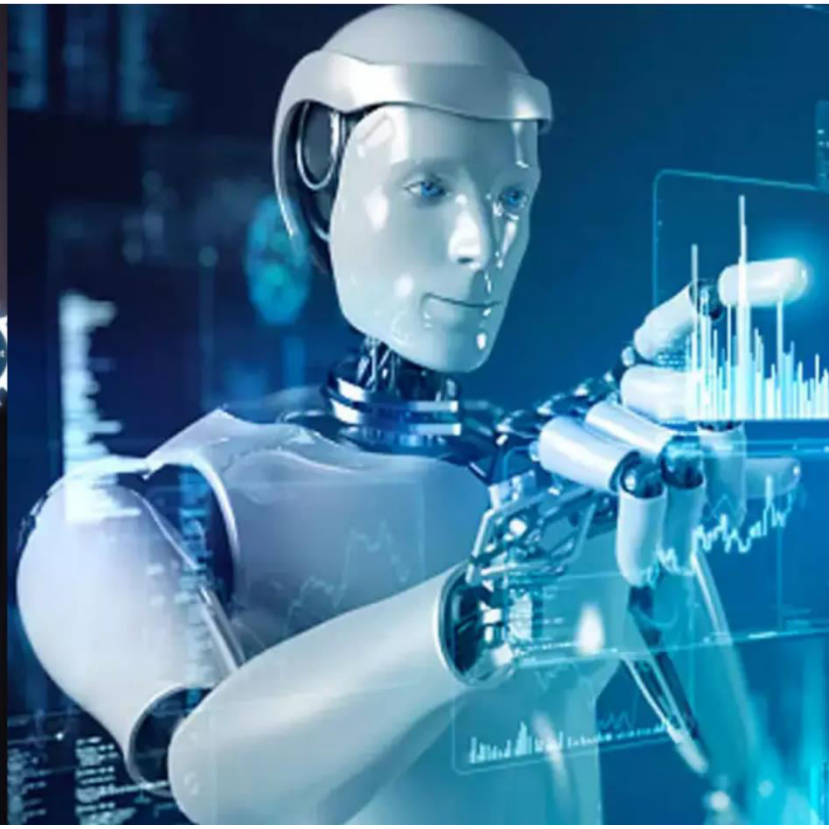
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

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, Special Issue 1, January 2026

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

International Conference on Emerging Technologies and Pharmaceutical Research (ICETPR 2026)

17–18 January 2026

Indala College of Engineering, Polytechnic & Pharmacy, Kalyan, MH, India

|Volume 15, Special Issue 1, January 2026|

Enhancement and Implementation of a Deep CNN BiLSTM Model with Attention Mechanism for Deepfake Image and Video Detection

Prof. Megha Joshi[1], Nidhi Khobarekar[2], Piyush Kowale[3], Hitesh Sharma[4], Aditya Sonawane [5]

Professor, Dept. of CE, Indala College of Engineering, Kalyan, Maharashtra, India [1]

Dept. of CE, Indala College of Engineering, Kalyan, Maharashtra, India [2]

Dept. of CE, Indala College of Engineering, Kalyan, Maharashtra, India [3]

Dept. of CE, Indala College of Engineering, Kalyan, Maharashtra, India [4]

Dept. of CE, Indala College of Engineering, Kalyan, Maharashtra, India [5]

ABSTRACT: Deep Learning Technology is increasingly being used to produce a large number of high-quality fake images and video realising considerable digital security and loss of public confidence. A better, enhanced model of deep learning for detecting deep fakes more accurately is being proposed by combining a Deep Convolutional Neural Network (CNN) with a Bi-directional Long Short Term Memory (BiLSTM) along with an attention mechanism. This model preprocesses the media content using methods of face detection, frame extraction, and then normalizing the media content which in turn will lead to improved quality of the content that is being supplied to the model. A deep Convolutional Neural Network will extract significant spatial features from the facial textures of the media while Bi-directional Long Short Term Memory networks capture and interpret temporal inconsistencies over video frames. The model provides an efficient and reliable means for much quicker and easier detection of deep fakes in both Cyber Security and Digital Forensics Applications.

KEYWORDS: Deepfake detection, Deep CNN, BiLSTM, Attention mechanism, Video forensics, Image forensics

I. INTRODUCTION

The accelerated progression in Artificial Intelligence capabilities, along with improvements made to Deep Learning tools over time, have radically changed how people create, edit and share Digital Media. Arguably, one of the more troubling consequences of these advances has been the advent of new Technologies involving Deepfake audio-visual information. Deep Fakes refer to artificially produced Audio or video footage presenting "Images" that closely resemble real people but do not represent any real person at all. Once created, they could be used to create Smart Face Swap Videos or Photos, Fake Lip-Synced Audio, Videos, or even Visual Distortions (e.g. using Face-Effect Filters) of other people's appearances. Henceforth, the potential benefits arising from these new Deepfake technologies for various Entertainment, VR, and Educational purposes may lead to them being misused to threaten Digital Security, Privacy, Political Stability, and Trust.

The increasing availability of advanced computational resources, along with complementary Open Source Deep Learning/AI Toolsets, has increased the opportunities for disadvantageous users to utilise the technology to create fraudulent Digital Media. As a result of this, Deepfakes have been increasingly misused by malicious actors for Identity Theft, Financial Fraud, Cyberbullying, and Disinformation Campaigns for Political Manipulation. These developments mean that conventional verification methods for Digital Media no longer provide adequate means of identifying manipulated media due to the phenomenal quality of realistic imaging of today's Deepfaked video and Audio content.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

International Conference on Emerging Technologies and Pharmaceutical Research (ICETPR 2026)

17–18 January 2026

Indala College of Engineering, Polytechnic & Pharmacy, Kalyan, MH, India

|Volume 15, Special Issue 1, January 2026|

II. SYSTEM MODEL AND ASSUMPTIONS

Either static facial images or video sequences can be used as the input for the system. The videos are uniformly sampled into frames and then undergo preprocessing steps including facial detection, normalisation, and enhancement so that each frame has the same level of quality across all frames. Each frame represents a very high dimensional tensor and will then pass through the deep convolutional neural network (Deep CNN) which will allow it to automatically learn hierarchical spatial features such as texture inconsistencies, blending artefacts, pixel distortions, and abnormal distributions of colours that are introduced during the creation of a deepfake. The spatial features that are extracted frames are then ordered in sequence and sent to the bidirectional long short-term memory (BiLSTM) network. The BiLSTM will process the feature sequences both forwards and backwards and this enables it to effectively learn long-term temporal relationships between the frames and identify motion anomalies including unnatural eye blinks, lip-sync mismatches, and inconsistent facial expressions that are commonly found in deepfake videos.

The results from the BiLSTM's processing of frames will then undergo further classification through the application of an attention mechanism to the output. The attention mechanism helps to assign a greater weight to frames and regions of the face that provide the highest levels of discriminative power while reducing the weight of frames and facial regions that do not provide enough information. This allows the model to pay more attention to evidence of forgery and therefore achieve a higher level of classification accuracy, faster convergence time, and provide greater insight into how it arrived at its conclusion. The output decision produced by the system will consist of either a positive or a negative classification that is made using a fully connected network with a sigmoid function to classify whether or not the frame is a deep fake.

III. EFFICIENT COMMUNICATION

In order for the proposed Deep CNN-BiLSTM with Attention Mechanism-Based deepfake detection system to function correctly, an established format for how information will be exchanged between system components must exist. This information includes of both multimedia data transmitted back and forth between all components and also learned features produced from training Neural Networks with multi-layered architectures via supervised learning methodologies. The first step in the communication process begins when users or surveillance feeds and/or various on-line services stream raw videos/images over the internet to a central system. This raw data is then sent via Ethernet cables directly to Preprocessing Modules. Within Preprocessing Modules various functions occur, such as, Face Detection, Frame Extraction, Normalizing Images, Enhancing Images. By optimally Preprocessing Images before sending them to Learning Modules this significantly reduces (noise) Un-Relevant Image Content and eliminates unnecessary data across Learning Modules.

Once Preprocessing has been completed, a alternate set of facial frames is transferred to and reused in the Deep CNN module, which excerpts spatial features including; Texture inconsistencies, Blending Vestiges and Pixel deformation. These uprooted Spatial Features are transferred directly in successional order to BiLSTM modules for farther training (via Bi-directional RNN) to be suitable to understand time- grounded dependences across videotape frames. Through Effective Communication, this system of connecting CNN and BiLSTM Modules ensures there's no point Losses between Spatial and Temporal point Sets. also, the preface of Attention Mechanisms into the system assists in perfecting Effective Communication because it assigns Advanced Weights to the most significantly Important Frame/ Facial Regions. By concentrating on only Critical Image Input Data for Spatial and Temporal point Identification, this dramatically decreases the quantum of Duplicate Feature Transmissions, which increases speed to reuse and improves performance of the overall system.

IV. SECURITY

The deepfake system proposes a deep CNN- BiLSTM model incorporating an attention medium employed descry both images and vids without compromising their security. An essential function of this system must be to maintain confidentiality, trustability, and responsibility (integrity and vacuity) for all multimedia information (data) reused on it and maintain the trust of druggies as to the delicacy of the affair produced.

The processing of image and videotape lines by the proposed system will involve storing multiple terabytes of multimedia lines containing particular relating data. To minimize the threat of unauthorized access to this data, all

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

International Conference on Emerging Technologies and Pharmaceutical Research (ICETPR 2026)

17–18 January 2026

Indala College of Engineering, Polytechnic & Pharmacy, Kalyan, MH, India

|Volume 15, Special Issue 1, January 2026|

datasets reused by the proposed system will be securely stored in an translated format within good digital surroundings. Access to the datasets will be rigorously limited to individualities having applicable access boons; also, wherever possible, data anonymization will be employed in order to avoid exposure of any tête-à-tête identifiable information. Eventually, regular translated backups will be performed at intervals established by the association in order to promote data integrity and permit the association to recover snappily from a data loss due to accidental omission, mechanical failure, or cyberattack.

Model security is critical because it represents an association's investment in intellectual property that may represent leveraged means. Trained parameters of the Deep CNN- BiLSTM model representing influence Source. Agents will be suitable to pierce this knowledge by penetrating the Director's intranet (or a element of it).

Communication security ensures that all information changed among the stoner interface, preprocessing module, deep literacy model, and affair interface is secure from unauthorized interception or tampering. All the data flowing through these communication channels are secured by the diligence most trusted encryption protocols including SSL/ TLS, precluding third- party wiretapping and" man- in- the- middle" attacks. Secure customer authentication commemoratives and secure session operation styles allow only vindicated druggies and services to communicate with the system. The combination of these two measures ensures that all data flows through the entire process securely and reliably.

V. DISCUSSION

A Deep CNN – BiLSTM model enhanced with an attention medium represents an advanced and holistic system for relating deepfake images and vids. In addition, the rapid-fire growth of the quality and availability of deepfakes has rendered homemade forensic ways and traditional forensic styles ineffective. This study addresses these issues by developing a unified frame that includes all three point literacy spatial, temporal, and contextual, enabling it to descry small artefacts of manipulation that the mortal eye cannot see. In addition to this, the combination of CNNs, BiLSTMs, and attention mechanisms results in substantial advancements over using one single module to descry deepfake vids because it allows the system to descry both pixel- position inconsistencies and sequence- position abnormalities in manipulated video. From the model performance discussion, it's apparent that CNNs form the base of the system by detecting all the spatial features of manipulated vids, including the deformations of textures, blending irregularities, unnatural lighting, and missing regions at facial boundaries. The crucial limitation of using spatial information alone exists in videotape- grounded deepfake discovery, where the temporal consonance of a videotape has been altered. To address this limitation, the BiLSTM module analyses a sequence of frames, in both forward and backward directions to descry the presence of unnatural temporal patterns, similar as mismatched lip movements, inconsistent facial expressions, or irregular blink patterns. The substantiation from both forward and backward directional analysis greatly improves the overall delicacy of findings, as there's generally a mismatch between the two directional analyses of the frames.

VI. CONCLUSION

The suggested Deep CNN- BiLSTM Model with the Attention Mechanisms is a comprehensive approach for combating the adding trouble of DeepFakes (image and/ or videotape). With its use of CNN to prize Spatial Features, BiLSTM for Temporal Sequence literacy, and the use of Attention for Intelligent point Selection/ Prioritization, this System can descry both nanosecond, pixel- position substantiation of manipulation as well as frame- position inconsistencies that are most attributed to ultramodern DeepFake technology. ThisMulti-Level Learning Architecture provides bettered delicacy and advanced degree of robustness conception against colorful Combinations of DeepFake Generation styles.

REFERENCES

1. Deepfake Transformer Models
   Zhang, Y. Li, “Motor- grounded Multi-modal Deepfake Discovery Using Cross Attention” IEEE Deals on Biometrics, 2024.
2. Lightweight Deepfake Detection
   Ranjan, A. Das, “Mobile- Light Deepfake Sensor for Real- Time Face Manipulation Analysis” Pattern Recognition

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

International Conference on Emerging Technologies and Pharmaceutical Research (ICETPR 2026)

17–18 January 2026

Indala College of Engineering, Polytechnic & Pharmacy, Kalyan, MH, India

|Volume 15, Special Issue 1, January 2026|

Letters, 2024.

3. CNN Attention Grounded Discovery
   Kim, R. Park, “mongrel CNN- Attention Network for High- delicacy Deepfake Image Identification” Computers & Security, 2023.
4. BiLSTM Based Video Deepfake Detection
   Singh, P. Jaiswal, “Temporal Deepfake Discovery Using BiLSTM with Sequence- position Attention” IEEE Access, 2023.
5. GAN- grounded Deepfake Analysis
   Mittal, S. Agarwal, “Evaluation of GAN- grounded Facial Manipulation and Discovery ways” Elsevier Information Fusion, 2023.
6. ultramodern Deepfake Dataset Expansion
   Li et al., “ Celeb- DF v2 bettered Dataset for DeepFake Video Detection” CVPR Workshops, 2022.
   EfficientNet for Fake Image Discovery
7. Wu, X. Chen, “Fake Image Discovery Using EfficientNet with Attention Refinement” Neural Computing and Applications, 2022.

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