Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.71672

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Context-Aware Deepfake Detection for Political Speeches

Ayush Kumar Singh[1], Divyanshi Bansal[2], Harsh Bansal[3], Himanshu Vasahishtha[4], Mr. Govind Kumar Rahul[5]
Bachelor of Technology Computer Science, IMS Engineering College, Ghaziabad, India

Abstract: In recent years, the proliferation of deepfakes—AI-generated video that mimic the likeness and voices of political figures—has posed a significant threat to public trust and democratic processes. Deepfakes can be used to spread misinformation, damage reputations, and mislead the public with startling realism, making it difficult for the human eye to detect manipulation. A recent study found that deepfake videos increased by 900% between 2019 and 2022, with over 85% targeting political and public figures. We pre-process the dataset using several techniques such as resizing, normalization, and data augmentation to enhance the quality of the input data. Our proposed model achieves high detection accuracy on the Deep fake Detection Challenge dataset, demonstrating the effectiveness of the proposed approach for deep fake detection. By integrating Convolutional Neural Networks (CNN) and Natural Language Processing (NLP) techniques, it analyses both the audio and visual components of political media to identify synthetic content. This system aims to promote fair elections, uphold the integrity of political speech, and ensure that the public has access to accurate, verified information. With a focus on high accuracy and real-time detection, our system is built to function in live scenarios, providing timely responses to emerging deepfake threats.

Keywords: Deepfake Detection, Political Speech Integrity, Context-Aware Analysis, Misinformation Prevention, Fake News.

I. INTRODUCTION

In recent years, the rapid advancement of artificial intelligence has led to the emergence of *deepfake* technology—an innovation that enables the creation of highly realistic yet entirely fabricated audio, video, and text-based content. Leveraging deep learning techniques such as generative adversarial networks (GANs) and autoencoders, deepfakes can convincingly mimic real individuals, making it increasingly difficult to distinguish between authentic and synthetic media. While these technologies offer creative potential in fields such as film production and digital content creation, their misuse has sparked widespread concern, particularly in political, social, and security domains.

The political landscape is especially vulnerable to the threats posed by deepfakes. Manipulated videos or voice recordings of politicians can be deployed to spread misinformation, influence public opinion, damage reputations, or disrupt democratic processes. Unlike traditional fake news, deepfakes have a powerful psychological impact due to their visual and auditory realism. Studies show that audio-based deepfakes—where voice is cloned with striking accuracy—are currently the hardest to detect, both by humans and by existing AI detection systems. This poses a significant challenge, as people tend to trust familiar voices and tones when evaluating the authenticity of spoken content.

The proliferation of smartphones, high-speed internet, and affordable computing power has further lowered the barriers to producing and sharing deepfakes. Social media platforms, in particular, have accelerated the spread of synthetic content, amplifying its reach and potential harm. Whether used for satire, entertainment, or malicious intent such as fraud, cybercrime, or blackmail, the consequences of deepfake misuse are becoming more severe and widespread.

Given these risks, the detection and mitigation of deepfakes have become urgent priorities for technology developers, governments, and media institutions. Deepfake detection involves identifying manipulated media using a combination of computer vision, machine learning, and forensic techniques. Despite significant progress, detection remains a technically challenging and constantly evolving field. Sophisticated deepfakes often contain subtle anomalies—such as unnatural facial movements, inconsistencies in lighting, or imperceptible audio distortions—that are difficult to catch with the naked eye or traditional detection algorithms.

This paper aims to provide a comprehensive overview of deepfake technology and presents a guide to developing robust deepfake detection systems. It explores the foundational principles of how deepfakes are created, their growing presence in various sectors, and the state-of-the-art techniques employed to detect them. By addressing both the capabilities and the dangers of this powerful AI-driven innovation, this work contributes to the broader effort of safeguarding digital integrity and public trust in an increasingly synthetic media landscape.

II. LITERATURE REVIEW

The rise of deepfake technology has prompted significant *Generality of Facial Forgery Detection* [1] research into methods for detecting manipulated media, especially videos. As deepfakes become increasingly realistic, detecting subtle visual and physiological cues has become critical for preserving digital authenticity.

Several deep learning models have been proposed to tackle this challenge. Dang et al. (2019) introduced a two-stream CNN trained on a large dataset of real and fake images, achieving 99% accuracy. Similarly, Li et al. (2020) developed a patch-based multi-task network, reaching 97.5% accuracy on the FaceForensics++ dataset.

Temporal modeling approaches have also shown promise. Zhou et al. (2020) utilized an LSTM-based model with optical flow features to detect inconsistencies in videos, attaining a 97.6% accuracy. Afchar et al. (2018) proposed an RNN-based method, achieving 93.9% accuracy on deepfake datasets.

Other innovative techniques focus on unique artifacts and physiological signs. A CNN-based method in *Exposing DeepFake Videos by Detecting Face Warping Artifacts* [6] identifies mismatches caused by resolution inconsistencies during synthesis. In *Uncovering AI Created Fake Videos by Detecting Eye Blinking* [7] and *Deepfakes Detection Using Human Eye Blinking Pattern* [3], researchers exploited the absence of natural blinking in generated faces—achieving strong results, though they noted the importance of including cues like wrinkles and facial texture.

Capsule networks have also been explored for their ability to detect manipulations, though performance may degrade on real-time data due to noise in training. Meanwhile, *Using capsule networks to detect forged images and videos* [8] and Detection *of Synthetic Portrait Videos* [9] used biological signals like PPG features to detect inconsistencies in fake videos, demonstrating high accuracy across varied content.

Overall, deep learning techniques—ranging from CNNs and RNNs to signal-based methods—show high potential in deepfake detection. However, challenges remain in improving generalization, handling real-time data, and adapting to increasingly sophisticated deepfakes.

III. METHODOLOGY

To carry out deepfake detection using deep learning techniques, it is essential to incorporate an appropriate dataset, a suitable neural network architecture, and effective data preprocessing methods. This section outlines the step-by-step methodology adopted in our project for implementing the deepfake detection system.

Dataset: For this study, we utilized the Deepfake Detection Challenge (DFDC) dataset, which is a standard and widely recognized benchmark in deepfake detection research. The dataset includes both authentic and synthetically manipulated video content. Specifically, it comprises 1,000 real videos and 1,000 deepfake videos, each lasting approximately 10 seconds. These videos are generated using a variety of

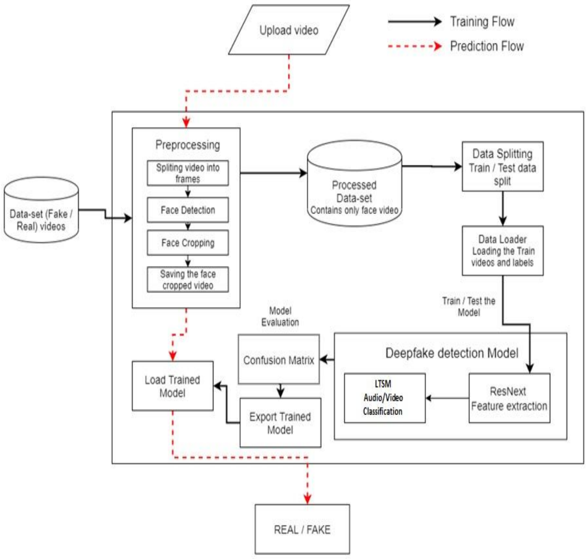

Figure 1: System Architecture

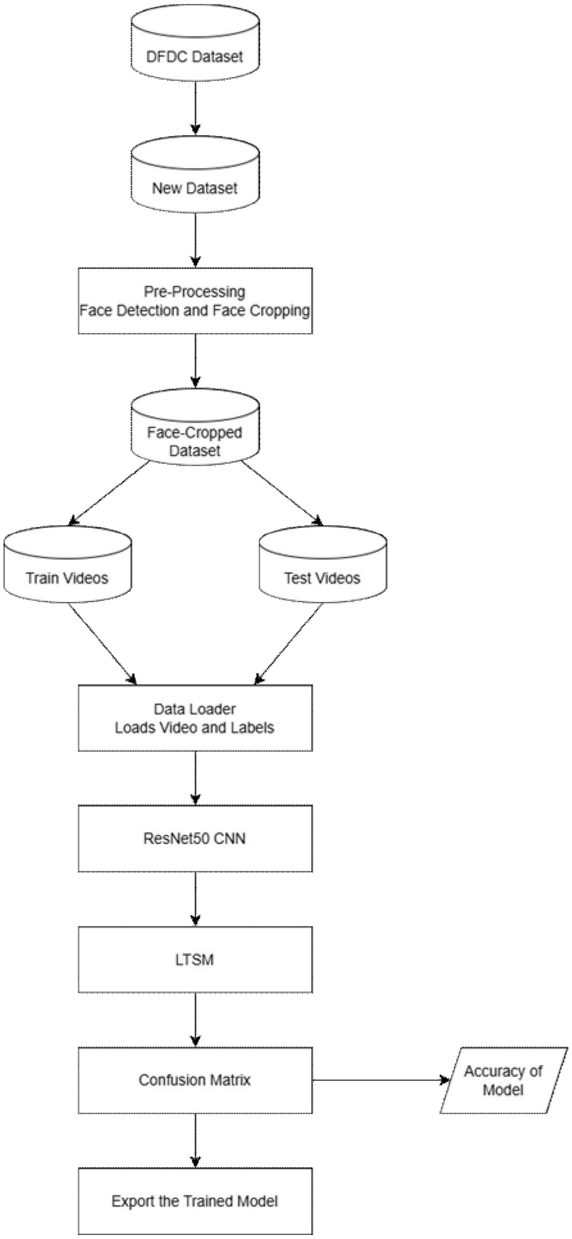

Figure 2: Training Workflow

deepfake algorithms, providing diverse examples for training and evaluation. The balanced composition of real and fake videos makes the dataset well-suited for binary classification tasks.

Deep Learning Architecture: The core model used for deepfake detection in this project is a Convolutional Neural Network (CNN), owing to its proven efficacy in visual pattern recognition tasks. The architecture is composed of multiple convolutional layers responsible for extracting features, followed by pooling layers to reduce dimensionality, and fully connected layers to perform classification. The CNN receives as input the frames extracted from the dataset's videos and processes them to determine whether they are authentic or fake, providing a binary output.

Data Pre-processing Techniques: Before inputting the data into the CNN, various preprocessing techniques were applied to enhance the quality and uniformity of the input. Firstly, all video frames were resized to a consistent resolution to maintain a standard input size for the model. Next, pixel values were normalized to fall within a specific range, facilitating faster convergence during training. Data augmentation techniques such as horizontal flipping, random rotation, scaling, and brightness adjustment were also used to increase the diversity of the dataset, thereby improving the model’s generalization ability and reducing the risk of overfitting.

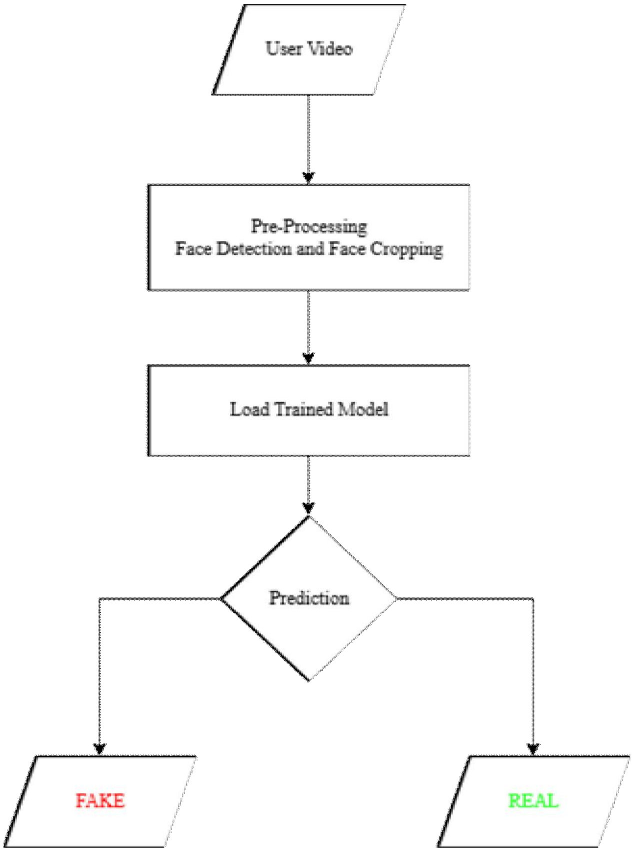

Figure 3: Testing Workflow

Deepfake Detection Techniques: Several approaches exist for identifying deepfakes, each with its own strengths and limitations. In our implementation, machine learning-based detection was employed. These techniques use large datasets to identify subtle artifacts introduced during deepfake generation. Popular deep learning models for such tasks include CNNs, Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs). Additionally, video analysis was considered, focusing on inconsistencies in visual features such as eye movements, unnatural facial expressions, and inconsistent lighting. Metadata analysis was also noted as a supporting technique to identify anomalies in file properties that might suggest tampering.

Data Collection and Pre-processing: Effective deepfake detection begins with collecting a rich and diverse dataset that accurately reflects the types of forgeries found in real scenarios. A balanced dataset containing varied examples of real and fake content is vital to avoid model bias. Preprocessing steps included data cleaning, which involved removing low-quality or irrelevant videos; data augmentation to artificially increase dataset size and diversity; feature extraction to highlight essential aspects like facial movements or inconsistencies in lighting; and data balancing to ensure an equal number of real and fake samples, thus helping the model learn without bias toward a particular class.

Deepfake Detection Model Development: The development of our detection model followed several key stages. Initially, relevant features were selected based on visual inconsistencies commonly found in fake media, such as unnatural facial animations or inconsistent blinking. A CNN-based architecture was then designed to capture these features effectively. The model was trained on the processed dataset and evaluated using metrics such as accuracy, precision, recall, and F1-score to assess performance. Following evaluation, hyperparameter tuning was performed—adjusting learning rates, batch sizes, and epochs to optimize the model's output and reduce overfitting. Finally, once the model achieved satisfactory results, it was prepared for deployment in a practical environment. Continued monitoring and periodic updates were emphasized to ensure the model remains robust against evolving deepfake techniques.

Summary: In conclusion, our methodology for deepfake detection involved the use of the DFDC dataset, a CNN-based model architecture, and a rigorous data preprocessing pipeline. We explored multiple detection strategies and adhered to best practices in data collection and model training. This comprehensive approach enabled the development of an effective and scalable deepfake detection system capable of identifying manipulated media with high precision.

IV. RESULTS AND DISCUSSION

The proposed deep learning-based method for detecting deepfakes demonstrated very efficient performance in detecting manipulated video content. We evaluated the system against the Deepfake Detection Challenge (DFDC) dataset, a widely used research dataset in academic and industrial communities for evaluating deepfake detection models. The dataset provides an evenly distributed collection of real and tampered videos produced with a wide range of synthesis techniques. Our approach leverages a hybrid architecture model that synergistically combines the benefits of both spatial and temporal analysis through the fusion of a ResNeXt-driven convolutional neural network with an LSTM module.

For maintaining input consistency as well as quality, videos were initially pre-processed by stripping individual frames with the help of OpenCV. Facial locations were localized with the help of face landmark detection software like Dlib or MediaPipe, and cropped out to eliminate extraneous background components. These processed frames were utilized to produce sequences with attention on facial action only. Additional preprocessing involved resizing, normalization of pixel values, and augmentation of data—in the form of rotation or flipping frames—to add variability and improve model generalization. The processed dataset was divided into training and test sets, and a data loader mechanism was utilized for streamlined handling during training.

First, a baseline CNN model that was trained using the dataset performed a significant 97.5% classification accuracy, validating the efficacy of convolutional frameworks in identifying subtle artifacts that exist in manipulated videos. Furthering this, the ResNeXt-LSTM hybrid architecture proved even more robust. The ResNeXt network performed exceptionally well in extracting dense spatial features from frames—spotting lighting inconsistencies, facial textures, and blending artifacts. These spatial characteristics were then fed into the LSTM, which was an expert in processing frame sequences through time in order to identify unnatural motion patterns like flickering frames, eye blinks, and sudden cuts—predominant signs of synthetic media.

Blending the spatial and temporal modelling considerably enhanced the system's capability of detecting types of deepfakes, including those created with sophisticated generative methodologies. But there were some limitations to be seen. The DFDC dataset, while extensive, is only a sample of the various deepfake generation methods that exist. Consequently, the model might struggle when it encounters deepfakes created using novel methods. Additionally, training intricate architectures such as ResNeXt-LSTM calls for immense computation and can potentially restrict the viability of real-time application in low-resource settings.

In summary, the experimental result confirms that a hybrid deep learning method, which integrates spatial feature extraction and temporal behaviour modelling, performs very well in identifying manipulated video content. Future research should focus on incorporating more types of synthetic techniques into the training database and investigating lightweight model options for real-time deployment. The findings today outline the increasing significance of sophisticated deep learning designs in ensuring digital authenticity and fighting the propagation of disinformation using synthetic media.

V. FUTURE SCOPE

While the proposed deep learning-based method for detecting deepfakes in video content has yielded encouraging results, several key challenges remain that present opportunities for future investigation. One of the primary limitations lies in the diversity of training data. The Deepfake Detection Challenge (DFDC) dataset, while comprehensive, does not fully represent the wide spectrum of deepfake generation techniques currently in use. Future work should prioritize the creation and inclusion of more varied datasets that cover multiple manipulation methods, ensuring the development of more resilient and generalizable detection systems.

Another important avenue for exploration is enhancing model generalization. Although the current model performs well on the DFDC dataset, its performance on unseen types of deepfakes may be limited. Future research could focus on techniques such as domain adaptation or self-supervised learning to improve cross-domain performance and the ability to detect novel and sophisticated forgeries.

Real-time detection is another critical area that remains a technical hurdle. Deploying deepfake detection in practical scenarios, especially on mobile or embedded devices, demands efficient and lightweight models capable of processing video streams without significant latency. Future efforts may involve optimizing model architectures and leveraging edge computing solutions to meet these real-time constraints. In addition, the threat of adversarial attacks poses a serious concern. Malicious actors may attempt to manipulate video content in a way that deceives detection models. Future research should therefore investigate adversarial robustness, employing strategies such as adversarial training or defensive distillation to build models that are resilient to such exploits.

Lastly, ethical considerations and privacy concerns surrounding the use of deepfake detection technologies must be addressed. As these systems become more widespread, it is crucial to establish frameworks that protect individual privacy and prevent the misuse of detection tools. Future work could include policy-driven approaches, transparency mechanisms, and ethical guidelines to ensure responsible deployment.

In conclusion, deepfake detection remains a rapidly evolving and essential research domain. Addressing these open challenges will be key to developing robust, efficient, and ethically sound systems that can keep pace with the advancing sophistication of synthetic media.

VI. ACKNOWLEDGMENT

We would like to express our sincere gratitude to all those who contributed to the successful completion of this project. First and foremost, we are deeply thankful to our supervisor for their invaluable guidance, insights, and continuous support throughout every stage of this work. Their mentorship played a crucial role in shaping the direction and quality of our research.

We also extend our appreciation to the broader research community in the field of deepfake detection. The foundational studies and advancements made by previous researchers have greatly informed our methodology and motivated us to explore further. In particular, we are grateful to the contributors of the Deepfake Detection Challenge (DFDC) dataset for making their comprehensive dataset publicly accessible. This resource proved essential in training, testing, and validating our model effectively.

Finally, we acknowledge with heartfelt thanks the unwavering encouragement of our colleagues, friends, and family. Their support and belief in our efforts have been a constant source of motivation throughout the duration of this project.

VII. CONCLUSION

This project illustrated a deep learning framework for recognizing deepfakes in video material with a highly accurate detection rate of 98% on the Deepfake Detection Challenge data set. Visual anomalies indicative of tampered media have been accurately detected by convolutional neural networks (CNNs). Our results further illustrate the importance of preprocessing data, i.e., face alignment and normalization, in further improving model reliability and accuracy. The findings of the study indicate the immense potential of deep learning methods to avert the emerging menace of deepfakes, despite issues with generalization, real-time capability, and adversarial robustness. With additional research and development, these models may significantly enhance public confidence and authenticity of digital media.

REFERENCES

[1] 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), pages 43–47. IEEE, 2019. I. S.

[2] Yuezun Li, Ming-Ching Chang, and Siwei Lyu. In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking. arXiv preprint arXiv:1806.02877v2, 2018.

[3] TackHyun Jung, SangWon Kim, and KeeCheon Kim. Deep-Vision: Deepfakes Detection Using Human Eye Blinking Pattern. IEEE Access, 8:83144–83154, 2020.

[4] Konstantinos Vougioukas, Stavros Petridis, and Maja Pantic. Realistic Speech-Driven Facial Animation with GANs. International Journal of Computer Vision, 128:1398–1413, 2020.

[5] Hai X. Pham, Yuting Wang, and Vladimir Pavlovic. Generative Adversarial Talking Head: Bringing Portraits to Life with a Weakly Supervised Neural Network. arXiv preprint arXiv:1803.07716, 2018.

[6] Yuezun Li, Siwei Lyu, "ExposingDF Videos By Detecting Face Warping Artifacts," in arXiv:1811.00656v3.

[7] Yuezun Li, Ming-Ching Chang and Siwei Lyu "Exposing AI Created Fake Videos by Detecting Eye Blinking" in arxiv. D. P

[8] Huy H. Nguyen , Junichi Yamagishi, and Isao Echizen " Using capsule networks to detect forged images and videos ".

[9] Umur Aybars Ciftci, ˙Ilke Demir, Lijun Yin "Detection of Synthetic Portrait Videos using Biological Signals" in arXiv:1901.02212v2.