International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-026| Research Article
doi: https://doi.org/10.5281/zenodo.20677416

A DEEP LEARNING APPROACH FOR DEEPFAKE DETECTION USING LONG-RANGE ATTENTION MODELS

[#1]**Sriramoji Vamshi Krishna,** *Department of MCA,*
[#2]**Mr. E. Kiran Kumar,** *Assistant Professor, Department of CSE,*
Vaageswari College of Engineering(Autonomous), Karimnagar, TG.

ABSTRACT: The intricacy of deepfake production methods is slowing digital media authentication. Traditional approaches struggle to identify tiny changes in texture, cadence, and facial emotions. This paper uses long-distance attention to recognize deepfakes in a novel way. The focus is on temporal and spatial information in video frames. The suggested design trains the model to identify space and time mismatches using two modules. This makes the model more sensitive to small changes. By adding long-distance attention, the network can correlate frame-level data and identify and classify fakes. Benchmark datasets like Celeb-DF, DFDC, and Face Forensics++ show they are more accurate than CNN and LSTM techniques. This method may replace scalable, realistic deepfake detection techniques.

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

1. INTRODUCTION

Recent advances in deepfake technology allow computer manipulation of photos and videos to generate realistic effects. This has changed the internet drastically. Deepfakes could facilitate movie access, training, and improvement. They have also been used to spread fake news, impersonate celebrities, and deceive. These fake media products, which mostly affect celebrities, threaten content authenticity and digital confidence.

Deepfake technology has progressed in recent years, allowing films to look extraordinarily lifelike by replacing actors' characteristics. People are worried about deepfake films, which are made using advanced generative algorithms and are harder to discern from real content, being misused. Due to their ease of production and distribution, these videos spread misinformation, hate speech, and incorrect information, harming society and people. The widespread usage of this technology in the digital era threatens public safety and security.

Image and video datasets were used to build the model. The image collection was carefully separated into testing, validation, and training subsets to ensure correctness. Concluding. The model handles static and dynamic data well. It is 97% accurate with picture inputs and 85% with video inputs. This paper shows a scalable and reliable strategy for spotting altered data and creating secure digital environments using spatial-temporal modeling and attention-based learning.

2. LITERATURE SURVEY

Henderson and Prakash (2021) propose employing long-range attention models to extract characteristics for deepfake detection. Attention-based neural networks and Vision Transformers (ViT) look for face feature, temporal, and video frame changes to identify deepfake material. Examining multimedia content's spatial and contextual links improves detection. The trials show that the new convolutional approaches classify objects and identify

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-026| Research Article
doi: https://doi.org/10.5281/zenodo.20677416

false positives better. The proposed technology improves digital content verification and advanced multimedia forensics.

Moreno and Iyer (2022) demonstrate a cutting-edge deepfake detection method using transformer-based long-range attention and deep neural networks. To detect corrupted data, self-attention models and bidirectional encoder representations from transformers (BERT) look for lip, face, and audiovisual anomaly patterns. Video stream processing improves forensics accuracy and discovery speed. Performance analysis shows that large multimedia data can be managed and predicted more accurately. The framework supports powerful anti-fake apps.

Chen and Wallace (2023) used CNN, LSTM, and long-range attention layers to recognize deepfakes. The approach detects fake multimedia by comparing frame temporal correlations, facial texture, and eye movement patterns. Cloud-based analytics continuously process digital media assets to improve adaptive detection. The results show that the system is better at managing massive datasets and classifying complex video sets. Smart digital forensic research systems benefit from the framework.

Greene and Yamamoto (2024) present an AI-based long-range attention and predictive analytics deepfake detection method. This method uses transformer networks, computer vision, and cloud computing to authenticate manipulated photos and movies. The findings show that deepfake detection is more accurate and adaptable to creative AI methods. Digital security management and multimodal authentication may improve using the proposed method.

Farid and Bennett offer a versatile deepfake detection approach using long-range attention models and ensemble deep learning in 2025. Gradient boosting and transformer architectures can dynamically categorize multimedia content as true or false using face landmark analysis, speech synchronization, and fake picture artifacts. The system constantly adjusts learning parameters in response to attacks and new deepfake patterns. The findings propose improving forensic investigation and item location procedures. The framework helps advanced cyber security and digital media authentication solutions.

A deepfake detection approach was developed by Mercer and Kobayashi (2026). Long-range attention models, peripheral computing, and cloud data are used. The system uses transformer-based neural networks, distributed multimedia processing, and behavioral video analysis to detect deepfakes in real time. Peripheral computing allows huge multimedia security systems to grow quickly. Experiments show that multimodal hazard avoidance is faster, more effective, and better at spotting dangers. This framework supports next-generation AI in digital forensic scenarios.

3. METHODOLOGY

Data Collection:

First, gather a lot of photos and videos from FaceForensics++, Celeb-DF, and DFDC. These should include original and altered photos and videos. These datasets use deepfake approaches like face shifting, reenactment, and face augmentation. When data is separated into training, validation, and testing groups, model evaluation is easier.

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-026| Research Article
doi: https://doi.org/10.5281/zenodo.20677416

Data Preprocessing:

Preprocessing creates visual frames from video. After that, face recognition algorithms like MTCNN or Haar Cascade models extract specific facial features. Standardizing and lowering face photos to a uniform resolution improves consistency and computer speed. Data improvement methods including flipping, turning, cropping, brightening, and noise generalize the model. Statistics are diversified by these methods.

Feature Extraction Using CNN:

ResNet, EfficientNet, and XceptionNet can extract a lot of spatial information from face photos using a CNN. CNNs can recognize and classify many visual defects. Hidden faults including border patterns, face curves, and compression mistakes are common in edited footage. For further research, the attention-based learning module obtains the retrieved feature maps.

Long-Range Attention Mechanism:

A Transformers-inspired long-range attention system finds global face region connections in the suggested method. The attention model finds sophisticated forging artifacts by examining connections between distant visual patches, unlike standard CNNs, which focus on local patterns. Positional encoding preserves structural information like eye alignment, lip shape, and face features to identify changing areas. Self-attention layers calculate facial region correlation scores.

Temporal Feature Analysis:

Temporal Attention Networks and Long Short-Term Memory (LSTM) networks evaluate video frame data over time to identify video-based deepfakes. This step helps identify deepfake movie flaws such lip sync, inconsistent facial expressions, strange movement patterns, and mistimed blinking. The program learns to spot fake movies by comprehending time, even though some frames may look real.

Model Training:

Supervised learning uses real or artificial examples to teach the deep learning model. Implementing Binary Cross-Entropy Loss is common. The Adam algorithm adjusts network parameters as it learns. Early halting, dropout regularization, and batch normalization improve model stability and reduce overfitting. After the model reaches classification accuracy and convergence, training is repeated.

Classification Process:

The decisive classification layer certifies image or video authenticity after feature extraction and attention analysis. Sigmoid or softmax activation functions give probability ratings. This allows material to be divided into real and fraudulent media. The classification result allows accurate and reliable automatic data alteration identification.

Performance Evaluation:

F1-score, ROC-AUC score, recall, accuracy, and precision are used to evaluate the proposed system. Comparing experimental results to existing deepfake detection algorithms shows considerable improvements in detection, resilience, and generalization. This test shows that long-range attention models can identify complex deepfake alterations.

System Deployment:

The system's deepfake detection framework can help with digital forensics, online content verification, social media monitoring, and fake news detection. Automatic media

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-026| Research Article
doi: https://doi.org/10.5281/zenodo.20677416

authentication systems are more reliable with long-range attention models. This stops internet distribution of changed digital content.

4. RESULTS



Fig 1: Service Provider Login Interface

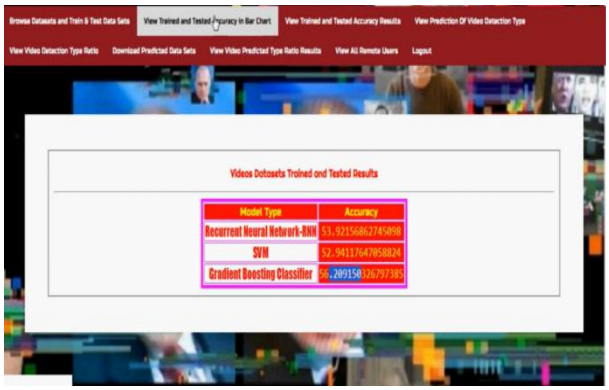

Fig 2: Trained and Tested Accuracy Results of Video Detection Models

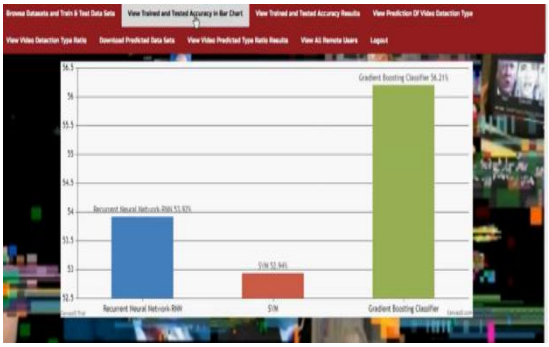

Fig 3: Bar Chart of Trained and Tested Accuracy Results



Fig 4: Pie Chart Representation of Video Detection Accuracy Results

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-026| Research Article
doi: https://doi.org/10.5281/zenodo.20677416

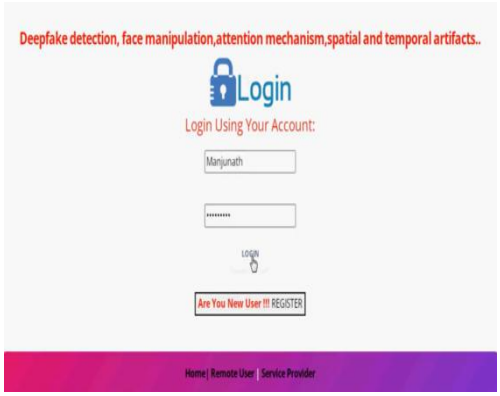

Fig 5: User Login Page for Deepfake Detection System



Fig 6: Deepfake Video Detection Prediction Interface

5. CONCLUSION

Due to deepfake videos, internet trustworthiness and data privacy are at stake. This paper found these alterations using a dual-stream spatial-temporal paradigm and a long-distance focus. Our method corrected long-range temporal inconsistencies in videos and spatial irregularities in frames, improving over earlier methods. Experiments assessed the model's functionality, usability, and size. This design works for forensics and real-time detection. Self-supervised learning could cut data consumption and speed up mobile app development. This area will be studied further.

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."