International Conference on Advances in Modern Technology of Research in Engineering Field (AIMTREF) April, 2026
Organized By: SKN Sinhgad Institute of Technology& Science, Kusgaon (Bk), Lonavala, Pune

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

Development of AI/ML Based Solutions for Deep Fake Detection

Prof. Chandani Lachke, Anushka Singh, Snehal Bhosarkar, Pradnya Jambhale
Computer Engineering, SKN Sinhgad Institute of Technology & Science, Lonavala,,

Absvtract- Deep learning has proven effective in a variety of tough issues, including computer vision, human-level control, and large data analytics. However, as deep learning technology advanced, software was developed that jeopardized national security, democracy, and privacy. Deepfake is a new technology that uses deep learning to create fake photos and videos that look very real. It's important to have tools that can automatically detect and check the quality of these AI-created images and videos. These systems help us quickly tell if a picture or video is real, edited, or fake, and they ensure that the quality is good and not misleading. An investigation of the strategies used to construct the most significant deepfakes, as well as the approaches proposed in the literature for detecting them. We provide a complete examination of the difficulties highlighted by deepfake technology, as well as recommendations for future and upcoming research opportunities. It also supports creating new and more reliable ways to handle deepfakes as they become more complex.

Keywords- CNN, Deep Learning, Face Detection, Face Recognition, Feature Extraction, Pre-Processing

I. INTRODUCTION

In recent years, Artificial Intelligence (AI) has revolutionized the way digital media is created and consumed. One of the most notable advancements is the emergence of deepfake technology, which uses deep learning algorithms to manipulate or generate realistic human faces in images and videos.

Deepfakes are created using techniques such as Generative Adversarial Networks (GANs) and autoencoders, which can swap faces, alter expressions, or synthesize entirely fake identities. While this technology has beneficial applications in entertainment and media, it also introduces serious risks including misinformation, identity theft, cybercrime, and political manipulation.

The increasing accessibility of deepfake tools has made it essential to develop robust detection systems. Traditional image processing techniques are insufficient to detect sophisticated manipulations. Therefore, AI/ML-based solutions have emerged as a powerful approach for identifying subtle inconsistencies in deepfake media.

This paper focuses on designing and implementing an AI/ML-based deepfake face detection system capable of accurately distinguishing real and fake images.

II. PROBLEM STATEMENT

The widespread use of deepfake technology has created significant challenges in verifying the authenticity of digital content. Existing detection methods often fail when dealing with high-quality deepfakes or unseen datasets.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

The problem addressed in this research is:

- To develop a robust and efficient AI/ML-based system capable of detecting deepfake faces with high accuracy.
- To ensure generalization across multiple datasets and real-world scenarios.
- To minimize false positives and false negative.

Objectives

- The main objectives of this research are:
- To design a deep learning-based model for deepfake face detection.
- To preprocess and extract facial features effectively.
- To train and evaluate the model using standard datasets.
- To improve detection accuracy using hybrid architectures.
- To analyse performance using evaluation metrics.

III. LITERATURE REVIEW

Researchers have explored multiple approaches for deepfake detection over time. Early techniques focused on identifying visible artifacts such as unnatural eye blinking, facial distortions, and inconsistent lighting conditions. While these methods were useful for detecting low-quality manipulations, they often failed against more advanced deepfake content.

With the advancement of deep learning, Convolutional Neural Network (CNN)-based models significantly improved detection performance by learning spatial features and identifying subtle inconsistencies in facial structures. To further enhance detection in videos, researchers introduced Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models, which analyse temporal patterns across frames to detect irregular motion or unnatural transitions.

More recent approaches include Capsule Networks, which aim to preserve hierarchical relationships between facial features, and hybrid models that combine both spatial and temporal analysis for better accuracy. Despite these advancements, detecting highly realistic deepfakes remains challenging, particularly when models are tested on unseen datasets, highlighting the need for more robust and generalized solutions.

IV. PROPOSED METHODOLOGY

System Overview

The proposed system consists of the following stages:

- Data Collection
- Preprocessing
- Feature Extraction
- Model Training
- Evaluation

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

Data Collection

Datasets used:

- FaceForensics++
- Celeb-DF
- DFDC (DeepFake Detection Challenge dataset)

These datasets contain both real and manipulated videos/images.

Preprocessing

Steps involved:

- Frame extraction from videos
- Face detection using MTCNN
- Face alignment and cropping
- Image resizing (e.g., 224×224)
- Data augmentation (rotation, flipping, noise addition)

Feature Extraction

The system extracts both:

- Spatial Features: Using CNN layers
- Temporal Features: Using sequence models (LSTM/GRU)

V. PROPOSED ARCHITECTURE

The proposed method's enhanced Convolutional Neural Network (CNN) architecture addresses problems in prior maps by deepfake detection methods. The CNN's ability to detect minute abnormalities and fine-grained spatial properties in and images and videos enables it to generalize more reliably across various types of deepfakes.

Convolutional Layers: These layers filter the input data and apply convolution to extract valuable properties. Each filter from the previous layer to the next. They classify based on the (CNNs) have played a key role in improving Deep Learning for computer vision. CNNs are special computer systems that learn from data and are particularly good at understanding images and videos. CNNs are great at analyzing visual data, like images and videos, because they automatically find recognizes unique patterns or features in data.

Pooling Layers: These layers decreases size of the feature making the data smaller after the convolution layers. Two common methods for doing this are maximum pooling and average pooling, which simplify the information by across picking the most important values or averaging them.

Flatten: Flattening is when data is turned into a long list of multi- numbers so it can be passed to the next step in the model. It increases takes the output from the previous step and arranges it into one straight line of features. images, Fully Connected Layers: Dense layers connect every neuron features obtained from previous levels.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

System Architecture

- Data Collection and Preprocessing: The dataset contains both real and fake videos.
- Feature Extraction: The video is broken down into individual pictures (frames). Then, faces are found in each of these pictures.
- Model Architecture: This part of the network is responsible for picking out important details from the cropped face images.
- Training: The model is taught using a set of training data that has the important features. The goal is for the model to learn the differences between real and fake videos by recognizing patterns in them.
- Output: The trained model is used to determine whether a new video is real or fake. The model looks at the features taken from the new video to make this prediction.

Advantages

- High detection accuracy
- Works on both images and videos
- Scalable and adaptable
- Robust against unseen data

Limitations

- Requires large datasets for training
- Computationally expensive
- May struggle with extremely high-quality deepfakes

Implementation

The proposed deepfake detection system is implemented using the Python programming language with the help of popular machine learning and image processing libraries. Deep learning frameworks such as TensorFlow or PyTorch are used to build and train the model, while OpenCV is utilized for video processing and face extraction.

Initially, video data is converted into frames, and faces are detected using algorithms like MTCNN. The extracted faces are then resized and normalized to a fixed dimension suitable for model input. The processed data is divided into training and testing sets.

A Convolutional Neural Network (CNN), optionally combined with an LSTM layer, is trained on the dataset to learn distinguishing features between real and fake faces. The model is trained using an optimizer like Adam and evaluated using performance metrics such as accuracy and precision.

Finally, the trained model is tested on unseen data to verify its effectiveness in detecting deepfake content.

VI. CONCLUSION

Deepfake technology poses a serious threat to digital trust and security. This paper presented an AI/ML-based approach for detecting deepfake faces using deep learning techniques. The proposed system

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

demonstrates high accuracy and robustness by combining spatial and temporal feature extraction. While challenges remain, continuous advancements in AI can help build more reliable detection systems to combat the misuse of deepfake technology.

REFERENCES

1. [1] Patel, Yogesh, Sudeep Tanwar, Rajesh Gupta, Pronaya Bhattacharya, Innocent Ewean Davidson, Royi Nyameko, Srinivas Aluvala, and Vrince Vimal. "Deepfake generation and detection: Case study and challenges." IEEE Access (2023).
2. Rana, Md Shohel, Mohammad Nur Nobi, Beddhu Murali, and Andrew H. Sung. "Deepfake detection: A systematic literature review." IEEE access 10 (2022): 25494-25513.
3. Lewis, John K., Imad Eddine Toubal, Helen Chen, Vishal Sandesera, Michael Lomnitz, Zigfried Hampel Arias, Calyam Prasad, and Kannappan Palaniappan.
4. "Deepfake video detection based on spatial, spectral, and temporal inconsistencies using multimodal deep learning." In 2020 4. IEEE Applied Imagery Pattern Recognition Workshop (AIPR), pp. 1- 9. IEEE, 2020.
5. Trinh, Loc, Michael Tsang, Sirisha Rambhatla, and Yan Liu. "Interpretable and trustworthy deepfake detection via dynamic prototypes." In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 1973-1983. 2021.
6. S. P and S. Sk, "DeepFake Creation and Detection
7. :A Survey," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2021, pp. 584-588, doi: 10.1109/ICIRCA51532.2021.9544522.
8. D. Pan, L. Sun, R. Wang, X. Zhang and R. O. Sinnott, "Deep fake Detection through Deep Learning," 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT), Leicester, UK, 2020, pp.134-143, doi: 10.1109/BDCAT50828.2020.00001.
9. A. Malik, M. Kuribayashi, S. M. Abdullahi and A.
10. N. Khan, "DeepFake Detection for Human Face Images and Videos: A Survey," in IEEE Access, vol. 10, pp.875718775,2022,doi:10.1109/ACCESS.2022.315118