International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 04 | Apr 2025 www.irjet.net p-ISSN: 2395-0072

DEEPFAKE DETECTION USING MACHINE LEARNING

Chandan Mahto[1], Prapti Khaparde[2], Priya Khatake[3] , Dr. Rajesh Kadu[4]

[1,2,3]*Student at Mahatma Gandhi Mission's College of Engineering And Technology, Navi Mumbai*
[4]*Associate Professor at Mahatma Gandhi Mission's College of Engineering And Technology, Navi Mumbai*

Abstract – ***With the increasing misuse of artificial intelligence for generating deepfakes in the form of fake images, cloned voices, and manipulated videos, ensuring media authenticity has become a significant challenge. This paper presents a unified machine learning-based multi-modal deepfake detection system capable of detecting forgeries in audio, image, and video formats. For image-based detection, a pipeline using MTCNN for face detection and InceptionResNetV1 for classification is used. Audio deepfakes are detected using CNN models trained on mel spectrograms derived from the ASVspoof 2019 dataset. For video analysis, a ResNet-based frame feature extractor and LSTM model are used to capture temporal inconsistencies, trained on the Celeb-DF dataset. All detection models are integrated into a single user interface using Streamlit, allowing users to input any media type and receive instant detection results. The system achieves high accuracy across all modalities and provides a practical, scalable solution for deepfake identification.***

***Key Words*: Deepfake Detection, Multi-Modal Detection, Audio Forgery, Image Manipulation, Video Deepfakes, CNN, LSTM.**

1.INTRODUCTION

The rapid advancement of AI-generated synthetic media popularly known as deepfakes has given rise to significant concerns around misinformation, identity theft, and manipulation of public discourse. Deepfakes can take the form of realistic synthetic videos, voice clones, or altered images that are often indistinguishable to the human eye or ear.

In this work, we propose a comprehensive machine learning-based system capable of detecting deepfakes in three distinct modalities: image, audio, and video. Our system is backed by well-known datasets (ASVspoof 2019 and Celeb-DF) and architectures (MTCNN, InceptionResNetV1, ResNet, LSTM), and it includes an intuitive UI built using Streamlit that allows users to upload the input which users wants to detect to check whether it is deepfake or not and respective model predict and gives the result back and display on the UI to users.

2. Motivation

In the old one or we can say previous deepfake detection or Most existing deepfake detection systems focus on a single domain, such as only audio or only video or only image. In real-world scenarios, however, manipulated content can appear in any format. The motivation behind this project is to build a multiple or separate model for respective , that is one for detect deepfake image , one for detect deepfake audio, one for detect deepfake video . This detection system that can detect deepfakes regardless of the input type and deliver results to the user in a simple and interactive way.

The primary goals of this system are:

1. To develop an accurate detection model for each media type that help the users to detect the deepfake image , audio ,video so that user can know the reality hidden behind the content and aware the public and save them from this manipulated content traps and make them responsible person .

2. To integrate all detection pipelines into a single user interface. Previously the user have to use different UI for detect other content , as there is no single UI where user can give image , audio , video at one place to detect the reality hidden behind the content . So we make the single UI where, user can give the input for respective model and can detect the content or input given by users to know whether it is real or deepfake.

3. To build a practical tool for general users, where users become the more informative regarding the deepfake content. This prevent the users from the traps of deepfake content . It is also makes the society more responsive regarding the deepfake content as the society people can check the reality of content and knows the real picture of the contents. The all three model are useful and give the accurate result as per the users input and response time is also less.

4. Aware the society regarding the deepfake content which is created using the newly emerged technologies like AI and ML. This deepfake detection system gives the users more clarity for the deepfake contents and gives them better idea for prevention and their relatives and other people from the menace situation.

3. Literature Survey

Andreas et al [1] this paper examines the realism of state-of the-art image manipulations, and how difficult it is to detect them, either automatically or by humans. After the collecting data it is manipulated, then the image is detected whether it is fake or real using CNNs convutional neural networks.

Yuezun Li et al [2] The need to develop and evaluate Deep Fake detection algorithms calls for large-scale datasets. However, current Deep Fake datasets suffer from low visual quality and do not resemble Deep Fake videos circulated on the Internet. The use of DNNs has made the process to create convincing fake videos increasingly easier and faster. In this work, they present a new large-scale and challenging Deep Fake video dataset, Celeb-DF3, for the development and evaluation of Deep Fake detection algorithms.

Brian et al [3] The DFDC is the largest currently and publicly available face swap video dataset. The dataset contains over 100,000 clips from 3,426+ paid actors. The dataset is created using several Deep fakes and GAN-based and non-learning techniques.

Kaede et al [4] In order to identify deep fakes, we introduce in this paper new synthetic training data dubbed self-blended images (SBIs). To replicate forging artifacts, SBIs are created by merging source and target photos that have been marginally altered from one authentic image.

Nicol'o et al [6] Take up the challenge of detecting face alteration in video sequences that use contemporary facial manipulation methods. Using more than 10,000 videos, the CNN approach is used to recognize false videos.

[7] proposes a method for detecting the appearance of facial forgery, which is used at the level of mesoscopic analysis. In fact, microscopic research based on image noise becomes illegal in the case of video with image noise degradation after video compression. Similarly, it is difficult for the human eye to classify fake images at a higher level, especially when images show human faces. Therefore, it is recommended to use a deep neural network with a sufficient number of layers as an intermediate method.

A deep learning framework was employed by the authors of the study [8] for audio-deep fake detection . The model separability is increased using a Long-short term memory (LSTM)-the based network is used to recognize events in sub-sampled signals

[10]proposed a method for using residual noise to be the difference between the original image and its noise free version. Residual noise has been shown to be useful in deep sensing due to its specificity and discrimination, which can be achieved through neural networks with adaptive learning. The method was tested on two datasets: low resolution FaceForensics++ videos and high resolution videos from the Kaggle Deepfake Detection Challenge (DFDC). In this article, we propose an adaptive learning based classifier that uses convolutional neural networks to learn the noise of real and fake videos.

4. Existing System

Currently existing system is not use single UI to display the result for audio, video, image detection. Users have to use different UI and model for detect the audio , image , video .

1. **Audio Deepfake Detectors**: Rely mostly on signal processing and spectrogram analysis and it is single model where user can give only input as audio and not other input like image.

2. **Image Detectors**: Use face detection followed by classification but don't address other media types where users have to face problem if he wants to detect other contents.

3. **Video Detection Models**: Focus on frame-level analysis or short sequences using LSTMs or 3D CNNs and it is not can be used for image or audio detection,

Limitations:

- Lack of a unified system that supports all input types.
- No interactive UI for real-time testing.
- Performance may degrade when deepfakes come from newer generation models not present in training data.

Overall while there are many existing system for Deepfake detection , many either lack real time adaptability or they might require complex infrastructure or may not provide user friendly outputs for everyday users to detect the contents.

5. Proposed System

The proposed system aims to a more advanced and an efficient way to give the users deepfake detection system for helps them to detect the deepfake content . So that users , will not have to face issue for detect the deepfake content and as we develop the single UI based on streamlit where users can give their desire content to check the reality hidden behind the content that means it will know the users that whether the given input is real or deepfake.

So basically , proposed system is a multi-modal deepfake detection platform built using machine learning and deep learning techniques, integrated with a Streamlit-based user interface.

Components:

Image Detection Module: Uses MTCNN for face detection and InceptionResNetV1 for binary classification (real/fake).This module help the users to give the input as image and this module on the basis of provided content it will display the result after detection , whether the provided content is real or not.

Fig 1.Real to Fake image

Audio Detection Module: Converts input audio into mel spectrograms and classifies them using a CNN model. Under this module, users can give the audio as input to detect the audio reality , whether the audio is deepfake or not.

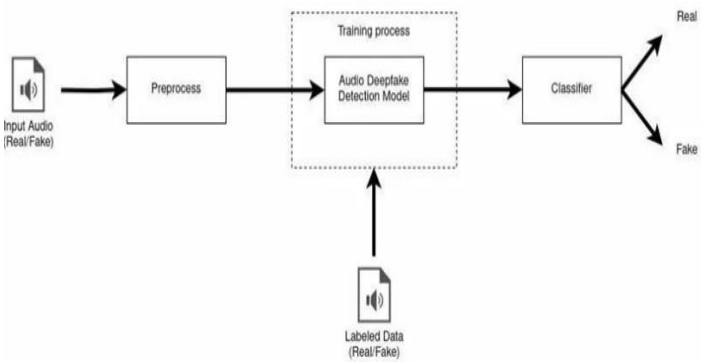

Fig 2. Audio detection Process

Video Detection Module: Extracts frames, uses ResNet to extract features, and applies LSTM to model the temporal relationship for classification. It will allow the users to detect the video , to know whether the given video is manipulated or not.

User Interface:

Built using Streamlit to allow users to:

1. Upload image/audio/video files
2. View predictions instantly

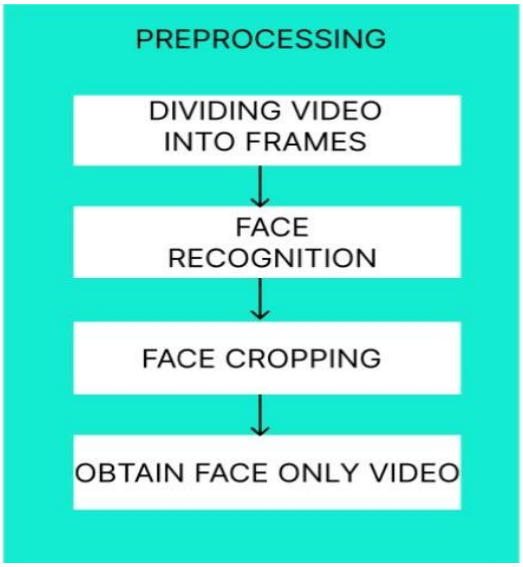

Fig 3: Video Detection Process

All three modules , are trained on respective required dataset , and gives the accuracy above 85, That helps the users to know the content is real or manipulated via new emerging technologies.

6. Methodology

6.1 Audio Deepfake Detection:

Dataset: For trained the model we used the ASVspoof 2019 Logical Access Dataset.
Preprocessing: Once the model is trained it will accept the audio as input and Audio converted into mel spectrograms for further analysis and result.
Model Architecture: CNN with convolutional, pooling, and dense layers.

6.2 Image Deepfake Detection

Face Detection: MTCNN to locate and crop facial regions. Once the image is cropped then it transfer to the model to detect the reality of image , whether is manipulated or not.
Model: InceptionResNetV1 fine-tuned for deepfake classification.

6.3 Video Deepfake Detection:

Dataset: For training the video detection model Celeb-DF dataset is used.

Preprocessing: Once the model is trained it accepts the video as input and Extract video frames and faces to check the reality of video.
Model: ResNet for spatial feature extraction; LSTM to capture temporal dynamics.

6.4 Streamlit UI Integration:

A simple and lightweight web interface which helps the users to upload and detect the content reality on one place . User selects input type (image/audio/video), uploads the file, and views the result in real time.

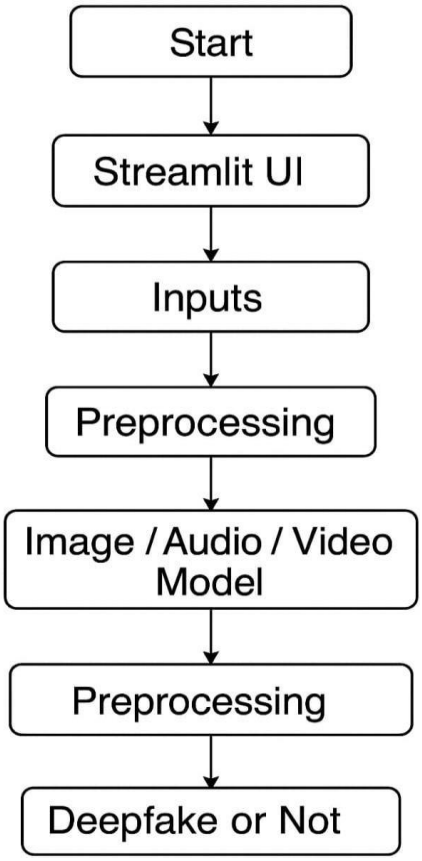

Fig 4. Overall Flowchart of Project.

7. Result

Modality	Dataset	Accuracy (%)
Image	Custom Dataset	92.3
Audio	ASVspoof 2019	84.7
Video	Celeb-DF	87.5

The image model shows high sensitivity to facial manipulation artifacts .The audio model performs well on various types of voice spoofing .The video model successfully captures both frame-level and sequence- level manipulation. Performance metrics, including precision, recall, and F1-score, were also calculated for each model. The results indicate that the system is highly effective at distinguishing between real and fake media, with minimal false positives and negatives.

We built an interactive web application using Streamlit, fully integrated with the trained model. The web application allows users to:

- Give the input as image,audio,video.
- Optionally provide spefic details for the image using grad-cam.
- Display the result as fake or real.

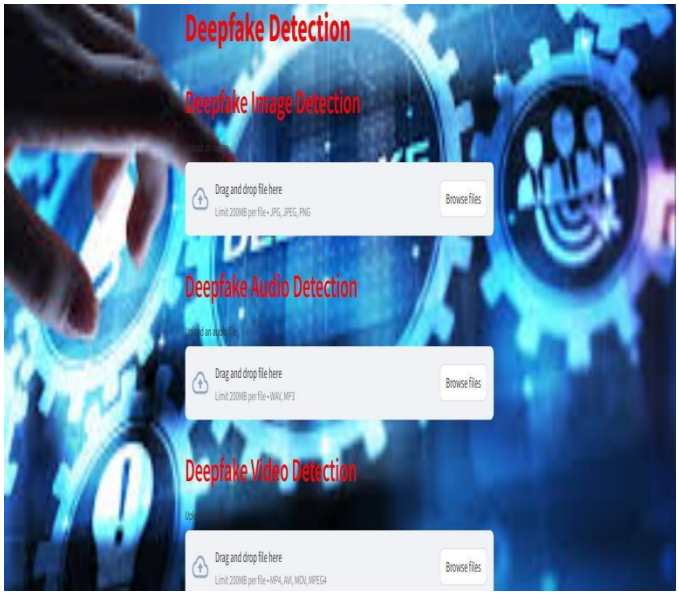

Fig 7.1: User Interface I

Fig 7.2: Image detection prediction result

Fig 7.3: Audio detection result

Fig 7.3: Video detection result

8. Conclusion

This paper introduces a unified deepfake detection system that supports audio, image, and video inputs. By using specialized machine learning models for each type and integrating them into an easy-to-use Streamlit interface, the system becomes a practical tool for detecting deepfakes in real-world scenarios. The multi-modal approach enhances robustness, and results show strong performance across all tested media formats. The system demonstrated high accuracy and real-time performance, making it a valuable tool for applications in security, media, and social platforms. The proposed approach represents a significant step toward creating unified, scalable solutions for deepfake detection across various media formats. This models make the society people more informative and responsible to reduce the fake content proliferation from the society and world and eliminate the spread of defamation on any person.

9. Future Scope

A) Model Fusion: Combine the outputs of different models for ensemble decision-making.

B) Cross-Dataset Generalization: Improve robustness against unseen data from other deepfake generation techniques.

C) Mobile App: Deploy lightweight versions for mobile use.
D)
D) Real-Time Streaming Detection: Extend video detection to live streams.

E) Explainability: Add explainable AI features to visualize manipulated regions or suspicious audio segments.

F) Models performance can be enhanced so it can be more accurate to detect the content reality.

10. References

[1] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Nießner. (2019). FaceForensics++: Learning to Detect Manipulated Facial Images. IEEE Conference Publication

[2] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi and Siwei Lyu. (2020). Celeb-DF:A Large-scale Challenging Dataset for Deep Fake Forensics. IEEE Conference Publication

[3] Brian Dolhansky, Joanna Bitton, Ben Pflaum, Jikuo Lu, Russ Howes, Menglin Wang, Cristian Canton Ferrer. (2020) The Deep Fake Detection Challenge (DFDC) Dataset. arXiv:2006.07397 Vol4

[4] Kaede Shiohara Toshihiko Yamasaki. (2022) Detecting Deep fakes with Self-Blended Images IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)

[5] Tolosana, R., Vera-Rodríguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection. ArXiv:2001.00179.

[6] Nicol'o Bonettini, Daniele Cannas, Sara Mandelli, Luca Bondi, Paolo Bestagini, Stefano Tubaro. (2020) Video Face Manipulation Detection Through Ensemble of CNNs. 25th International Conference on Pattern Recognition (ICPR)

[7] Darius Afchar, Vincent Nozick, Junichi Yamagishi and Isao Echizen. MesoNet: a Compact Facial Video Forgery Detection Network in arXiv:1809.00888v1 [cs.CV] 4 Sep 2018.

[8] A. Abbasi, A. R. R. Javed, A. Yasin, Z. Jalil, N. Kryvinska, and U. Tariq,"A large-scale benchmark dataset for anomaly detection and rare event classification for audio forensics," IEEE Access, vol.10, pp. 38885–38894,
2022.

Abbasi et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS [9] Z. Khanjani, G. Watson, and V. P. Janeja, "How deep are the fakes? focusing on audio deepfake: A survey," arXiv preprint arXiv:2111.14203,
2021.

[10] A. Malik, M. Kuribayashi, S. M. Abdullahi and A. N. Khan, "DeepFake Detection for Human Face Images and Videos: A Survey," in IEEE Access, vol. 10, pp. 18757- 18775, 2022, doi: 10.1109/ACCESS.2022.3151186. S. Hochreiter and J. Schmidhuber, "Long short-term memory," N.

[11] Raza A, Munir K, Almutairi M, "A novel deep learning approach for deepfake image detection" Applied Sciences 2022 Sep 29.

[12] Suratkar S, Kazi F, "Deep fake video detection using transfer learning approach" Arabian Journal for Science and Engineering. 2023 Aug 2021

[13] Khalil, Hady A., and Shady A. Maged. "Deepfakes creation and detection using deep learning." 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) IEEE, 2021

[14] Gupta G, Raja K, Gupta M, Jan T, Whiteside ST, Prasad M. "A Comprehensive Review of DeepFake Detection Using Advanced Machine Learning and Fusion Methods" Electronics. 2023 Dec 25 2020

[15] Passos LA, Jodas D, Costa KA, Souza Júnior LA, Rodrigues D, Del Ser J, Camacho D, Papa JP. "A review of deep learning-based approaches for deepfake content detection" Expert Systems. 2022

[16] Chen B, Li T, Ding W. "Detecting deepfake videos based on spatiotemporal attention and convolutional LSTM". Information Sciences. 2022 Jul 1

[17] Masud U, Sadiq M, Masood S, Ahmad M, Abd El-Latif AA. "LW-DeepFakeNet: a lightweight time distributed CNNLSTM network for real-time DeepFake video detection" Signal, Image and Video Processing. 2023 Nov;17

[18] Saikia, Pallabi, et al. "A hybrid CNN-LSTM model for video deepfake detection by leveraging optical flow features." 2022 international joint conference on neural networks (IJCNN). IEEE, 2022 .

[19] Al-Dhabi, Yunes, and Shuang Zhang. "Deepfake video detection by combining convolutional neural network (cnn) and recurrent neural network (rnn)." 2021 IEEE international conference on computer science, artificial intelligence and electronic engineering (CSAIEE). IEEE, 2021

[20] Zhang T. "Deepfake generation and detection, a survey" Multimedia Tools and Applications. 2022 Feb

[21] Rebello, Lian, et al. "Detection of Deepfake Video using Deep Learning and MesoNet." 2023 8th International Conference on Communication and Electronics Systems (ICCES). IEEE,