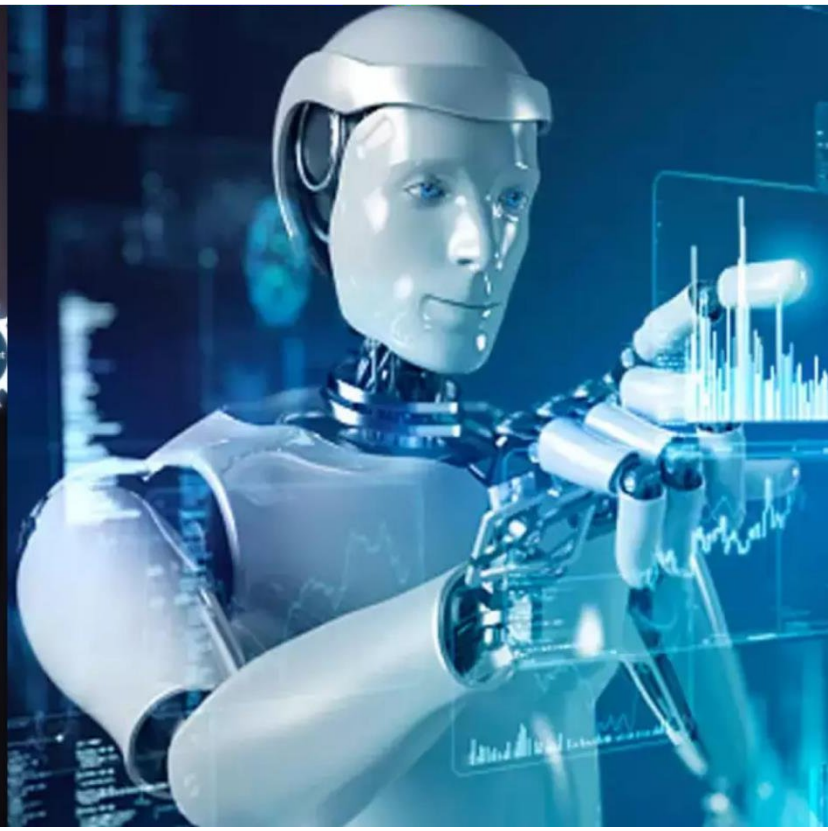
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 4, April 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504210|

Deep Learning Based Deepfake Detection System

Pallavi Mahadev Biradar, Prof. Anil Gujar

Department of Computer Engineering, TSSM's Bhivarabai Sawant College of Engineering and Research, Pune, India

ABSTRACT: Recent years have seen a fast advancement in deepfake technology, making it possible to produce extremely lifelike fake photos and movies that are frequently indistinguishable from actual content. This presents significant concerns in areas like political influence, disinformation, and identity theft. The subtle temporal and spatial irregularities found in forged material are frequently missed by many current detection methods, which solely concentrate on static images or individual video frames. In order to accurately distinguish between real and false media, this study suggests an improved deepfake detection system that integrates Multi-Head Self-Attention (MHSA) mechanisms with Bi-Directional Long Short-Term Memory (BiLSTM) networks. The system expands the capabilities of the BMNet framework to enable deepfake detection based on both images and videos. While self-attention processes assist in focusing on localized face changes in both photos and video sequences, temporal discrepancies are collected across frames for video analysis. The goal of this system is to offer a reliable, user-friendly deepfake detection tool that supports digital content verification on several platforms. Journalists, content moderators, and the general public can benefit from the solution's high accuracy and generalization, particularly in the fight against false information and upholding digital trust.

KEYWORDS: Deepfake, deepfakeimage/video detection, bi-directional long short-term memory network, deep learning.

I. INTRODUCTION

Robust detection is necessary because deepfakes pose a severe threat to digital trust. By concentrating just on visual artifacts and neglecting crucial temporal irregularities between video frames, current CNN-based techniques fall short. This is resolved by our BMNet design, which combines geographical and temporal analysis.

Conventional deepfake detectors primarily rely on Convolutional Neural Networks (CNNs) for spatial analysis, focusing on finding visual irregularities and subtle artifacts within single frames. While effective for basic manipulation, these models falter against advanced forgeries because they treat frames independently, ignoring the sequential relationship inherent in video. This results in a critical inability to model temporal coherence across video sequences, leaving the systems vulnerable to sophisticated deepfakes that strategically employ inconsistent movements or flicker artifacts across time to bypass static detection methods. The current reliance on solely spatial cues creates a generalization gap that modern countermeasures must urgently address.BMNet was created to get around these existing system constraints. In order to capture thorough sequential dynamics and long-range relationships, it first extracts features using a CNN base and feeds those features into a Bidirectional Long Short-Term Memory (BiLSTM) layer. Importantly, in order to maximize the discriminative capability of the model, a Multi-Head Self-Attention Mechanism (MHSAM) is subsequently used to intelligently weigh and concentrate on the most forgery-indicative features and locations. By successfully modeling minute spatiotemporal artifacts, this innovative combination of BiLSTM and MHSAM guarantees excellent, robust detection.

II. LITERATURE SURVEY

Rue Marconi [1],Multi-channel techniques have been put out recently to increase PAD systems' resilience. The efficacy of these techniques is frequently restricted since only a small amount of data is available for additional channels. In this study, we introduce a new PAD framework that combines a novel loss function with RGB and depth channels.

Mathilde Caron [2],In this paper, we investigate whether self-supervised learning gives Vision Transformer (ViT) [19] additional characteristics that make it different from convolutional networks (convnets). We note that self-supervised ViT features contain explicit information about the semantic segmentation of an image, which does not

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504210|

emerge as clearly with supervised ViTs or convnets, in addition to the fact that adapting self-supervised methods to this architecture works particularly well.

AnukulMuley [3],ATM security is ensured by a variety of technologies, including RFID, fingerprint, facial recognition, iris scan, OTP, reference number, random keypad, and others. Card and PIN numbers are needed for authentication in a typical ATM system, which leads to security problems including misplaced cards, stolen pin numbers, card cloning, shoulder surfing, phony keyboards, skimming, etc. The development of anti-spoofing based facial recognition for ATM systems employing liveness detection is the primary emphasis of this work.

Jianyu Xiao [4],The security of facial recognition systems is greatly enhanced by the facial Anti-Spoofing (FAS) techniques. In short-range situations, like as phone unlocking and face payment, the current FAS techniques function well. However, because of the inconsistent image quality, it is still difficult to enhance the generalization of FAS in long-distance settings (such as surveillance).

RaghavendraMudgalgundurao [5],In order to identify presentation attacks of printed and digitally replayed attacks, this study suggests a pixel-wise supervision on DenseNet. The authors advocate using pixel-wise supervision to exploit minute clues on a range of artifacts, including as moiré patterns and printing artifacts. Several handmade and deep learning models are used to display the baseline benchmark utilizing a newly constructed in-house database from an operational system with 886 users with 433 genuine, 67 print, and 366 display attacks.

Akshay Kumar [6],This study suggests a facial recognition technology for smart ATM transactions. The user must enter their registered mobile number in order to proceed with the transaction, after which an OTP will be created. The camera installed atop the ATM will then take a picture of the user's face, which will be compared with the picture linked to the contact number using a specially trained YOLO algorithm.

SiddhikaPokharkar [7],In order to increase the security of ATM transactions, this study proposes a dual-factor authentication system that combines face recognition and PIN verification. The suggested solution greatly lowers the risk of unwanted access by using a Convolutional Neural Network (CNN) for real-time face recognition.

Sangeetha S [8],The review also analyses research gaps and future directions in areas like multimodal fusion, explainable AI, collaborative learning, online adaptation etc. Overall, this review provides useful insights into the capabilities and limitations of existing literature which can guide future research on secure face recognition systems for attendance marking.

III. METHODOLOGY

The hybrid spatio-temporal deep learning pipeline used in the suggested deepfake detection system is intended to detect both visual artifacts and time-based discrepancies found in edited films. As explained below, the entire working process is broken down into methodical steps.

1. Dataset Collection
A sizable benchmark dataset is gathered from publicly accessible sources and includes both real and deepfake films. Numerous editing techniques, including face switching, lip-sync change, and expression replication, are included in the dataset. To make sure the model learns robust and generic features, videos with different lighting conditions, poses, backdrop complexity, and resolutions are used.

2. Video Frame Extraction
Every video is broken down at a set frame rate into a series of picture frames. Extracting consecutive frames aids in maintaining motion continuity and sequential linkages because videos are temporal data. The main input for spatiotemporal modeling is these frame sequences.

3. Data Preprocessing
The retrieved frames go through a number of preprocessing steps to guarantee consistency and boost learning effectiveness:

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504210|

• Resizing: A consistent resolution appropriate for the CNN input layer is applied to every frame.
• Normalization: To stabilize gradient updates, pixel intensity values are adjusted to a standard range.
• Face Localization: To eliminate background noise, facial parts are identified and clipped.
• Face Alignment: To standardize orientation, important face landmarks are aligned.
• Data augmentation: To avoid overfitting and enhance generalization, methods like flipping, rotation, and brightness variation are used.

4. Spatial Feature Extraction

The retrieved frames go through a number of preprocessing steps to guarantee consistency and boost learning effectiveness:
• Resizing: A consistent resolution appropriate for the CNN input layer is applied to every frame.
• Normalization: To stabilize gradient updates, pixel intensity values are adjusted to a standard range.
• Face Localization: To eliminate background noise, facial parts are identified and clipped.

5. Temporal Sequence Modeling

To ensure consistency and increase learning efficacy, the retrieved frames undergo many preprocessingsteps:
• Resizing: Each frame is given a uniform resolution suitable for the CNN input layer.
• Normalization: Pixel intensity values are modified to a standard range in order to stabilize gradient updates.
• Face Localization: Parts of the face are recognized and trimmed to remove background noise.

6. Attention-Based Feature Enhancement

The BiLSTM output is subjected to a Multi-Head Self-Attention mechanism in order to increase detection precision.
• Gives several attributes priority weights.
• Concentrates on areas vulnerable to forgery
• Hides background information that isn't relevant
• Improves spatiotemporal discriminative cues
The model can assess several representation subspaces at once thanks to its multi-head design, which enhances contextual comprehension.

7. Feature Fusion and Classification

The fully connected thick layers that receive the improved features are as follows:
• Learning of high-level representations
• The application of non-linear transforms Overfitting is avoided by dropout regularization. Lastly, binary classification is carried out by a Softmax activation function:
• Class 0: Actual Video
• Class 1: Video Deepfake

8. Model Training

Supervised learning is used to train the network. Binary Cross-Entropy Loss is the loss function.
• Adam optimizer for effective convergence
• Gradient descent in mini-batch processing
• Backpropagation: Reduces classification error by updating weights
Until validation performance stabilizes, training goes on.

9. Performance Evaluation

Standard evaluation metrics are used to gauge the proposed system's effectiveness:
• Accuracy: Total correctness
• Accuracy and dependability of deepfake detection
• Recall: The capacity to identify altered videos
• F1-Score: A balance between recall and precision
To validate advances, comparative trials are carried out against traditional CNN-only models.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504210|

IV. PROPOSED SYSTEM

The BMNet design uses a hybrid strategy to overcome CNNs' temporal restrictions. The Bidirectional Long Short-Term Memory (BiLSTM) network, which is intended to record thorough sequential dependencies throughout the video timeline, receives the spatial data that are first extracted using a CNN backbone. A Multi-Head Self-Attention Mechanism (MHSAM), which intelligently weighs the most important spatio-temporal cues, further improves this temporal modeling and greatly increases the model's discriminative ability and robustness against subtle, time-based forgeries.

V. CONCLUSION

In conclusion, by effectively incorporating crucial spatiotemporal analysis, the BMNet framework marks a substantial development in deepfake detection technology. BMNet overcomes the crucial flaw of earlier, spatially-constrained CNN models thanks to the combination of the MHSAM for intelligent feature discrimination and the robust BiLSTM for sequence modeling. This innovative hybrid technique demonstrates that it produces better results and increased resistance to increasingly complex media manipulation, offering a more dependable basis for detecting and thwarting deepfake technologies.

REFERENCES

[1] Rue Marconi 19, CH- 1920, Martigny, Switzerlan" Cross Modal Focal Loss for RGBD Face Anti-Spoofing", published year arXiv:2103.00948v1 [cs.CV] 1 Mar 2021.
[2] Mathilde Caron12 Julien Mairal2 Hugo Touvron13 Ishan Misra1 Piotr Bojanowski1 Herv´e Jegou1 Armand Joulin" Emerging Properties in Self-Supervised Vision Transformers", published year arXiv:2104.14294v2 [cs.CV] 24 May 2021.
[3] Anukul Muley1, Akash Bendre2, Priti Maheshwari3, Shanmukh Kumbhar4, Prof.BhagyashreeDhakulkar," Anti-Spoofing Based Secured Transaction Using Facial Recognition And 2FA", 2022.
[4] Jianyu Xiao 1, Wei Wang 2, Lei Zhang 1 and Huanhua Liu," AMobileFaceNet-Based Face Anti-Spoofing Algorithm for Low-Quality Images", published year Electronics 2024, 13, 2801. https://doi.org/10.3390/ electronics13142801.
[5] RaghavendraMudgalgundurao, PatrickSchuch, KiranRaja, RaghavendraRamachandraNaserDamer "Pixel-wisesupervisionforpresentationat tackdetectiononidentitydocumentcards", Accepted:7June2022.
[6] Akshay Kumar1, Pooja Joshi2, Anju Bala3, PravinkumarSudhakar Patil4, Dilip Kumar Jang Bahadur Saini5 and Kapil Joshi," Smart Transaction through an ATM Machine using Face Recognition", Accepted 30 October 2023.
[7] Siddhika Pokharkar1, Bhagyashri Bhalerao2, Pratiksha Dolas3, Dr.Anand Khatri," A Real Time Face Detection & Security System for ATM Machine." 2025.
[8] Sudha V, Sangeetha S, Ganesh Shankar S," FACE RECOGNITION-BASED ATTENDANCE SYSTEMS USING ANTI-SPOOFING METHODS", 2024.
[9] DeeptiVarshney, Birendra Kumar Sharma, Mamta Bansal "Secure Watermarking to Protect Colour Images on Social Media from Misuse", Electronic Systems and Intelligent Computing. Lecture Notes in Electrical Engineering, vol 860, June 2022, Springer, Singapore, ISBN: 978-981- 16-9487-5.
[10] DeeptiVarshney "Secure Watermarking Technique for Color Images using Aadhar Number, DWT and SVD" Turkish Journal of Computer and Mathematics Education,Vol.12 No.6 (2021), 4252-4268.
[11] "Robust Watermarking Technique for Sharing Family Photos on Social Media using Aadhar Number and DCT", International Journal of Recent Technology and Engineering, ISSN: 2277-3878 , Volume-9 Issue-3, September 2020. Page No.:381-387.
[12] "Watermarking for images using Alphanumeric Technique", International Journal of Recent Technology and Engineering, ISSN: 2277-3878 , Volume-8 Issue-6, March 2020. Page No.: 2639- 2645.
[13] "A Singular Value Decomposition Based Robust Image Watermarking Scheme" International Journal of Computer Science & Communication, Vol. 7,Sep 2015 - March 2016, pp 89-94, ISSN: 0973-739.
[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, ''Generative adversarial networks,'' Commun. ACM, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504210|

[15] B. Chesney and D. Citron, ‘‘Deep fakes: A looming challenge for privacy,democracy, and national security,’’ California Law Rev., vol. 107, p. 1753, Jan. 2019, doi: 10.2139/ssrn.3213954.
[16] K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Deep residual learning for image recognition,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[17] M. Tan and Q. V. Le, ‘‘EfficientNet: Rethinking model scaling for convolutional neural networks,’’ in Proc. Int. Conf. Mach. Learn.,Jan. 2019, pp. 6105–6114.
[18] E. Sabir, J. Cheng, A. Jaiswal,W.AbdAlmageed, I. Masi, and P. Natarajan, ‘‘Recurrent convolutional strategies for face manipulation detection in videos,’’ Interface (GUI), vol. 3, no. 1, pp. 80–87, Jan. 2019.
[19] V. S. Nguyen, H. Kim, and D. Suh, ‘‘Attention mechanism-based bidirectional long short-term memory for cycling activity recognition using smartphones,’’ IEEE Access, vol. 11, pp. 136206–136218, 2023, doi: 10.1109/ACCESS.2023.3338137.
[20] X. Yang, Y. Li, and S. Lyu, ‘‘Exposing deep fakes using inconsistent head poses,’’ in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2019, pp. 8261–8265, doi: 10.1109/ICASSP.2019.8683164.
[21] W. Liu, T. She, J. Liu, B. Li, D. Yao, Z. Liang, and R. Wang, ‘‘Lips are lying: Spotting the temporal inconsistency between audio and visual in lipsyncingDeepFakes,’’ 2024, arXiv:2401.15668.
[22] T. Jung, S. Kim, and K. Kim, ‘‘DeepVision: Deepfakes detection using human eye blinking pattern,’’ IEEE Access, vol. 8, pp. 83144–83154, 2020, doi: 10.1109/ACCESS.2020.2988660.
[23] O. A. H. H. Al-Dulaimi and S. Kurnaz, ‘‘A hybrid CNN-LSTM approach for precision deepfake image detection based on transfer learning,’’ Electronics, vol. 13, no. 9, p. 1662, Apr. 2024, doi: 10.3390/electronics13091662.
[24] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner, ‘‘FaceForensics++: Learning to detect manipulated facial images,’’ in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 1–11, doi: 10.1109/ICCV.2019.00009.
[25] F. Chollet, ‘‘Xception: Deep learning with depthwise separable convolutions,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 1800–1807, doi: 10.1109/CVPR.2017.195.
[26] M. Zou, B. Yu, Y. Zhan, S. Lyu, and K. Ma, ‘‘Semantics-oriented multitask learning for DeepFake detection: A joint embedding approach,’’ 2024, arXiv:2408.16305.
[27] C. Zhao, C. Wang, G. Hu, H. Chen, C. Liu, and J. Tang, ‘‘ISTVT: Interpretable spatial–temporal video transformer for deepfake detection,’’ IEEE Trans. Inf. Forensics Security, vol. 18, pp. 1335–1348, 2023, doi: 10.1109/TIFS.2023.3239223.
[28] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, ‘‘Active shape models-their training and application,’’ Comput. Vis. Image Understand., vol. 61, no. 1, pp. 38–59, Jan. 1995.
[29] T. F. Cootes, G. J. Edwards, and C. J. Taylor, ‘‘Active appearance models,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 6, pp. 681–685, Jun. 2001, doi: 10.1109/34.927467.
[30] P. Dollár, P. Welinder, and P. Perona, ‘‘Cascaded pose regression,’’ in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 1078–1085, doi: 10.1109/CVPR.2010.5540094.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details