e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:03/March-2024 Impact Factor- 7.868 www.irjmets.com

UNMASKING DIGITAL DECEPTION: DETECTING DEEPFAKES THROUGH DEEP LEARNING

Patlolla Deeksha Reddy*1, Sthalam Sai Charan*2, Sammeta Yashwanth Naidu*3, Dr. M. Sambasivudu*4

*1,2,3Student Of B.Tech Computer Science And Engineering, Department Of Computer Science And Engineering, Malla Reddy College Of Engineering & Technology, Hyderabad, Telangana, India.

*4Associate Professor, Department Of Computer Science And Engineering, Malla Reddy College Of Engineering & Technology, Hyderabad, Telangana, India.

ABSTRACT

The rapid advancement of deep learning technologies has facilitated the creation of convincing face swaps in videos, commonly known as "Deep Fakes" (DF). In recent months, the accessibility of free deep learning-based tools has significantly contributed to the proliferation of this technology. Despite visual effects being employed for decades to manipulate digital video content, recent developments in deep learning have notably enhanced the accessibility and realism of deceptive content. The increased computational power of deep learning algorithms, particularly in the realm of deep fakes, has made the creation of AI-generated media more attainable. However, this accessibility has raised concerns due to the potential misuse of realistic face-swapped deep fakes. These malicious activities include inciting political unrest, orchestrating terrorist acts, spreading revenge porn, and extorting money. In response to these challenges, a novel deep learning technique has been introduced to effectively distinguish between authentic and AI-generated videos. This approach utilizes a Rest Next Convolutional Neural Network (CNN) to extract frame-level characteristics. Subsequently, these characteristics are employed to train a Long Short-Term Memory (LSTM)-based Recurrent Neural Network (RNN). This innovative system can automatically detect deep fakes involving replacement and reenactment, ensuring practicality in real-world scenarios and enhancing model performance on real-time data. The evaluation of this approach involved using a comprehensive and meticulously curated dataset. Various well-established datasets, including the Deep Fake Detection Challenges, Face-Forensic++, celeb-DF, and other pre-existing datasets, were combined to create a robust dataset. This thorough compilation ensures a diverse and representative set of examples for training and testing the deep learning model, enhancing its ability to accurately detect deep fakes across different contexts and scenarios.

I. INTRODUCTION

The rising availability of high-speed internet and sophisticated smartphone cameras has led to a global growth in social networking platforms and media-sharing portals. This has resulted in an unprecedented period of accessibility, allowing for the simple production and diffusion of digital media. At the same time, as computing power has increased, deep learning has reached previously unimaginable levels. However, this revolutionary technological environment introduces new challenges, particularly with regard to "DeepFake" content created by powerful deep generative adversarial models. These fraudulent audio and video recordings, which are widely distributed on social media, are a common source of spam and incorrect information, endangering public safety. Developing powerful DeepFake detection technology is crucial to addressing this expanding issue. To address this, we introduce DF Videos, a novel deep learning technique capable of accurately distinguishing between actual and counterfeit AI-generated videos. Such technology is critical because it detects and prevents DeepFakes from propagating over the internet. It is essential to comprehend the workings of generative adversarial networks (GAN) in order to identify Deepfake. A GAN makes a new movie in which the target's face is replaced with that of another person (referred to as the "source") using a video and an image of a certain person (referred to as the "target"). Deep adversarial neural networks, trained on target films and face images, make up the heart of Deepfakes. These networks transfer faces and expressions from the source to the target automatically. After then, the input image is replaced in each frame of the segmented video. Auto encoders are then utilized to rebuild the frames.

Our deep learning-based detection method replicates how GANs make DeepFakes. It uses the qualities that DeepFake films have by default, as well as the DeepFake algorithm's production constraints and processing limits. The method can only be used to synthesize fixed-size face images, which require an affine warping to match the facial layout in the source. The warped face area and the surrounding context have different resolutions, which causes noticeable artifacts to appear in the DeepFake video output. Once the video has been segmented into frames, we utilize a ResNext Convolutional Neural Network (CNN) to extract features in order to detect these artifacts. Furthermore, we use a Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) to record temporal irregularities introduced by GAN while reconstructing the DeepFake. Resolution inconsistencies in affine face wrappings are intentionally mimicked to train the Res Next CNN model. By carefully addressing the inherent limitations and artifacts produced throughout the fabrication process, our approach essentially aims to increase the accuracy of DeepFake detection.

II. LITERATURE REVIEW

There is a growing need for fake video analysis, detection, and intervention due to the serious threat that the spread of deepfake films poses to democracy, justice, and public confidence.

A specific convolutional neural network model is used to compare generated face areas and surroundings in order to identify deepfake videos. This technique is described in "Exposing DF videos by detecting face wrapping artifacts[1]".This method focuses on finding artifacts from low-resolution photos produced by existing deepfake algorithms.

One such method, described in "Exposing AI created fake videos by detecting eye blinking [2]," uses the physiological signal of eye blinking—which is not well represented in artificially created fake videos—to expose fraudulent face footage. Although the method's performance appears promising, the suggested methodology recommends taking into account more factors like wrinkles and dental enchantments for a more thorough deep fake identification.

"Using capture network to detect forged images and videos [3]" Introducing an approach that uses a capsule network to detect forged, modified photos and videos in a variety of contexts, including replay attack detection and computer-generated video detection. The suggested strategy prioritizes training on noiseless and real-time datasets to improve performance in real-world scenarios.

"Identifying synthetic portrait videos through biological signals [4]" centers on the extraction of biological signals from the facial regions of real and phony portrait video pairings. The method entails performing transformation to compute spatial coherence and temporal consistency while recording signal capitalistic in feature sets and PPG maps. The derived authenticity probabilities are used to judge if a video is fraudulent or genuine.

"Fake Catcher" is a technology made to accurately identify fraudulent content, regardless of the video's creator, content, resolution, or quality. However, the loss of findings pertaining to the preservation of biological signals is caused by its lack of discriminator. For Making a differentiable loss function that adheres to the suggested signal processing stages is the suggested method for improving detection efficiency.

III. METHODOLOGY

Problem Definition:

clearly define the problem of deep fake detection, emphasizing the significance and potential consequences of deep fake proliferation.

Literature Review:

Review of the Literature: To comprehend the current state of deep fake detection methodology, approaches, and difficulties, thoroughly review the literature. Identify holes in current research that your methodology intends to fill.

Data Collection and Preprocessing:

Gather a diverse and representative datasets containing both authentic and deep fake videos. Preprocess the datasets by cleaning, normalizing, and augmenting the data to ensure a balanced and robust training set.

Model Selection:

Choose an appropriate deep learning architecture for deep fake detection, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), or hybrid models. Detection. Depending Consider per-trained models or create a custom architecture based on your data set's individual requirements.

Feature Extraction:

Extract significant features from the datasets that suggest deep fake qualities. Try Experiment with different feature extraction techniques, taking into account both spatial and temporal features.

Training The Model:

Divide the dataset into sets for testing, validation, and training. Using the deep learning model with the training data, fine-tuning hyperparameters as needed. If Implement ways to rectify any existing class imbalances.

Evaluation Metrics:

Define relevant assessment criteria, such as accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve, to assess your model's performance.

Cross Validation:

Implement cross validation techniques to ensure the robustness and generalization of the model.

Hyper parameter Tuning:

Optimize hyper parameters through systematic tuning to enhance the models performance.

Interpret ability and Explain ability:

Incorporate methods for interpretation and explanation to make the deep fake detection process more clear and understandable.

Publishable Results:

Present the outcomes of your deep fake detection methods in an understandable and succinct manner. Process provides a comparative analysis of current approaches, highlighting the advantages and originality of yours.

Ethical Consideration:

Address ethical concerns related to deep fake creation and detection in your publication. Discuss potential implications and applications of your methodology.

Code and Model Sharing:

Transparency and cooperation within the research committee are promoted by freely disseminating the code and trained models.

Peer Review and Validation:

Send in your methodology for a peer assessment to ensure that it is sound and practical.

Continuous Improvement:

Encouraging feedback and continuously improve your deep fake detection methodology based on community input and emerging challenges in the field.

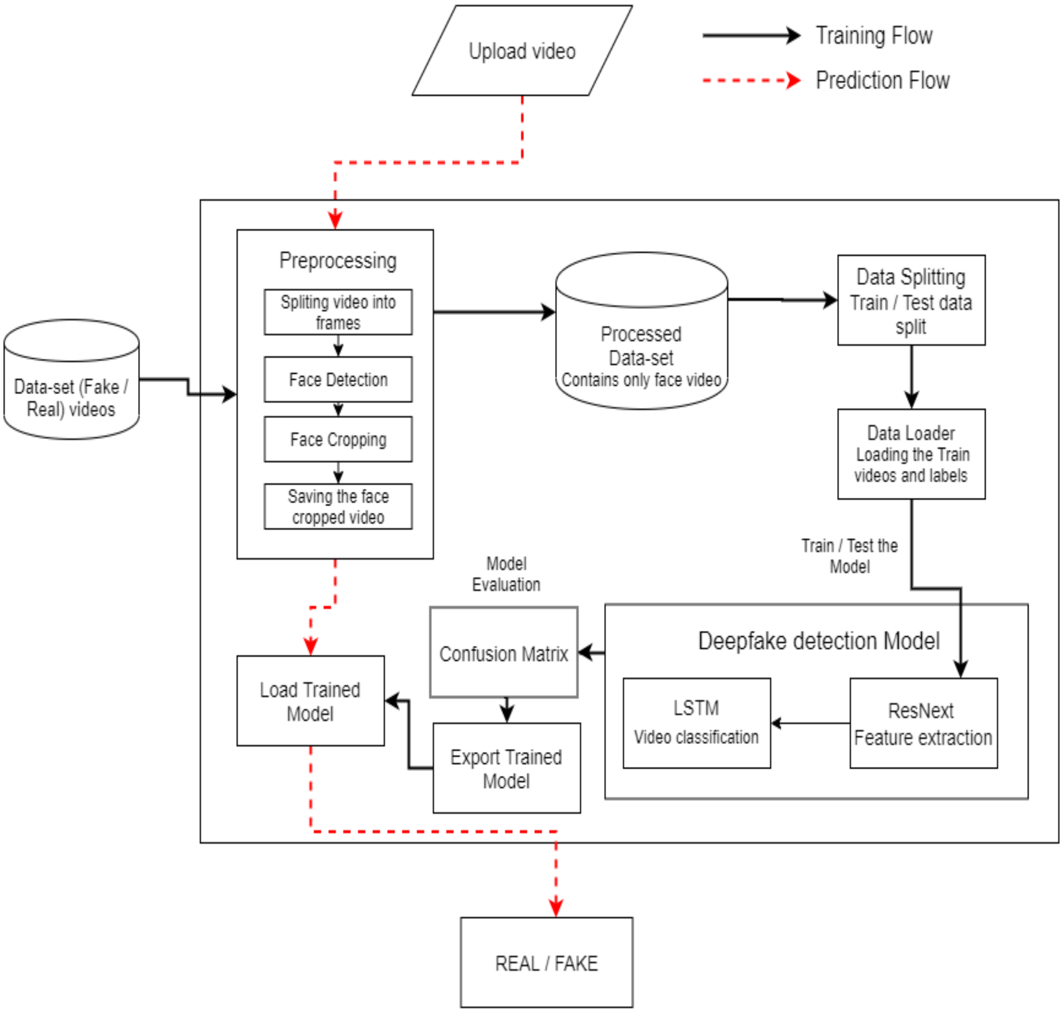


Figure 1: System Architecture Diagram

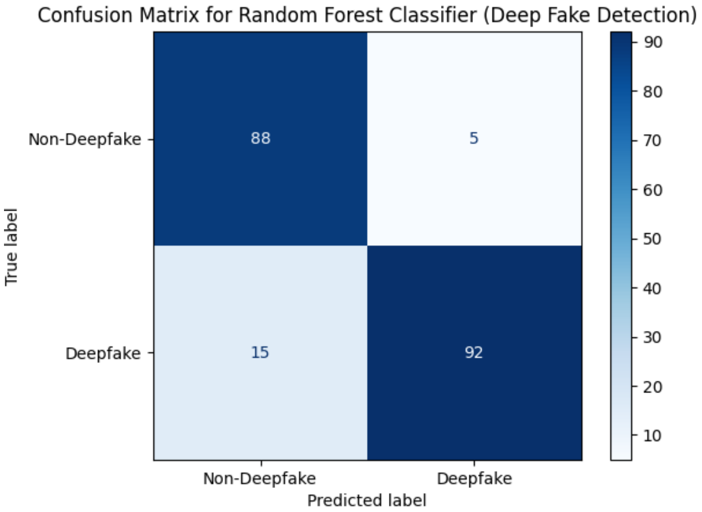


Figure 2: Confusion Matrix Random Forest Diagram

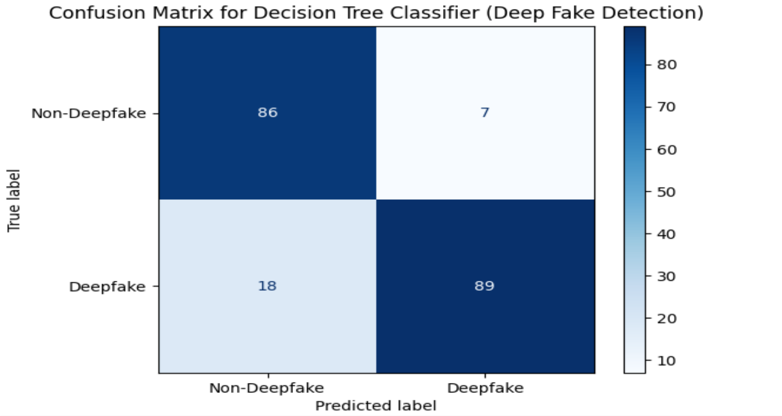


Figure 3: Confusion Matrix Decision Tree Diagram

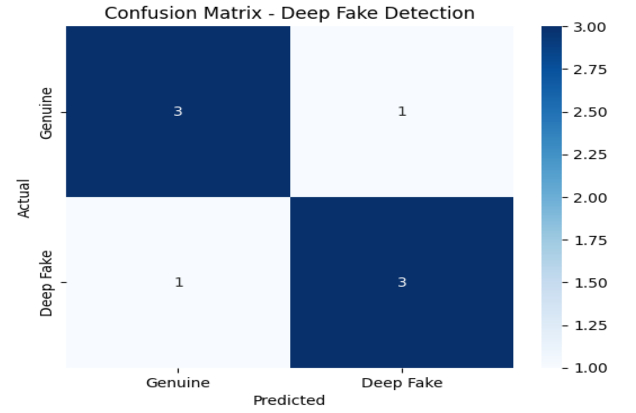
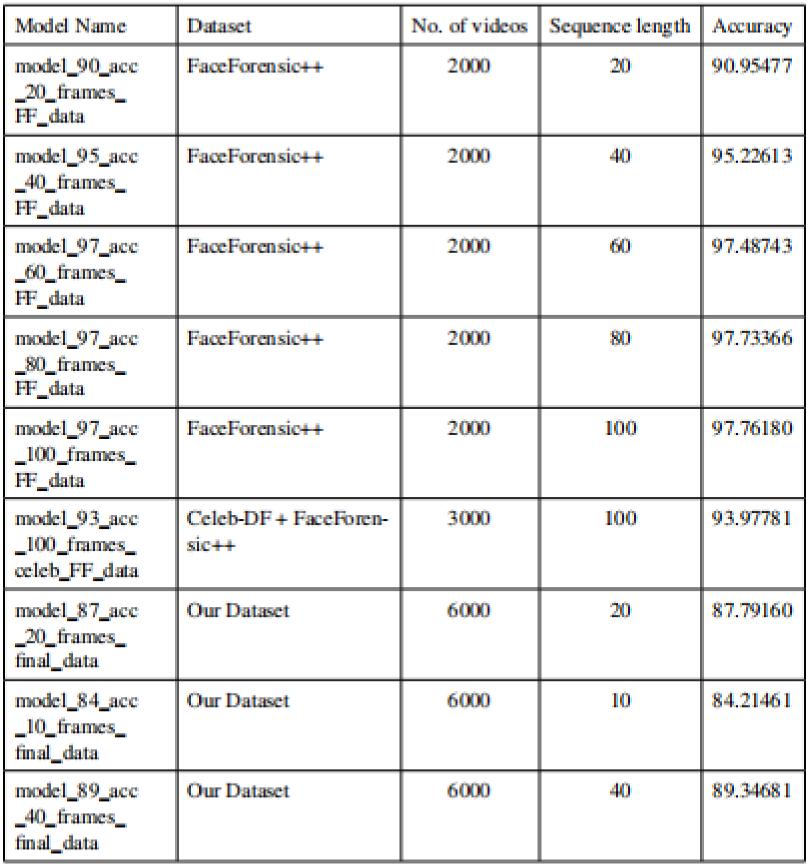


Figure 4: Confusion Matrix Deep Fake Detection Diagram

Table 1: Table Model Results



Model Name	Dataset	No. of videos	Sequence length	Accuracy
model_90_acc _20_frames_ FF_data	FaceForensic++	2000	20	90.95477
model_95_acc _40_frames_ FF_data	FaceForensic++	2000	40	95.22613
model_97_acc _60_frames_ FF_data	FaceForensic++	2000	60	97.48743
model_97_acc _80_frames_ FF_data	FaceForensic++	2000	80	97.73366
model_97_acc _100_frames_ FF_data	FaceForensic++	2000	100	97.76180
model_93_acc _100_frames_ celeb_FF_data	Celeb-DF + FaceForensic++	3000	100	93.97781
model_87_acc _20_frames_ final_data	Our Dataset	6000	20	87.79160
model_84_acc _10_frames_ final_data	Our Dataset	6000	10	84.21461
model_89_acc _40_frames_ final_data	Our Dataset	6000	40	89.34681

IV. RESULTS AND ANALYSIS

We're tackling the urgent problem of deep fake videos head-on by developing an advanced neural network method that distinguishes between fake and real videos. Our strategy, utilizing a pre-trained ResNext CNN and LSTM networks, boasts high accuracy, even with short video snippets, analyzing one second of data at a speedy 10 frames per second.

Our method is versatile across different time intervals, demonstrating adaptability for various video lengths and complexities. Looking ahead, we aim to turn our methodology into a user-friendly browser plugin, seamlessly integrating it into users' browsing experiences for widespread use.

While celebrating our success, we acknowledge the need for improvement. We aspire to expand our algorithm beyond facial manipulation detection to encompass full-body deep fakes, enhancing its effectiveness against deceptive video content.

In essence, our journey persists. Fueled by innovation and determination, we stand ready to face the ever-evolving challenges of digital deception, committed to advancing and remaining resilient.

The image shows a person standing in a room, with the text “FAKE 100.0%” and a red bounding box around their face. This indicates that a deep learning model has identified the person's face as being digitally altered or completely synthesized, commonly referred to as a "deepfake".

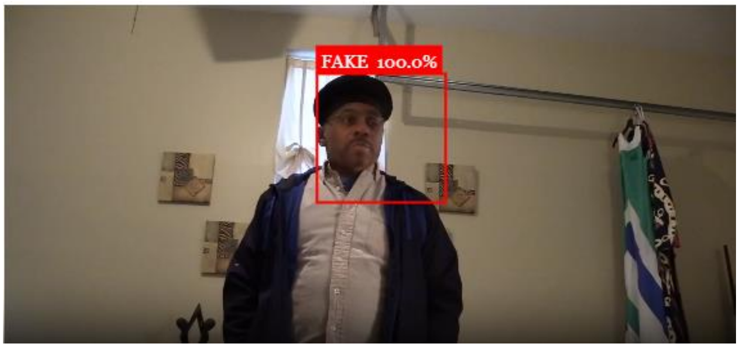


Figure 5: Deep Faked video

V. CONCLUSION

In our presentation, we unveiled a cutting-edge neural network method meticulously designed to discern the authenticity of videos, drawing a clear line between genuine content and deepfake fabrications. The distinguishing feature of our approach lies in its ability to supply the model with a quantifiable classification confidence, ensuring precise predictions, even when analyzing a mere one-second snippet of video data at a brisk rate of 10 frames per second.

At the core of our methodology is the utilization of a pre-trained ResNext CNN model, strategically employed to extract intricate frame-level features. Concurrently, we leverage Long Short-Term Memory (LSTM) networks for the nuanced temporal sequence processing, enabling the detection of subtle alterations between consecutive frames (t and t-1). The processing of frame sequences, spanning intervals of 10, 20, 30, 40, 60, 80, and 100 frames, adds a layer of versatility to our model, showcasing its adaptability across varying video lengths and complexities.

Our inspiration stems from the realm of deepfake images generated through Generative Adversarial Networks (GANs) and autoencoders. The incorporation of frame-level detection, utilizing the ResNext CNN and LSTM, is a pivotal aspect of our technique. With meticulously specified parameters, our method excels in distinguishing between authentic and deepfake videos, providing exceptional accuracy in real-time scenarios. It stands as a testament to our commitment to staying at the forefront of technological innovation, particularly in the face of challenges posed by deceptive visual content.

VI. REFERENCES

[1] Rossler, Andreas, et al. investigate "FaceForensics++: Learning to Detect Manipulated Facial Images," a study published on arXiv under the identifier 1901.08971, which provides insights on advances in facial manipulation detection.

[2] The Deepfake Detection Challenge Dataset, a critical resource for academics, is available for study and application via Kaggle's platform, as reported on March 26, 2020, at the following URL: [Deepfake Detection Challenge Data](https://www.kaggle.com/c/deepfake-detection-challenge/data).

[3] A substantial dataset to test and improve DeepFake detection methods is proposed by Li, Yuezun, et al. in "Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics," which is published at arXiv:1909.12962.

[4] The global spread of a Deepfake video featuring Mark Zuckerberg on the day of a major House AI hearing, as reported by Fortune.com on June 12, 2019, highlights the sociological and political ramifications of DeepFake technology. You can review this incident at [Mark Zuckerberg's Deepfake](https://fortune.com/2019/06/12/deepfake-mark-zuckerberg/).

[5] On March 26, 2020, Creative Bloq produced "10 deepfake examples that terrified and amused the internet," a showcase illustrating the dual-edged nature of DeepFake technology, which is accessible at [Deepfake instances](https://www.creativebloq.com/features/deepfake-examples).

[6] TensorFlow, Keras, and PyTorch are foundational DeepLearning tools that provide substantial libraries and frameworks to facilitate the building of models capable of identifying and analyzing DeepFakes, with resources available on their respective websites as of March 26, 2020.

[7] "Face aging with conditional generative adversarial networks," a creative study by Antipov, G., et al., that was published on arXiv (1702.01983) in February 2017, examines facial alteration and shows how GANs may be used to simulate aging in a realistic way.

[8] The IEEE Conference on Computer Vision and Pattern Recognition in June 2016 featured a work by Thies, J., et al. titled "Face2Face: Real-time face capture and reenactment of RGB videos," which illustrates the potential and ramifications of real-time facial reenactment technology.

[9] Applications like as FaceApp and Face Swap, which are available as of March 26, 2020, present actual instances of facial modification technology, emphasizing the need for robust detection techniques.

[10] A November 1, 2019 article on Forbes.com highlights the serious consequences of DeepFakes, especially with regard to revenge porn and how it disproportionately affects women. It also emphasizes the urgent need for strong defenses against these kinds of abuses.

[11] The interactive article "The rise of the deepfake and the threat to democracy" from The Guardian captures the larger social worries about DeepFake technology and considers how it can erode democratic processes.

[12] In their study "Exposing DeepFake Videos By Detecting Face Warping Artifacts," published in arXiv:1811.00656v3, Li, Yuezun, and Lyu, Siwei offer a technique for spotting facial warping irregularities that can be used to identify DeepFakes.

[13] In a similar vein, their paper "Exposing AI Created Fake Videos by Detecting Eye Blinking," which is available at arXiv:1806.02877v2, presents a novel method for identifying DeepFakes that centers on the typical human blinking pattern.

[14] Nguyen, Huy H., et al. demonstrate the potential of this architecture in recognizing manipulated content by investigating the usage of capsule networks for the identification of fabricated photos and movies, as reported in arXiv:1810.11215.

[15] In "Deepfake Video Detection Using Recurrent Neural Networks," Güera, D., and Delp, E. J. (2018) presented at the 15th IEEE International Conference on Advanced Video and Signal-Based Surveillance, discuss how RNNs are used to detect DeepFake movies.

[16] Laptev, I., et al.'s analysis of actual human activities from movies, which was presented at the IEEE Conference on Computer Vision and Pattern Recognition in June 2008, helps us grasp motion patterns, which are important for spotting abnormalities in altered videos.

[17] Ciftci, Umur Aybars, et al. suggest using biological signals to detect synthetic portrait movies, a new strategy documented in arXiv:1901.02212v2, which expands the DeepFake detection toolbox.

[18] In December 2014, Kingma, D. P., and Ba, J., published "Adam: A method for stochastic optimization," which is a key contribution to deep learning optimization methods. It may be accessed at arXiv:1412.6980.

[19] PyTorch has greater details about the ResNext Model, a potent convolutional neural network architecture, on their official hub, which is available for testing and deployment as of April 6, 2020.

[20] GeeksforGeeks provides an explanation of the COCOMO model, which is crucial for comprehending software engineering and development project estimations and provides insightful information about project management techniques.

[21] The International Journal for Scientific Research and Development and other scholarly journals keep adding to the conversation about DeepFake Video Detection with Neural Networks by highlighting the continuous efforts to research and develop solutions to lessen the difficulties presented by DeepFake technology.