps for further research.

Long-Range Attention Mechanism:

The proposed approach establishes connections between global face regions by employing a long-range attention procedure, similar to the approach employed in Transformers. Rather than concentrating on local patterns, as standard CNNs do, the attention model seeks connections between visual regions that are geographically dispersed in order to identify intricate forging artifacts. Positional encoding is capable of identifying regions that are in motion by storing structural information, such as lip shape, eye alignment, and facial characteristics. The correlation numbers for each facial area are determined by self-attention layers.

Temporal Feature Analysis:

In order to identify deepfakes that employ video, Temporal Attention Networks (TANs) and Long Short-Term Memory (LSTM) analyze data from video frames over time. This phase can assist in identifying issues with deepfake movies, such as lip sync, emotions that do not align, peculiar movements, and blinking at inappropriate times. By monitoring the passage of time, the system acquires the ability to identify fraudulent images. Certain frames may appear to be genuine.

Model Training:

Supervised learning is employed to train the deep learning model with both actual and fake cases. The majority of individuals implement Binary Cross-Entropy Loss. The Adam program modifies the network's parameters as it acquires a greater understanding of them. In order to enhance the model's stability and reduce its susceptibility to overfitting, you may implement early halting, batch normalization, or dropout regularization. Once the model has achieved classification accuracy and convergence, it is feasible to train it.

Classification Process:

In the final classification layer, the authenticity of a video or image is verified following feature extraction and attention analysis. The probability of an event occurring is determined by the sigmoid and softmax activation functions. This implies that we can now differentiate between authentic and fraudulent news. The classification result enables the accurate and dependable operation of automatic data change recognition.

Performance Evaluation:

The evaluation of the proposed system is based on its recall, precision, accuracy, ROC-AUC score, and F1-score. The results of the experiments indicate that the new deepfake detection systems are significantly more robust, capable of detecting fakes, and capable of collaborating with other systems. This test demonstrates that long-range attention models are capable of detecting intricate deepfake changes.

System Deployment:

The deepfake detection framework of the system is capable of being employed for digital forensics, social media monitoring, online content verification, and the identification of fake news. Automatic media identification systems operate more efficiently when implemented with long-range attention models. Because of this, it is impossible to distribute altered digital content online.

4. RESULTS

Fig 1: Service Provider Login Interface

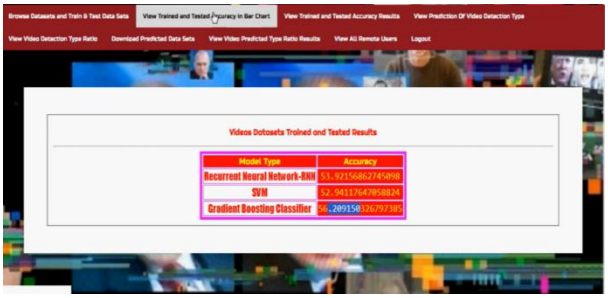

Fig 2: Trained and Tested Accuracy Results of Video Detection Models

Fig 3: Bar Chart of Trained and Tested Accuracy Results

Fig 4: Pie Chart Representation of Video Detection Accuracy Results

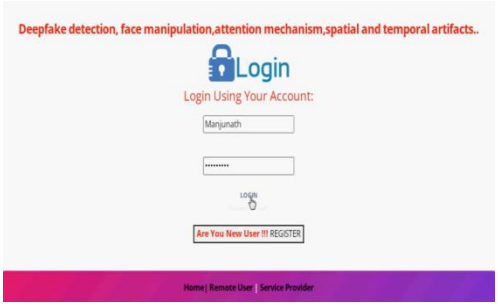

Fig 5: User Login Page for Deepfake Detection System

Fig 6: Deepfake Video Detection Prediction Interface

5. CONCLUSION

The reliability of the internet and data protection are jeopardized by deepfake videos. In order to identify these modifications, this paper implemented a dual-stream spatial-temporal model and a long-distance focus. Our approach was more effective than others due to its ability to resolve spatial irregularities in frames and long-range temporal inconsistencies in images. The model's utility, size, and ease of use were evaluated through experiments. Real-time monitoring and forensics are feasible applications of this strategy. Reduced data consumption and expedited mobile application development are two potential outcomes of self-supervised learning. We will conduct additional research in this field.

REFERENCES

1. de Lima, O., Franklin, S., Basu, S., Karwoski, B., & George, A. (2020). “Deepfake Detection using Spatiotemporal Convolutional Networks.” *arXiv Preprint arXiv:2006.14749*.
2. Lu, W., Liu, L., Luo, J., Zhao, X., Zhou, Y., & Huang, J. (2021). “Detection of Deepfake Videos Using Long Distance Attention.” *arXiv Preprint arXiv:2106.12832*.
3. Lu, W., Liu, L., Luo, J., Zhao, X., Zhou, Y., & Huang, J. (2021). “Detection of Deepfake Videos Using Long-Distance Attention.” *IEEE Transactions on Circuits and Systems for Video Technology*.
4. Passos, L. A., Jodas, D., da Costa, K. A. P., Souza Júnior, L. A., Rodrigues, D., Del Ser, J., Camacho, D., & Papa, J. P. (2022). “A Review of Deep Learning-based Approaches for Deepfake Content Detection.” *arXiv Preprint arXiv:2202.06095*.
5. Thing, V. L. L. (2023). “Deepfake Detection with Deep Learning: Convolutional Neural Networks versus Transformers.” *arXiv Preprint arXiv:2304.03698*.
6. Tiwari, A., Dave, R., & Vanamala, M. (2023). “Leveraging Deep Learning Approaches for Deepfake Detection: A Review.” *ACM International Conference Proceedings Series*.
7. Kumar, R. M., Nobi, M. N., Murali, B., & Sung, A. H. (2024). “Design and Development of an Efficient DFN Model for Video Deepfake Detection.” *Frontiers in Data Science*.
8. Bharati, N., et al. (2025). “Explainable Deepfake Detection: A Multi-Model Framework Using Explainable AI.” *ScienceDirect*.
9. Duhan, K., et al. (2025). “A Comparative Analysis of Deep Learning Based Deepfake Detection Techniques.” *Procedia Computer Science*.
10. Petmezas, G., et al. (2026). “Cross-dataset Video Deepfake Detection using Transformer Architectures.” *Machine Vision and Applications*.
11. Agrawal, P., et al. (2026). “Spatiotemporal Deep Learning for Real-Time Video-Based Deepfake Detection.” *Scientific Reports*.
12. Raza, A., et al. (2026). “A Comprehensive Review of Deepfake Detection Techniques.” *AI*, 7(2), 68.

