IRJCS: International Research Journal of Computer Science ISSN: 2393-9842
Volume 12, Issue 11, November 2025 https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1211.08

Deep Fake Detection Using Deep Learning

Dr.Sumathi P
Department of CSE,
Sri Sairam College of Engineering, Bengaluru, India
sumathip.cse@sairamce.edu.in
Tanuja KN, Shri Lakshmi NK, Sahana S, Supritha PS
Department of CSE,
Sri Sairam College of Engineering, Bengaluru, India
sce22cs110@sairamtap.edu.in, sce22cs080@sairamtap.edu.in
sce22cs098@sairamtap.edu.in, sce22cs113@sairamtap.edu.in

Publication History
Manuscript Reference: IRJCS/RS/Vol.12/Issue11/NVCSXI10087
Research Article | Open Access | Double-Blind Peer Reviewed Article ID: IRJCS/RS/Vol.12/Issue11/NVCSXI10087
Received:23,October 2025, Revised: 09, October 2025, Accepted: 31October 2025 Published Online: 21November 2025
https://www.irjcs.com/volumes/Vol12/iss-11/08.NVCSXI10087.pdf
Article Citation:Dr.Sumathi,Tanuja,Shri,Sahana,Supritha(2025),Deep Fake Detection Using Deep Learning, IRJCS: International Research Journal of Computer Science, Volume 12, Issue 11 of 2025 pages 673-677
Doi:>https://doi.org/10.26562/irjcs.2025.v1211.08
BibTeX Key Dr.Sumathi@2025Deep
IRJCS papers should be cited as IRJCS (International Research Journal of Computer Science, AM Publications, India 2025, ISSN 2393-9842, https://doi.org/10.26562/ijrcs.2025.v1211.08 The journal's official abbreviation is IRJCS.
Orcid: https://orcid.org/0009-0004-9398-7488

Abstract: The emerging Deep fake Detection System has been a major threat to digital media integrity since its introduction in 2017; early applications of the technology have been largely based on applications of face- swapping techniques using auto encoders and generative adversarial networks (GANs).Owing to the fast pace of development of these synthetic media generation techniques, these methods have outpaced more traditional detection methods, leading to an urgent need for more adaptive solutions. The system creates a solution to the problem by designing a new self-learning detection system that does not require pre-existing datasets. Compared to traditional methods, which have to be provided with large amounts of labelled data, our implementation uses incremental machine learning algorithms that are incrementally refined based on user feedback and confidence-based learning mechanisms. By leveraging transfer learning and online learning paradigms, the platform learns to adapt to emerging deep fake variations in real-time. Results of current implementation show promising results in detecting artifacts from eye blinking patterns, facial micro-expressions and audio-visual synchronization errors. Further work will be done to incorporate transformer-based architectures for better temporal analysis, add text deep fakes detection. The project will serve to build an open source, continually improving defence against synthetic media manipulation that will further increase the authenticity and security of digital media.
Keywords: Res-Next Convolution neural network and also RNN and LSTM.

I. INTRODUCTION

The major AI threats in social media are deep fakes, which have been used for malicious intents: political distress, fake events, revenge porn, and blackmail. We are going to use AI in deep fake detection. Our approach will involve the usage of an LSTM-based neural network combined with a pre-trained Res Next CNN: for the analysis of video frames to extract features that assist in identifying whether it is a deep fake or real video. Our model is trained on large datasets provided by Face Forensic++, Deep fake Detection Challenge, and Celeb-DF that include a wide variety of deep fake videos. Abroad set of deep fake videos allows our model to learn various scenarios. The approach leverages strengths from both LSTMs and ResNext CNNs for effective deep fake detection. The developed front-end application will allow users to easily upload videos and get the results of the classification along with the confidence of the model regarding the result. This application simplifies everything and makes usage easy and time-consuming. But by pitting AI against AI, we also seek to mitigate some of the potential risks of deep fakes. We strive to give people a reliable way of detecting manipulated videos, in an attempt to restore faith in digital media. And as detection technology gets better, so will our defenses against deep fakes put to malicious use.

II. LITERATURE REVIEW

Deep fake detection mechanism has become a major area of research in leading industry in these recent years, leading to the development of numerous models and approaches leveraging deep learning. A research study on this has been published in the year 2022 titled "Deepfake Technology using different Methodologies": A literature survey paper has presented a detailed summary of deep learning-based strategies for identifying fake media. It has been summarized in different models such as CNN, RNN, LSTM, and hybrid architectures through evaluating them on benchmark datasets like Face Forensics++, Celeb-DF, and UADFV. The work provided a comprehensive comparison of detection algorithms, datasets, and performance metrics, offering valuable insight into the strengths and limitations of existing models.

ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1211.08

In another work of this project Deepfake Detection using the Deep Learning methods by research scholars in 2024, a hybrid CNN-LSTM framework was proposed for the detection of manipulated political videos. Temporal examination of each frame was done using a self-compiled dataset for training and validation. While temporal inconsistencies were well captured by this method, its scope was limited due to the usage of a small dataset and no benchmarking with other state-of-the-art models.

The 2022 paper "Deepfake Detection: A Systematic Literature Review" tabulated and analyzed results from 112 research papers published between 2018 and 2020. The paper classified the detection techniques into deep learning-based, machine learning-based, statistical, and block chain-oriented approaches. The provided review therefore gave a structured categorization and broad overview of the area but no recent research and experimental evaluation was included, hence limiting its applicability for current developments. Another important related work is "Deepfake Detection Using Deep Learning Methods: A Systematic and Comprehensive Review" (2023); it analyzed detection methods in several modalities: image, video, audio, and hybrid forms. It focused on CNN architectures, dataset diversity, and modern research directions according to 34 reviewed works.

This paper contributed vast categorization of detection techniques and related datasets but did not focus on real-time applications or deployment aspects. A 2024 contribution titled "Deepfake Video Detection Using RNN" proposed a two-stage CNN-LSTM framework that utilized InceptionV3 for extracting spatial features and LSTM for temporal analysis. This approach achieved an accuracy of nearly 97% on the restricted dataset of 600 samples, making this one of the very first models that showed high accuracy in detecting deep fakes temporally. However, this contribution faced limitations due to its restricted dataset and older architecture design compared to modern transformer- based systems. In 2023, "Implicit Identity Driven Deepfake Face Swapping Detection" proposed an identity-based approach to learn personal identity features rather than visual artifacts. This model effectively recognized inconsistencies between genuine and manipulated videos by implicit feature learning, making it resilient to common manipulations. However, this was more suitable for face-swapping scenarios and lacked adaptation in the cases of other types of manipulation.

The paper called "A Survey on Multiple data Deepfake Detection like detecting Audio, Video & Image": Tools, Trends, Challenges, and Future Directions" (2024) provided a broad overview of detection tools and techniques based on multimedia. Major challenges discussed included dataset bias and generalization issues, along with growing sophistication in fake content. The review, while highlighting the ongoing researchtrends and gaps, did not include experimental findings or propose new models. In “Hybrid AHA-PLO Metaheuristic Feature Selection for Robust Deepfake Video Detection,” further improvements were made in the publication year 2025 by developing a hybrid optimization technique of AHA combined with PLO to select the best features. It improves detection speed and builds accuracy while reducing redundant data. Results showed improved computational efficiency and increased model reliability; however, this system was developed and tested with just a few datasets, leaving scope for its broader evaluation.

A 2022 work, "Coot Bird Optimization-Based E Skip-ResNet Classification for Deepfake Detection," combined the CBO optimization with the E Skip-ResNet architecture. In this way, the model guaranteed faster convergence and better feature extraction, which resulted in a higher accuracy of classification on benchmark datasets. However, the real-world effectiveness of the model remained very uncertain because of limited validation outside a controlled environment. Lastly, “Deep Dect: A Facial Deepfake Video Detection Application Using Ensemble Learning” proposed, in 2024, a real-time detection system that employed ensemble learning, leveraging several deep learning models. This system leveraged both CNN and LSTM architectures in detecting spatial and temporal inconsistencies of reliable real-time performances. High dependency on hardware capability and model complexity remained its high limitation, which had implications for efficiency in deployment. Overall, these studies altogether contribute to the understanding and advanced level of deepfake detection through deep learning, ranging from the result of hybrid and optimized models to the demand for scalability, generalization, and real-world testing.

III. PROBLEM STATEMENT AND SOLUTION

Problem: With Convincing manipulations of digital images and videos have been demonstrated for several decades through the use of visual effects, recent advances in deep learning have led to a dramatic increase in the realism of fake content and the accessibility in which it can be created. These so-called AI-synthesized media (popularly referred to as deep fakes). With different concepts today the you this facing a lot due to these AI generated Images, Videos, Audio and many more. Due to these people are facing many kind of challenges like cyber bullying, blackmailing, identity stealing, Misusing of trust, Mental and physical health issues and etc... Creating deep fake images generating videos, has become very easy now a days. But recognizing them has been very difficult task towards the society. So we have come up with the solution that is our project called Deepfake Detection System.

Solution: In this regard a deep-learning-based detection framework could be designed that would take advantage of both convolutional neural networks for extracting spatial features and recurrent neural networks for extracting temporal features. Transfer learning and attention mechanisms would be helpful in extending its generalization Capability on unseen datasets. Furthermore, model parameter tuning through PSO or CBO can be integrated in order to achieve superior results. This brings better accuracy, robustness, and adaptability in identifying deepfake content across diverse video and image sources.

IRJCS: International Research Journal of Computer Science
Volume 12, Issue 11, November 2025
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1211.08

IV. METHODOLOGY

1. Motive to Use:

The approach is mainly motivated by devising a robust and intelligent deep learning frame work that is able to detect deep fakes from different datasets and manipulation types with high accuracy. Temporal dependencies inside the sequence of video frames are captured by an RNN, such as LSTM, combined with CNNs for spatial feature extraction. Transfer learning can be performed with pre-trained models like ResNet or Efficient Net to facilitate improved efficiency in learning with high accuracy. Besides, some meta heuristic-based optimization algorithms such as PSO are also utilized for the fine-tuning of hyper parameters toward optimal model performance. The proposed model is trained and tested using the bench mark dataset Face Forensics++ and DFDC, which ensures not only high precision but also good generalization and robustness against a variety of manipulations of deepfakes.

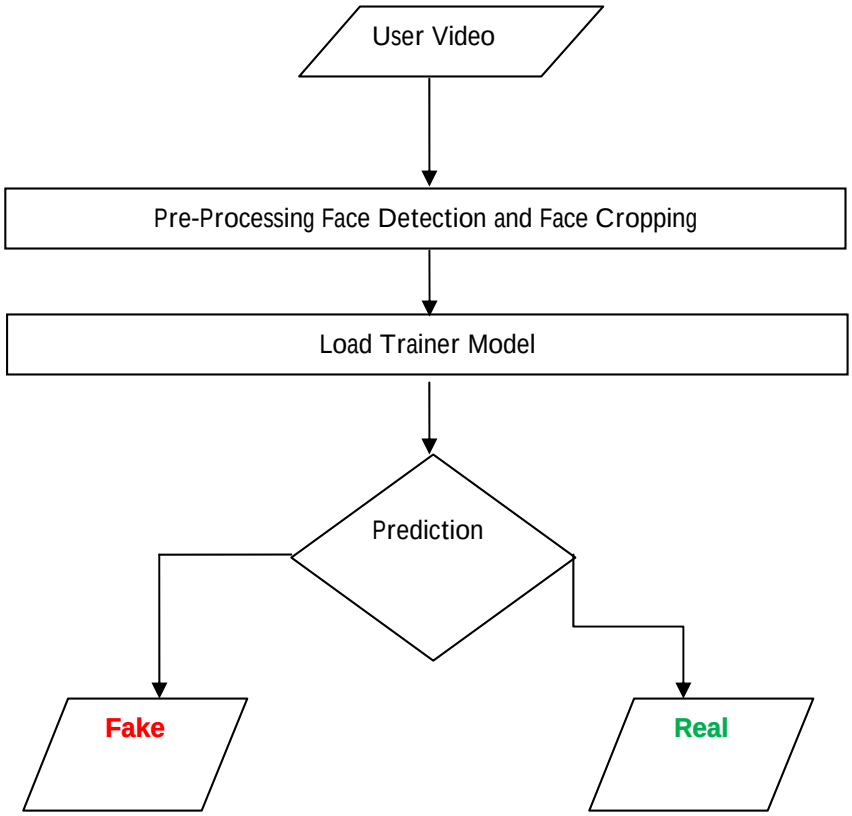

Fig1: Block Diagram

2. Flow of Real-Time Interaction

The actual flow of the deep fake detection system starts with the acquisition of data, where it receives the media input, say, a live video stream or an uploaded image. The Pre-processing of face detection, extraction of frames, resizing, and normalization for keeping all these things should be in formal pattern. Then, once the data is prepared, it goes through a feature extraction module, where the convolutional layer scapture intricate spatial features like skin texture, in consistencies in lighting, and facial land marks. Next comes temporal feature analysis involving models like LSTM or GRU focused on finding inconsistencies in motion or blinking patterns overtime. Then, the processed features are finally classified as "real " or "fake" with the help of a trained neural network classifier. The proposed system works in real time, leveraging GPU Acceleration and optimized model inference. Finally, the detection results will be represented through a user interface, highlighting manipulated regions or providing authenticity scores. Continuous monitoring of data will ensure adaptive learning and improvement with increasing time, hence enhancing overall reliability and accuracy.

3. Definition & System Requirements

The Deepfake detection refers to the automatic process of detecting and classifying manipulated digital media-such as synthetic video or images-created with deep learning models like GANs. It mainly tries to detect subtle in consistencies that could be induced by swapping faces, cloning voices, or modifying facial expressions. The system requires high-performance hardware to function effectively. This includes high-performance with CUDA operations support for fast deep learning computation. This system requires some of the software components include Python3.10+,Tensor Flow/PyTorch,OpenCV, and scikit-learn for feature processing and model training. It is recommended to run this software on a machine with at least 16GB RAM, while storage shall be ample enough to hold several datasets of interest, such as Face Forensics++,CelebF, or DFDC.

ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1211.08

Then morely requires of system having high data speed of internet which is supported to OS options will include Windows10/11, Linux Ubuntu, and macOS. Regarding security, encryption, firewalls, and secure logging of data are in place to ensure data integrity. In addition, for real-time deployment, Flask/Django APIs and cloud infrastructure such as AWS or GoogleCloud may be utilized to reach scalability and global accessibility.

4. Choose Components & System Modelling

The deepfake detection model is conceptualized as a multi-stage deep learning architecture. The first stage consists of input preprocessing, which detects facial regions from frames using algorithms like MT CNN or Dlib. Spatial feature extraction is carried out in the second stage using a CNN to capture facial artifacts, blending mismatches, and pixel-level discrepancies. Temporal modeling through RNNs or LSTM captures frame-to-frame variations due to video manipulations in the third stage. This is followed by a fully connected layer that combines spatial and temporal features to feed into a softmax classifier to output a probability score regarding the authenticity of the media. The optimization module utilizes PSO or Genetic Algorithms, among other algorithms, to obtain optimal hyper parameters. In addition, transfer learning by including pre- trained networks such as ResNet, Efficient Net, or Xception enhances generalization with minimum training time. The model architecture is completely trained and validated using bench mark datasets for stability across divers escenarios.

5. Develop and Integrate Code

The implementation step is where the deep fake detector system is developed in Python. Preprocessing of the data involves face detection, cropping, resizing, and normalization using OpenCV. These models can be created either with Tensor Flow or PyTorch by incorporating CNN for capturing spatial features and LSTMs or GRUs for temporal features.
This model can be improved in the future in generalization by incorporating the rotation, flipping, and the addition of noise during training. Then this model will be deployed by one of the Flask or Fast API, exposing a RESTful API that can be used in real-time. It uses Mongo DB/Postgre SQL for efficient storage of detection results along with meta data. Therefore the system architectural view is modular, updates of model weights or other components can be done independently without affecting the whole architecture. Version control with Github supports continuous integration, testing, and deployment via CI/CD pipelines that ensure scalability and ease of maintenance.

6. Building and Deploying the System

The development phase implements the Python programming language for data pre-processing scripts, utilizes the Open CV library to detect, crop, and normalize faces. Deep learning modules coded in Tensor Flow or PyTorch create CNN and RNN architectures. Hence, a training pipeline consists of data augmentation: rotation, flipping, and addition of noise to enhance model generalization. Once the model is trained and set up for integration into a real-world inference pipeline with either Flask or Fast API. This Phase of integration connects towards the different topics like- pre-processing, feature extraction, classification, and visualization into a single workflow. The architecture implements a RESTful API structure for communication between the detection model and the user interface. The integration of the database through Mongo DB or PostgreSQL considers efficient storage and results retrieval. The overall code base is modular and scalable, and any updates made to model weights or algorithms do not impact the overall architecture. Version control systems like Gitor CI/CD pipelines maintain continuous integration and deployment.

7. Test and Optimize

The final testing of the system for performance, accuracy, and reliability is done after its deployment. Unit testing, integration testing, and system validation will be carried out using both synthetic and real-world datasets. Data such as accurate prediction, precision checking, recall, F1-score, and ROC-AUC are computed for model evaluation. It also involves stress testing to access the system to work under many different conditions and cross-dataset validation to measure the general capacity of the project. Optimization steps will involve tuning hyper parameters of batch size, learning rate, dropping cost through techniques using grid search or evolutionary algorithms. Techniques to do model compression, such as pruning and quantization, will be applied so as to bring down the inference time with the least possible reduction in accuracy. Retraining models at regular intervals with new datasets gives relevant updates on evolving deep fake generation techniques. Moreover, feedback from test runs will be analyzed and continuous optimization cycles implemented. This will eventually result in a highly efficient, scalable, and accurate real-time deepfake technology that can identify even complex manipulations from diverse input types.

V. HARDWARE AND SOFTWARE

Hardware:

The instances will be used for the CPU servers or google clouds, utilized for the process of training and deploying a deep fake detection model. End-User Devices: These allow users to upload or stream videos using laptops, PCs, or smart phones. Network Infrastructure: This requires the stable internet connection for real-time processing and showing the accurate results.

Software:

Backend Requirements: Python with one of Flask or FastAPI for Application Program Interfance integration.
Deep Learning Framework: Tensor Flow or PyTorch is used for the construction and training of the CNN- LSTM models.
Frontend Requirements: A simple HTML,CSS, JS interface for uploading the media files and displaying results.
Database required: MongoDB or SQL is used to store the outputs and logs.
Libraries: Pre-processing, feature extraction, and analysis are done by the use of mathematical library functions.

IRJCS: International Research Journal of Computer Science
Volume 12, Issue 11, November 2025
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1211.08

VI. FUTURE SOLUTION

In the near future, deep fake detection solutions will be increasingly intelligent, adaptive, and transparent, thanks to the power of advanced deep learning techniques. This will be enabled through various approaches, including multimodal fusion and GNN, which will allow systems to process visual, audio, and contextual cues all at once, thereby opening ways for improvements in the detection precision across diverse media formats. Moreover, efforts will be directed at enhancing model generalization so that detectors remain good effective opposite to newly emerging and unseen deep fake generation methods. A part from this, explanations and transparency will be another vital area of development that enables the user and policy maker to understand how detection models make decisions, building trust in automated systems. This, in turn, entails a need for robust training and evaluation through large-scale and diverse real-world datasets. In particular, incorporating mechanisms of human-in-the-loop, whereby human judgment supplements AI predictions, will raise the level of precision and accountability substantially. These put together will result in robust, reliable systems that protect people, organizations, and society from the deleterious effects of deep fakes.

REFERENCE

1. Coot Bird Optimization-Based E Skip-Res Net Classification for Deepfake Detection Krishnan et al. (2022)
2. Deep fake Detection Algorithm Based on Improved VisionTransformer Heo,Y.J.,Yeo,W.H., & Kim, B. G. (2023)
3. Artificial Rabbits Optimization with Transfer Learning Based Deepfake Detection Model for Biometric Applications lazwari,S.,Alsamri,M.O.J.,Alamgeer,M.,Alabdan,R.,Alzahrani,I., Rizwanullah, M., & Osman, A. E. (2024)
4. DeepDect: A Facial Deepfake Video Detection Application Using Ensemble Learning Tay,W. X., Na Chua, H., Jasser, M. B., Issa, B., & Wong, R. T. (2024)
5. Video Deep fake Detection Using Improved Deep Neural Networks with Particle Swarm Optimization Cunha, L., Zhang, L., Sowan, B., Lim, C. P., & Kong, Y. (2024)
6. Deepfake Media Detection Using a Hybrid CNN–RNN Model and Particle Swarm Optimization PSO Algorithm Al-Adwan, A., Alazzam, H., Al-Anbaki, N., & Alduweib, E. (2024)
7. A Novel Block chain-Based Deep fake Detection Method Using Federated and Deep Learning Models Heidari, A., Navimipour, N. J., Dag, H., Talebi, S., & Unal, M. (2024)
8. Hybrid AHA-PLO Metaheuristic Feature Selection for Robust Deep fake Video Detection Koçak, A., & Alkan, M. (2025)
9. A Survey on Multimedia-Enabled Deepfake Detection: State-of-the-Art Tools and Techniques, Emerging Trends, Current Challenges & Limitations, and Future Directions Khan,A.A.,Laghari,
10. A.A.,Inam,S.A.,etal.(2025) Implicit Identity Driven Deep fake Face Swapping Detection Huang,B.,Wang,Z.,Yang,J.,Ai, J., Zou, Q., Wang, Q., & Ye, D. (2023)
11. Learning a Self-Supervised Domain-In variant Feature Representation for Generalized Audio Deep fake Detection Xie, Y., Cheng, H., Wang, Y., & Ye, L.(2023)
12. Temporal Feature Prediction in Audio-Visual Deep fake Detection Gao,Y.,Wang,X.,Zhang, Y., Zeng, P., & Ma, Y. (2024)
13. Deepfake Detection: Analysing Model Generalisation Across Architectures,Datasets and PreTraining Paradigms Khan, S. A., & Dang-Nguyen, D. T., 2023
14. Deepfake Detection Using Deep Learning Techniques Li,Y.,etal.(2022)
15. Deepfake Detection Based on Multi-Modal Fusion Zhang,Y.,etal.(2023)