Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 3, pp 14-19, March 2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.303003
Available Online at https://www.jaiet.com

AI-Based Deepfake Video Detection Framework for Forensic Analysis & Digital Evidence Validation

[1]S Jubeda Banu, [2]K Irshad, [3]S Shafiya Bi, [4]M Vasudeva, [5]G Manikanta, [6]K C Yashwanth Kumar

[1,2,3,4,5,6]Department of Computer Science Engineering (Cyber Security), GATES Institute of Technology, Gooty, Andhra Pradesh, India
E-mail: [1]jubedashaik123@gmail.com, [2]irshadahmedk2004@gmail.com, [3]shafiyabi10@gmail.com, [4]vasudeva1791@gmail.com, [5]gmani9496@gmail.com, [6]chitrayaswanth202@gmail.com

Abstract: With the rapid advancement of Artificial Intelligence (AI), realistic deepfake videos are being created using deep learning techniques. The manipulated videos can cause significant harm in social, political, and security scenarios. Deepfake videos are created using deep learning techniques to change or replace a person's facial features, voice, and expressions. The detection of deepfake videos has become challenging with the improvement of deepfake video generation techniques. This project aims to develop a deep learning- based approach to detect deepfake videos using inconsistencies in facial features and patterns. The system integrates Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM). CNN is applied to extract spatial features from each video frame, focusing on facial features, while LSTM is applied to detect abnormal patterns in videos using temporal dependencies between frames. The model is trained and tested using the Celeb-DF dataset to recognize and classify real and deepfake videos accurately.

Keywords: Convolutional Neural Networks, Long Short-Term Memory, Deepfake Detection, Digital Media Forensics, Face Forensics, Transfer Learning.

I. INTRODUCTION

Forging of data is nothing new in this era having a backbone made up of artificial intelligence and machine learning. Distortion of reality is becoming a huge problem these days as more and more fake images and videos are emerging everyday on the internet. These deepfakes are getting better with time, to the extent that they cannot be distinguished as fake or real by the human eye, hence are increasingly resistant to detection. While deep-fake technology will bring with it certain benefits, it also will introduce many harms. In an era brimming with so much truth decay, nothing is more dangerous than people taking such videos at their face value. These deepfakes can be used for various malicious purposes such as defamation of celebrities, creating political bias, personal sabotage, intimidation and exploitation, false propaganda, piracy and other vengeful activities. The most targeted feature in a deepfake is the face. Thus, many algorithms and techniques can be used to identify manipulation of faces. These face manipulations can be of two types- Expressions and Identity. In the first type, the expressions of one person are transferred to another in real time. In identity manipulation, faces of two people are swapped. This type of manipulation can be used to spread false information among public by swapping with the face of a famous person.

In this paper, techniques are used to detect deepfakes using deep learning models such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) along with image preprocessing techniques. The system analyzes video frames to identify inconsistencies in facial features and motion patterns. The CNN model extracts spatial features from each frame, while the LSTM network captures temporal relationships between consecutive frames to detect abnormal facial movements. The model is trained on a dataset containing both real and fake videos and finally classifies the input as real or fake.

The algorithm works in the following steps:

Dataset Preparation:

A dataset of real and fake videos is collected and converted into frames.

Feature Extraction:

Faces are detected, aligned, and processed using a pre-trained CNN model (VGG16/OxfordNet) to extract facial features.

Temporal Analysis and Classification:

The extracted features are passed through an LSTM

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 3, pp 14-19, March 2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.303003
Available Online at https://www.jaiet.com

network to analyze frame sequences, and the trained model classifies the videos as real or fake after post-processing.

II. DEEPFAKE CREATION

There are numerous ways in which deepfakes can be created. The internet is flooded by softwares to create deepfakes that can be used by users with various technical skills ranging from novice to professional. It is so easy to create deepfakes that any layman can create them, having zero knowledge on the technicalities of the subject. The videos created may be very basic or elementary that can be recognised as fake by just a glance, to highly manipulated videos that cannot be detected even by a keen observer.

These deepfakes are created using artificial intelligence and deep learning methods. They rely on a type of convolutional neural network called auto-encoder which is used for encoding the input image by applying dimension reduction and image compression, and a decoder which reconstructs the image from the constructed representation by the encoder. The auto encoder is a self-supervising algorithm as it uses targets provided by itself to train on. An upgrade to this method is GAN, i.e. Generative Adversarial Network, an unsupervised deep learning algorithm, which further improves the quality of deepfake created.

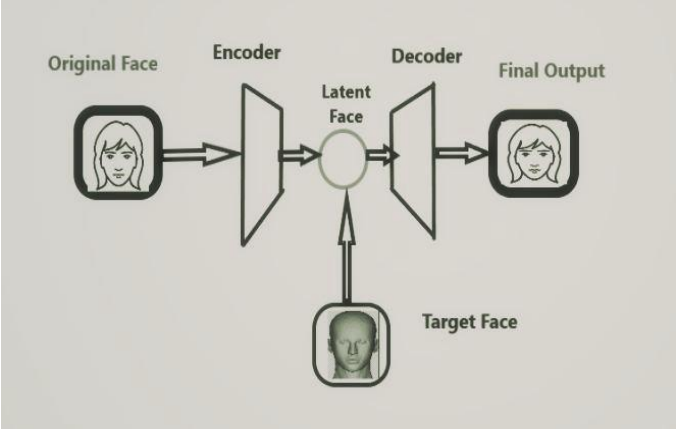


Figure 1: Deepfake Generation using auto encoder-decoder

A GAN is made up of two competing neural networks- generator and discriminator. The generator creates images as close to the real images as possible. The discriminator is then fed a training set containing both real and generated images and it tries to distinguish between them. As it continues to train, the generator makes images which are wrongly classified by the discriminator and the latter gets better at discriminating these images. So they form a pair that learn from each other and improve over time. In this way better quality deepfakes can be created. Deep Convolutional GANs (DCGANs) are even more effective as it uses convolutional layers to increase its efficiency.

An existing image of a person can be replaced with someone else's face by superimposing the latter onto the given image using artificial neural networks. There is also a method called FaceSwap which can be used to swap faces in manipulated images and videos. It uses image compression to adjust the superimposed image onto the given image. The colour of the two faces is also matched. The emerging artificial intelligence technologies can even replace expressions of one person to another in real time. The first application in the direction to create deepfake was FakeApp which allowed users to swap faces with other person. More such applications have been built over time such as FaceSwap, DeepFaceLab, DFaker, FaceSwap-GAN, DeepFake-tf, and many more.

III. RELATED WORK

This section aims to discuss and analyse various techniques that have been used for deepfake detection. Various attempts have been made to detect deepfakes which use deep learning at their core. These approaches work either on detecting faults in video or in separate frames of the video. Approaches which involve image analysis target various parameters like face warping artifacts [4], eye blinking rate [3] and head movements [6]. In 2018, "MesoNet" [5] was developed which used Inception model[11] to detect faults at mesoscopic level. Convolutional Neural Networks (CNN) have thus shown excellent feature extraction properties that can be used by a model to detect deepfake videos.

Various other approaches have used CNN along with other learning models like Recurrent Neural Network (RNN), Long Short-Term Memory Networks (LSTM) and Capsule Network[7] to further improve the accuracy by detecting temporal discrepancies and have shown good results on dataset containing videos generated by FaceSwap and deepfake.

Although various improvements have been made, more robust models are needed to detect deepfakes of lower quality and maintaining considerable accuracy remains a challenge with constantly improving methods in deepfake creation.

This paper aims to discuss successful feature extraction and processing required to detect discrepancies in the deepfake videos.

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 3, pp 14-19, March 2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.303003
Available Online at https://www.jaiet.com

IV. DEEPFAKE DETECTION

DeepFakes use specialized techniques that modify specific regions of the face, which are used as a base for superimposition. The algorithm works in a similar way for generating different deepfakes, often leaving certain discrepancies during the editing process. Factors such as compression artifacts, lighting variations, and temporal inconsistencies like abnormal lip and eye movements [3] can be used to train models for detecting deepfake videos. Among the various detection methods, Convolutional Neural Networks (CNN) have become a popular choice due to their strong capability in image and video analysis. CNNs can effectively extract spatial features from images, which helps in identifying manipulated facial regions.

To further improve detection, Long Short-Term Memory (LSTM) networks can be used along with CNN. While CNN extracts spatial features from individual frames, LSTM analyzes temporal relationships between consecutive frames, enabling the model to detect unnatural motion patterns and inconsistencies across the video sequence.

For practical implementation, transfer learning can be used to improve model performance. Transfer learning utilizes pre-trained neural network weights and fine-tunes them on a different dataset for a specific task. Several pre- trained models such as VGGNet, Xception, Inception, and ResNet are widely available. In this work, a fine-tuned model based on a pre-trained VGGNet is proposed for facial image analysis and deepfake detection. Frameworks such as Keras in Python provide efficient tools for implementing neural networks with transfer learning. The proposed model uses pre- processed video frames and applies transfer learning with a fine-tuned VGGNet along with an LSTM layer to effectively detect deepfake videos.

Uses pre-trained weights of the neural network for training a fine tuned version of same on different dataset for some specific application. Various pre-trained models have been made open source like VGG Net, Xception, Inception and ResNet. A fine-tuned model trained on a pre- trained convolution network like VGG Net is proposed which is specifically used for Image Analysis on human faces.

Various resources like keras for python provide easy functionality for implementing neural networks for transfer learning. The model uses the pre-processed frame (image) set from the original dataset and implements transfer learning on a fine tuned VGG Net model for detection of DeepFakes.

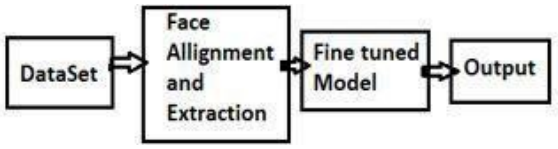


Figure 3: Process flow for deepfake detection

Dataset

The Celeb-DF dataset consists of 408 real and 795 synthesized videos that are made using modified DeepFake generation algorithm. The average length of the videos is 13 seconds with 30fps frame rate. The synthesized videos consist of low visual artifacts so that the quality of videos is greater than the previous datasets, where the videos are of high resolution and high number of visual artifacts that makes it easier to detect deepfakes. This dataset consists of lower quality deepfakes that makes the problem more challenging.

Table I: Contents of Dataset

1.Synthesized Videos	795
2.Real Videos used for creating deepfakes	158
3.Real Videos from Youtube	250

Figure 4: Generated fake samples from Celeb-DF dataset

Figure 5: Generated real samples from Celeb- DF dataset

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 3, pp 14-19, March 2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.303003
Available Online at https://www.jaiet.com

V. PROPOSED SYSTEM

This paper proposes a hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) based classifier for detecting deepfake videos using extracted video frames as input. The proposed system leverages CNN for spatial feature extraction and LSTM for temporal feature learning across sequential frames.

The video dataset is first pre-processed before training the model. The preprocessing stage involves face extraction and face alignment. Deepfake artifacts are mostly visible around facial boundaries; therefore, face extraction isolates the facial region from each video frame so that only relevant features are analyzed. Face alignment is used to normalize variations in head pose and orientation across frames in the video sequence.

After preprocessing, frames from each video are organized into temporal sequences and passed to the proposed CNN-LSTM architecture. The CNN component is responsible for extracting spatial features such as facial texture, color inconsistencies, and warping artifacts. Multiple convolutional layers followed by ReLU activation and max-pooling layers are used to capture hierarchical spatial representations of facial features.

The extracted feature maps from the CNN are then flattened and passed to the LSTM layer, which learns temporal dependencies between consecutive frames. Deepfake videos often contain temporal inconsistencies such as unnatural facial movements or flickering artifacts, which can be captured effectively by the LSTM network. The LSTM layer processes sequential frame features and learns long-term dependencies within the video.

Following the LSTM layer, batch normalization and dropout layers are introduced to improve generalization and reduce overfitting. Batch normalization stabilizes the learning process by normalizing the input distributions across layers, while dropout randomly disables neurons during training, thereby preventing the model from relying too heavily on specific features.

The final classification stage consists of a dense layer with two output nodes, representing the two classes: Real and Fake. A Softmax activation function is applied to produce probability scores for each class.

The model is trained using the Adam optimizer, which combines the benefits of momentum and adaptive learning rates. Adam provides faster convergence and improved performance compared to optimizers such as Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp). An optimal learning rate of 0.001 is used to ensure effective feature learning and stable convergence.

Transfer learning can also be incorporated by initializing the CNN layers with weights from a pretrained model trained on large image datasets. During training, the CNN extracts low-level spatial features such as edges, textures, and color patterns, while the LSTM captures temporal inconsistencies across frames. These features are analyzed to detect anomalies introduced during deepfake generation, including compression artifacts, facial warping distortions, and subtle color mismatches.

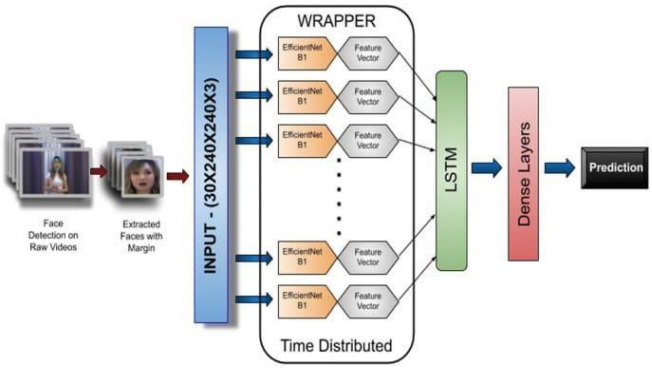


Figure 6: Architecture of the fake video detection

Fake Video Detection

CNN-LSTM Based Detection: While CNN models analyze individual frames, they cannot fully capture temporal inconsistencies across video frames. To address this limitation, hybrid architectures combining CNN and LSTM are used.

In this approach:

1. Video frames are extracted from the input video.

2. Faces are detected and aligned using libraries such as dlib.

3. A CNN network extracts spatial features from each frame.

4. The extracted features are passed to an LSTM network that analyzes frame sequences and learns temporal relationships.

5. A fully connected dense layer classifies the video as Real or Fake.

This hybrid model improves detection accuracy by analyzing both frame-level artifacts and temporal

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 3, pp 14-19, March 2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.303003
Available Online at https://www.jaiet.com

inconsistencies, such as unnatural facial movements or flickering artifacts between frames.

VI. RESULT ANALYSIS

The performance of the proposed CNN–LSTM based fake video detection model is evaluated using metrics such as accuracy, precision, recall, and F1-score. The model is trained on preprocessed video frames after face extraction and alignment.

During training, the accuracy increases and loss decreases, indicating effective learning. The CNN extracts spatial features from frames, while the LSTM captures temporal inconsistencies across video sequences.

The results show that the model successfully detects deepfake artifacts such as facial distortions, texture inconsistencies, and frame- level irregularities. Overall, the proposed CNN-

LSTM model provides reliable and accurate detection of fake videos.

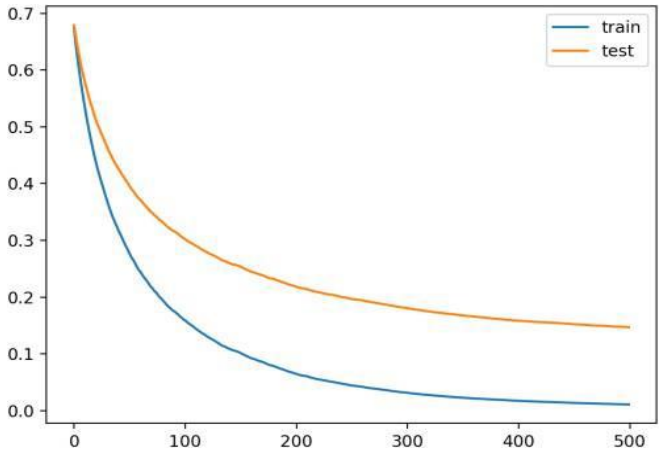
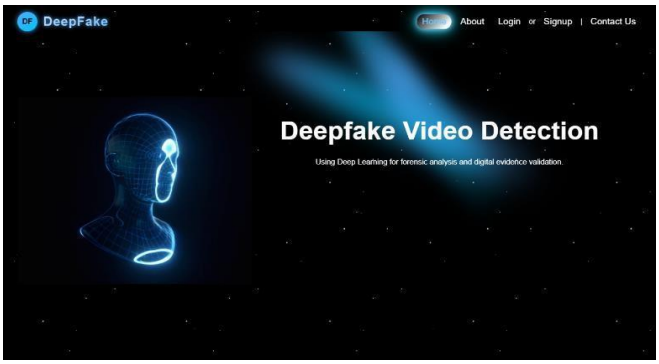
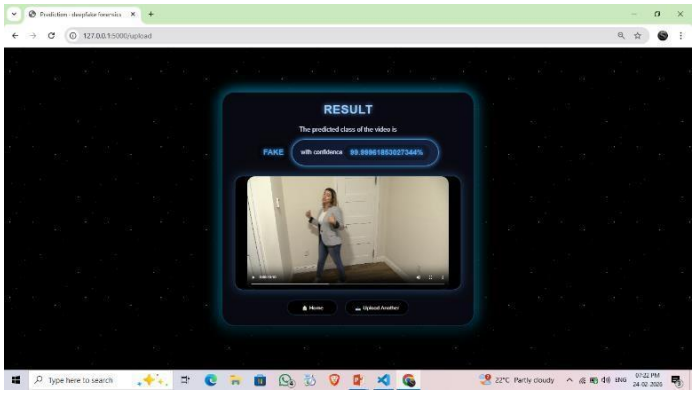


Figure 7: Test vs. Train curve for multiple epochs





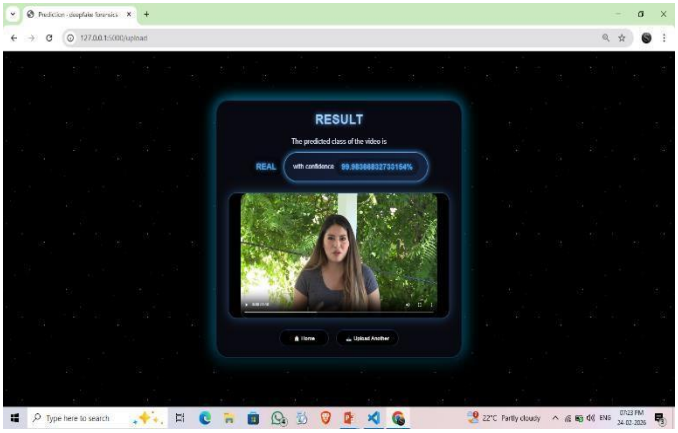


Figure 9: Detection of Fake and Real videos

VII. CONCLUSION

The results show that the CNN–LSTM architecture can successfully detect real and fake videos with good accuracy. By combining CNN for feature extraction and LSTM for sequence learning, the model provides a reliable approach for deepfake detection.

Future Research Directions:

Future work can focus on improving detection accuracy using larger datasets, advanced deep learning models, and real-time detection systems to handle more sophisticated deepfake videos.

REFERENCES

[1] Karen Simonyan and Andrew Zisserman, “Very Deep Convolutional Networks for Large – Scale Image Recognition”, ICLR 2015, arXiv:1409.1556v6 [cs.CV] 10 Apr 2015

[2] Chesney, Robert and Citron, Danielle Keats, Deep Fakes: A Looming Challenge for Privacy, Democracy, and

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 3, pp 14-19, March 2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.303003
Available Online at https://www.jaiet.com

National Security (July 14, 2018). 107 California Law Review (2019, Forthcoming); U of Texas Law, Public Law Research Paper No. 692; U of Maryland Legal Studies Research Paper No. 2018-21

[3] Yuezun Li, Ming-Ching Chang and SiweiLyu,In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking, arXiv:1806.02877v2 [cs.CV] 11 Jun 2018

[4] Yuezun Li, and SiweiLyu, "Exposing DeepFake Videos By Detecting Face Warping Artifacts", arXiv:1811.00656v3 [cs.CV] 22 May 2019. https://doi.org/10.1016/S0969-4765(19)30137-7

[5] DariusAfchar, Vincent Nozick, Junichi Yamagishi and Isao Echizen, "MesoNet: a Compact Facial Video Forgery Detection Network", arXiv:1809.00888v1 [cs.CV] 4 Sep 2018. https://doi.org/10.1109/WIFS.2018.8630761

[6] Xin Yang ,Yuezun Li and SiweiLyu, "Exposing Deep Fakes Using Inconsistent Head Poses" , ICASSP 2019 - 2019 IEEE ICASSP,17 May 2019.

[7] Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen, "Use of a capsule network to detect fake images and videos", arXiv:1910.12467v2 [cs.CV] 29 Oct 2019

[8] FalkoMatern , Christian Riess and Marc Stamminger, "Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations ", 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW). https://doi.org/10.1109/WACVW.2019.00020

[9] Jessica and Silbey Woodrow Hartzog , "The Upside of Deep Fakes", Maryland Law Review, Volume 78 issue 4,2019

[10] Schwartz, Oscar (12 November 2018). "You thought fake news was bad? Deep fakes are where the truth goes to die". The Guardian.

[11] Sik-Ho Tsang, "Review: Inception-v3 — 1st Runner Up (Image Classification) in ILSVRC 2015", https://medium.com/@sh.tsang/review-inception- v3-1st-runner-up-image-classification-in-ilsvrc- 2015-17915421f77c.

[12] Pavel Korshunov, Sebastien Marcel, "DeepFakes: a New Threat to Face Recognition? Assessment and Detection", citing arXiv: 1812.08685[cs.CV], 20 Dec 2018.

[13] David Güera,Edward J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks", 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS).

Citation of this Article:

S Jubeda Banu, K Irshad, S Shafiya Bi, M Vasudeva, G Manikanta, & K C Yashwanth Kumar. (2026). Anomaly Detection on Railway Track and Safety Monitoring Using Deep Learning. *Journal of Artificial Intelligence and Emerging Technologies (JAIET).* 3(3), 14-19. Article DOI: https://doi.org/10.47001/JAIET/2026.303003

***** End of the Article *****