IJIRIS: International Journal of Innovative Research in Information Security
Volume 10, Issue 04, May 2024
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives

# Revolutionizing Deepfake Detection and Realtime Video Vision with CNN-based Deep Learning Model

**Pavan Kumar R B,**
Assistant Professor, Dept. of CSE
East Point College of Engineering and Technology,
Bengaluru, Karnataka, INDIA
pavankumarrb@gmail.com
**Pruthvi H V**
Department of Computer Science Engineering
East Point College of Engineering and Technology,
Bengaluru, Karnataka, INDIA
pruthvipruthvi9342@gmail.com
**Salma Begum N**
Department of Computer Science Engineering
East Point College of Engineering and Technology,
Bengaluru, Karnataka, INDIA
salmabegumn13@gmail.com
**Shreya C**
Department of Computer Science Engineering
East Point College of Engineering and Technology,
Bengaluru, Karnataka, INDIA
shreyakunder17@gmail.com
**Sujana Yadav S**
Department of Computer Science Engineering
East Point College of Engineering and Technology,
Bengaluru, Karnataka, INDIA
sujanasyadav@gmail.com

**Publication History**
**Manuscript Reference No:** IJIRIS/RS/Vol.10/Issue04/MYIS10089
Research Article | Open Access | Double-Blind Peer-Reviewed | Article ID: IJIRIS/RS/Vol.10/Issue04/MYIS10089
Received: 07, April 2024|Revised:16, April 2024 | Accepted: 27, April 2024 Published Online: 07, May 2024 Volume 2024
Article ID MYIS10089 http://www.ijiris.com/volumes/Vol10/iss-04/10.MYIS10089.pdf
**Article Citation:**Pavan,Pruthvi,Salma,Shreya,Sujana(2024). Revolutionizing Deepfake Detection and Realtime Video Vision with CNN-based Deep Learning Model. International Journal of Innovative Research in Information Security, Volume 10, Issue 03, Pages 173-177
**doi:>** https://doi.org/10.26562/ijiris.2024.v1004.10
**BibTex key:** Pavan**@2024Revolutionizing**

Copyright: ©2024 This is an open access article distributed under the terms of the Creative Commons Attribution License; which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Abstract:** The rapid advancement of deep learning models that can generate and synthesis hyper-realistic videos known as Deepfakes and their ease of access have raised concern on possible malicious intent use. Deep learning techniques can now generate faces, swap faces between two subjects in a video, alter facial expressions, change gender, and alter facial features, to list a few. These powerful video manipulation methods have potential use in many fields. However, they also pose a looming threat to everyone if used for harmful purposes such as identity theft, phishing, and scam. In this work, we propose a Convolutional Vision Transformer for the detection of Deepfakes. The Convolutional Vision Transformer: Convolutional Neural Network (CNN) and Vision Transformer (ViT). The CNN extracts learnable features while the ViT takes in the learned features as input and categorizes them using an attention mechanism. We trained our model on the DeepFake Detection Challenge Dataset (DFDC) and have achieved 91.5 percent accuracy, an AUC value of 0.91, and a loss value of 0.32. Our contribution is that we have added a CNN module to the ViT architecture and have achieved a competitive result on the DFDC dataset.

**Keywords:** DeepFake, Auto Encoder, Convolutional Neural Network (CNN), Artifical Neural Network (ANN), Vision Transformer(ViT), Convolutional Vision Transformer (CVT)

## I. INTRODUCTION

Technologies for altering images, videos, and audios are developing rapidly. Techniques and technical expertise to create and manipulate digital content are also easily accessible. Currently, it is possible to seamlessly generate hyper-realistic digital images with a little resource and an easy how-to-do instructions available online. Deepfake is a technique which aims to replace the face of a targeted person by the face of someone else in a video. It is created by splicing synthesized face region into the original image. The term can also mean to represent the final output of a hype-realistic video created.

Deepfakes can be used for creation of hyper-realistic Computer Generated Imagery (CGI), Virtual Reality (VR), Augmented Reality (AR), Education, Animation, Arts, and Cinema. However, since Deepfakes are deceptive in nature, they can also be used for malicious purposes. Since the Deepfake phenomenon, various authors have proposed different mechanisms to differentiate real videos from fake ones. As pointed by, even though each proposed mechanism has its strength, current detection methods lack generalizability. The authors noted that current existing models focus on the Deepfake creation tools to tackle by studying their supposed behaviors. For instance, inconsistencies in eye blinking to detect Deepfakes. However, it is now possible to mimic eye blinking. System that generates videos of talking heads with natural facial expressions such as eye blinking. Also proposed a model that can generate facial expression from a portrait. Their system can synthesis a still picture to express emotions, including a hallucination of eye-blinking motions.

## LITERATURE SURVEY

This conference paper provides an extensive overview of Revolutionizing Deepfake Detection and Real-time Video Vision with CNN-based Deep Learning Model, with a specific focus on Deepfake. It covers various aspects on detecting facial manipulation in videos using deep learning. It introduces an ensemble of Convolutional Neural Network (CNN) models, particularly EfficientNetB4, enhanced with attention mechanisms and siamese training. The goal is to identify various facial manipulation techniques like deepfakes. Experiments assess model performance on FF++ and DFDC datasets, using metrics like AUC and Log Loss. The attention mechanism highlights key facial regions, and Siamese training improves discrimination capabilities. Results show that assembling models outperforms the baseline XceptionNet, with discussions on explain ability, feature visualization, and network architecture independence. In our conference paper explores the role of Revolutionizing Deepfake Detection and Real-time Video Vision with CNN-based Deep Learning Model. DeepFakes detection involves detecting and identifying images, videos, audio, and text that have been generated or manipulated using artificial intelligence techniques .DeepFakes are often created to deceive and manipulate viewer's byinserting fake information into real events or spreading misinformation or creating fake news. The use of DeepFakes posesa significant threat to the authenticity and credibility of social media, making it essential to develop reliable and effective DeepFakes detection methods .The DeepFakes detection is primarily focused on identifying and combating the spread of fake news and misinformation. The current surveys can help prevent the potential harm caused by the misuse of DeepFakes and ensure that the information presented is accurate and trustworthy. It explores Manipulated-face detection involves the identification of images that have been manipulated or tampered with by AI algorithms. The increasing availability of AI algorithms has made it easier to create manipulated images for fraudulent purposes, such as face swapping, and facial manipulation .Video conferencing DeepFakes detection. Left: A video call attendant is being actively authenticated with the live patterns shown on the screen. Right: A real person's cornea will produce an image of the pattern shown on the screen while a real-time DeepFake cannot. The current surveys indicate that the development of reliable and robust manipulated-face detection methods is essential to safeguard the authenticity and integrity of digital media and prevent harm caused by the misuse of manipulated images. As previously mentioned, it's important to note that face manipulation and GAN-based face generation are distinct techniques. While GANs generate faces from scratch, face manipulation involves altering or modifying existing face images using various techniques.

## PROPOSED SYSTEM

The designed Artificial Neural Network (ANN) model combines convolutional layers for spatial feature extraction with transformer layers to capture temporal dependencies in video frames. The convolutional layers identify spatial patterns, while the transformer layers process sequential information for robust deep fake detection. The model includes dense layers for classification and is trained to distinguish between authentic and manipulated video content. The architecture leverages both spatial and temporal aspects, enhancing its ability to detect sophisticated deep fake techniques. The model is optimized using categorical cross entropy loss and the Adam optimizer for effective training. The Convolutional Neural Network (CNN) model utilizes convolutional layers to extract spatial features from video frames, focusing on facial patterns and anomalies. Max-pooling layers enhance hierarchical feature representation, aiding in deep fake detection. The model incorporates multiple convolutional blocks to progressively capture complex visual patterns. Flattening is applied before dense layers for classification, enabling the model to discern between genuine and manipulated video content. Training involves optimizing with categorical cross entropy loss and Adam optimizer for efficient convergence.

IJIRIS © 2014-24, AM Publications -All Rights Reserved

## II. SYSTEM ARCHITECTURE

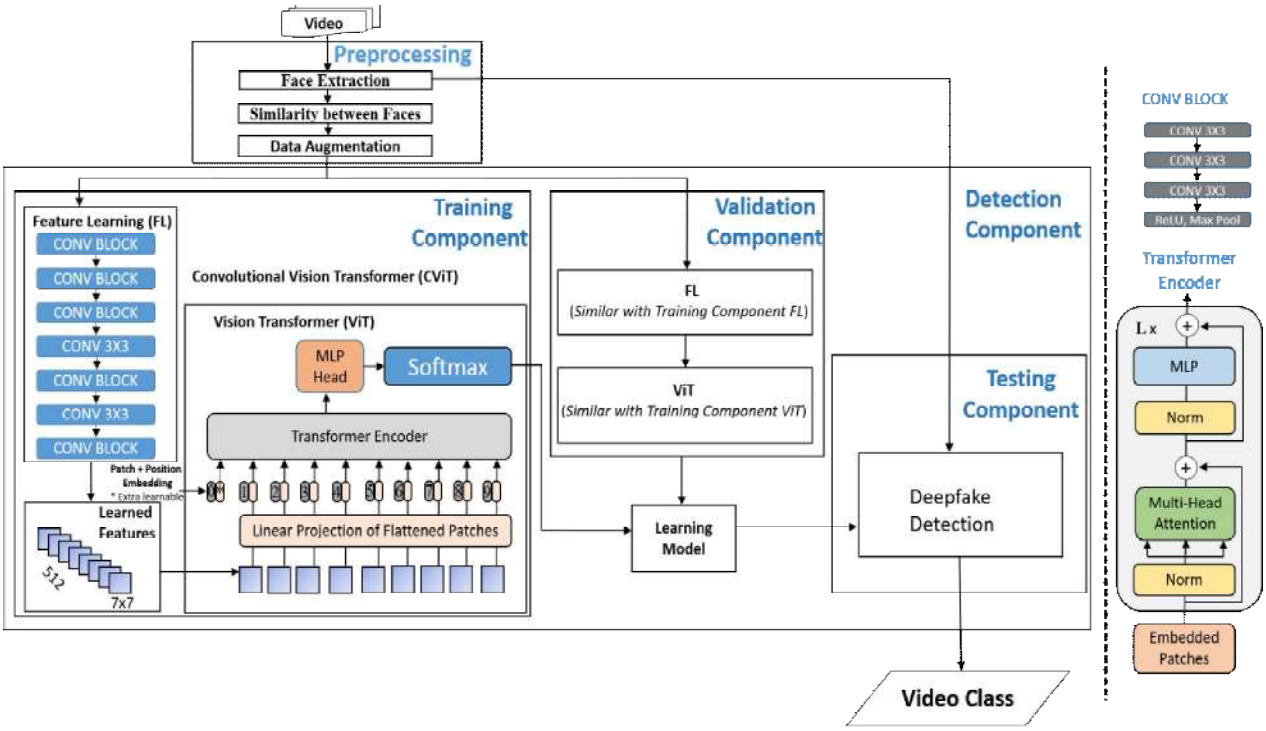


Figure 1: System Architecture

## III .SYSTEM MODULES

### The Deepfake Video Detection Model

The above figure in system architecture explains the model of video deepfake face detection.

### The Training Component

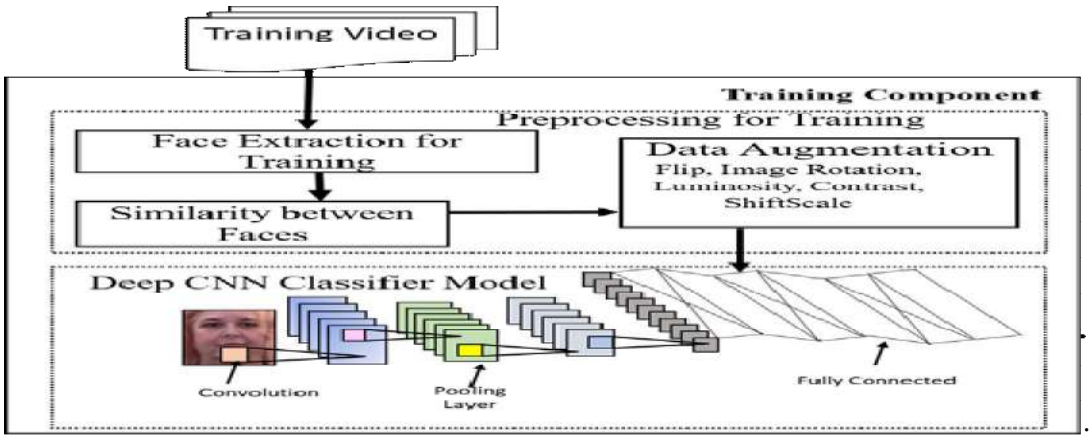


Figure 2 : Training Video

The project titled "Revolutionizing Deepfake Detection and Real-Time Video Vision with CNN-based Deep Learning Model" encompasses several key modules aimed at achieving its objectives. Firstly, the Data Collection and Preprocessing Module focuses on gathering a diverse dataset of real and deepfake videos, standardizing their formats, resolutions, and annotating them for training purposes. Subsequently, the Deep Learning Model Architecture Module is tasked with designing an optimized convolutional neural network (CNN) architecture tailored for deepfake detection and real-time video processing. This involves exploring various CNN architectures, such as ResNet or VGG, and experimenting with different layers and optimization techniques. The Training Module involves training the deep learning model on the annotated dataset, leveraging techniques like transfer learning and data augmentation to enhance performance and generalization capabilities. Real-Time Video Processing Module is responsible for developing algorithms for efficient real-time video analysis and processing, including frame extraction, feature extraction, and temporal analysis, optimized for low-latency performance. The Deepfake Detection Module integrates the trained model into the system for identifying deepfake videos, incorporating thresholding mechanisms and post-processing techniques to refine detection results.

The User Interface (UI) Module focuses on designing an intuitive interface for interacting with the system, allowing users to upload videos, initiate real-time processing, and view detection results across different devices. Lastly, the Monitoring and Maintenance Module involves implementing mechanisms for monitoring system performance, detecting anomalies, and facilitating regular updates and maintenance tasks to keep the system effective and up-to-date with the latest deepfake detection techniques. Together, these modules form a comprehensive framework poised to revolutionize deepfake detection and real-time video vision through advanced CNN-based deep learning models.

## IV. RESULTS

## V. SCOPE OF FUTURE APPLICATION

The scope of future deepfake detection applications Initially, the Data Acquisition and Preprocessing Module is essential for gathering a diverse dataset comprising both real and deepfake videos, followed by cleaning and preprocessing tasks to prepare the data for training. The Deep Learning Model Architecture Module focuses on designing a suitable Convolutional Neural Network (CNN) architecture, configuring the training pipeline, and defining appropriate parameters. Subsequently, the Training Module involves training the CNN model using the prepared dataset, validating its performance, and evaluating key metrics such as accuracy and F1-score. Real-Time Video Processing Module is pivotal for integrating the model with live video streams, enabling real-time deepfake detection, and optimizing processing efficiency. The User Interface Module ensures a user-friendly interaction platform, featuring real-time feedback and control options. Post-Processing and Reporting Module deals with refining detected deepfakes and generating comprehensive reports on detection results. Deployment and Scalability Module strategizes the deployment across different environments and considers scalability factors for efficient operation. Lastly, Maintenance and Updates Module facilitates periodic model retraining, software updates, and maintenance tasks to keep the system robust and up-to-date with evolving deepfake techniques. Together, these modules form a comprehensive framework to address the challenges of deepfake detection and real-time video processing using CNN-based deep learning models.

IJIRIS © 2014-24, AM Publications -All Rights Reserved 

## VI .CONCLUSION

The project was aimed at designing and developing a generalized model for Deepfake video detection utilizing Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). Equal emphasis was placed on data preprocessing during both training and classification phases, ensuring the model's robustness and accuracy. Leveraging the largest available dataset for Deepfake detection, our novel approach yielded promising results. The developed Deepfake detection model demonstrates potential in identifying Deepfakes intended for malicious intents, providing a crucial tool in combating disinformation and fraudulent activities. With a reported accuracy of 93 percent on the DFDC dataset, our model signifies a significant advancement in Deepfake detection technology. Furthermore, its versatility allows for effective detection across various settings, enhancing its applicability in real-world scenarios. The meticulous attention given to data preprocessing during training and classification significantly contributes to the model's overall accuracy, underscoring the importance of comprehensive data preparation in Deepfake detection systems.

## REFERENCES

[1]. Bobby Chesney and Danielle Citron, "Deep Fakes A Looming Challenge for Privacy, Democracy, and National Security," 2019. [Online]. Available: https://ssrn.com/abstract=3213954. [Accessed 28 October 2019].
[2]. Umur Aybars Ciftci, Ilke Demir and Lijun Yin, "FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals," arXiv preprint rXiv:1901.02212v2, August 2019.
[3]. Generated Media, Inc., "Gallery of AI Generated Faces," 2019. [Online]. Available: https://generated.photos/faces.
[4]. Yuezun Li, Ming-Ching Chang and Siwei Lyu, "In Ictu Oculi: Exposing A Generated Fake Face Videos by Detecting Eye Blinking," arXiv preprint arXiv:1806.02877v2, 2018.
[5]. Konstantinos Vougioukas, Stavros Petridis and Maja Pantic, "Realistic Speech-Driven Facial Animation with GANs," International Journal of Computer Vision, vol. 128, p. 1398–1413, May 2020.
[6]. Joshua Brockschmidt, Jiacheng Shang and Jie Wu, "On the Generality of Facial Forgery Detection," in 2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems Workshops (MASSW), Monterey, CA, USA, IEEE, 2019, pp. 43-47.
[7]. Polychronis Charitidis, Giorgos Kordopatis-Zilos, Symeon Papadopoulos and Ioannis Kompatsiaris, "A Face Preprocessing Approach For Improved Deepfake Detection," arXiv:2006.07084v1, June 2020.
[8]. Tero Karras, Samuli Laine and Timo Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," arXiv:1812.04948, December 2018.
[9]. Thanh Thi Nguyen, Cuong M. Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen and Saeid Nahavandi, "Deep Learning for Deepfakes Creation and Detection," arXiv preprint arXiv:1909.11573v1, September 2019.
[10]. Sakshi Agarwal and Lav R. Varshney, "Limits of Deepfake Detection: A Robust Estimation Viewpoint," arXiv:1905.03493v1, May 2020.
[11]. Brian Dolhansky, Joanna Bitton, Ben Pflaum, Jikuo Lu, Russ Howes, Menglin Wang and Cristian Canton Ferrer, "The DeepFake Detection Challenge Dataset," arXiv:2006.07397v3, June 2020.

IJIRIS © 2014-24, AM Publications -All Rights Reserved 