International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 5, Sep-Oct-2026, ISSN (Online): 2395-566X

Hybrid InceptionV3-LSTM Framework for Real-Time Deepfake Video Detection

Prof. Vikram Singh[1], Naresh Kumar[2]
[1]Professor, Department of Computer Science & Engineering Chaudhary Devi Lal University, Sirsa (Haryana), India
[2]Research Scholar, Department of Computer Science & Engineering Chaudhary Devi Lal University, Sirsa (Haryana), India

Abstract — The rapid generation of deepfake videos poses significant challenges to real-time detection systems, which often encounter a wide range of quality variations and novel manipulation techniques. In order to deal with the challenges faced we propose a hybrid InceptionV3-LSTM network that encapsulates spatial feature extraction and action sequence over time. This architecture employs InceptionV3 pretrained on ImageNet for modeling frame-level artifacts, then stacks a bidirectional LSTM to jointly learn the temporal problem from multiple sequences. The model displays robustness under high compression and when tested on low compression, it shows a decline of just 2.5% in accuracy which implies that the model is sensitive. According to temporal analysis, LSTM proves capable of detecting manipulations as illustrated due to reconstruction loss variance being higher by 3.2× of low-fidelity deepfakes. The deepfake inference data under tool-based evaluation shows this detection. The InceptionV3-LSTM's full-fledged version achieves 150 ms latency and 96% accuracy while real-time performance granted 20 ms/frame latency using MobileNet-LSTM with no accuracy loss (96%). The solution presented in this study can be deployed, balancing detection accuracy and computational efficiency. However, limitations still exist for novel manipulation techniques and multi-modal integration could be an avenue for future work. The objective of this experimental study is to achieve the enhanced detection accuracy using separate training and testing datasets of deepfake videos which can help get a better security performance and robustness. By integrating these advances, the application allows users to check the authenticity of people and deep fake videos. A framework for real-time detection of visual deepfake videos analyzes pixel-level artifacts including blending boundaries, double edges, eyes, ears, nose, head pose inconsistencies and other possible parameters sub-pixel noise patterns unique to manipulation videos.

Keywords— Deepfake Video Detection, Synthetic Media Detection, Generative Adversarial Network, Recurrent Neural Network, Long Short-Term Memory (LSTM), Spatiotemporal analysis.

I. INTRODUCTION

Synthetic media produced using artificial intelligence also called deepfakes have become more widespread than ever. These pose a danger to digital trust and public discussion and to truth. Deepfake plays with the deep learning architectures particularly generative adversarial networks (GANs), variational autoencoders (VAEs) and most recently diffusion models in order to manipulate or create new multimedia content that are hardly distinguishable from real human behavior. There are a series of interdependent technical and operational challenges to real-time deepfake video detection, including technical architectural and operational issues that serve as deterrents against the deployment of viable scalable and future-proof deepfake forensic systems [1]. The previous pixel-level artifacts anachronisms in compression or lighting no longer suffices as a telltale sign of image manipulation since modern deepfake generators can synthesize plausible images with high fidelity and temporal coherence. Thus, current detection approaches should consider physical (spatial) so called content based feature modeling as well as non-physical (temporal) features that detect nuanced physiological and behavioral variances embedded within the synthetic videos. For example, natural human behaviors like eye blinking patterns, gaze behavior, micro-facial movements and not to mention head motion dynamics often deviate in deepfakes as training data is commonly incomplete or algorithms themselves have limitations in terms of replicating biological signals that are spontaneous [2,3]. Although currently developed deepfake detectors have shown technical success they do face shortcoming in coping with novel manipulation methods and a variety of deepfake creation methods. Temporal inconsistencies which are responsible for such forgery often remain undetected in the existing frame-based detection systems while the pure temporal models lack in spatial feature extraction needed to differentiate between subtle facial artifacts [57]. Recent studies have found that deepfake videos may show noticeable irregularities in eye movements, particularly a lower blinking frequency or asynchronous movement of the eyes. These irregularities can provide useful cues for detecting manipulated video content. The CycleGAN-S+ framework is a deep learning-based approach designed for deepfake video detection. It uses convolutional neural networks (CNNs) to extract fine-grained spatial features from video frames. In

addition to spatial information, the framework also considers the temporal information carried over from previous frames. This temporal information can be modeled using recurrent architectures such as Long Short-Term Memory (LSTM) networks, or transformer-based models. By combining spatial and temporal features, the framework can identify subtle patterns that may help distinguish deepfake videos from genuine ones. [4,5]. Explainable AI (XAI) can improve the transparency of deepfake detection by helping forensic analysts understand classification results through tools such as saliency maps and anomaly heat maps. However, several challenges still exist including adapting to new generative models, handling adversarial attacks and deploying detection systems on edge devices. The rapid development of fake media therefore highlights the need for more reliable detection methods. To address these issues researchers are increasingly using advanced deep learning techniques that can effectively capture both spatial and temporal patterns in manipulated media [6,7]. Although existing deepfake detection methods have shown promising results they still face difficulties in handling new manipulation techniques and different types of deepfake generation methods. Frame-based approaches may overlook changes that occur across consecutive frames while methods that mainly focus on temporal information may not capture the fine spatial details needed to detect subtle facial artifacts [56]. This research aims for a better approach for detection real media from fake media by making most of the strength proposed in both architectures. As deepfake generation methods continue to evolve in sophistication, the demand for solutions that not only achieve high detection performance but also do so efficiently continues to grow. Traditional methods rely on high computational cost, which sometimes limits their real-world applicability. However, our approach offers a balance of accuracy and resource overhead that enables practical real-world deployment where high accuracy can be achieved without demanding an unreasonable number of resources. Spreading of the Deep fakes over the social media platforms have become very common leading to spamming and peculating wrong information over the platform. Just imagine a deep fake of our prime minister declaring war against neighboring countries or a Deep fake of reputed celebrity abusing the fans. These types of the deep fakes will be terrible and lead to threatening misleading of common people. Existing deepfake detection approaches primarily rely on single-model architectures trained on fixed datasets which often fail to generalize to unseen manipulation methods and real-world conditions. To overcome such a situation deep fake detection is very important. So, we describe hybrid InceptionV3-LSTM framework for real-time deepfake video detection that can effectively distinguish AI generated fake videos (deepfake videos) from real videos. It's incredibly important to develop technology that can spot fakes so that the deep fakes can be identified and prevented from spreading over the social media.

1. Important of Hybrid InceptionV3-LSTM Framework for Real-Time Deepfake Video Detection

Hybrid InceptionV3-LSTM framework for real-time deepfake video detection is increasingly recognized as vital components in deepfake video detection systems because they address several critical limitations of traditional models by enhancing transparency trust and practical utility. Deep learning model detectors often achieve high accuracy in classifying content as real or fake but without explainability their internal decision processes remain inadequate to users. In deepfake detection Intelligence AI Agent helps user to distinguish between real and fake video. An Intelligence AI Agent-Based Framework for Deepfake Video Detection leverages autonomous multi-agent systems powered by advanced AI models to autonomously detect and deepfake video threats in real-time video streams. the Hybrid InceptionV3-LSTM Framework in visual deepfake videos analyzes pixel-level artifacts including blending boundaries, double edges, eyes, ears, nose, head pose inconsistencies and other possible parameters sub-pixel noise patterns unique to manipulation videos.

2. Background of Deepfakes Video

Deepfake videos represent a transformative yet perilous advancement in synthetic media, originating from early 19th-century photo manipulations that evolved into digital video tampering by the late 20th century. The pivotal breakthrough arrived in 2014 when Ian Goodfellow introduced Generative Adversarial Networks (GANs), pitting a generator against a discriminator to produce hyper-realistic fakes, far surpassing prior autoencoder methods. This technology exploded publicly in late 2017 when a Reddit user named "deepfakes" shared open-source tools like fake app and Face swap enabling amateurs to swap celebrity faces into pornography via accessible GitHub code and consumer GPUs. platforms from deepfacelab and online communities on Reddit and YouTube or social media to political disinformation and fraud. Deepfake technology is, in its essence, a result of advancements made in two disparate fields artificial intelligence (AI) and computer graphics both of which have long sought to synthesize and manipulate digital imagery with seamless precision. The transition from manual editing methods that demanded hours of work from skilled individuals to automated systems powered by sophisticated algorithms mark a remarkable advancement. Manipulation of video content for entertainment or malicious purposes is not new though but it's with the ease, speed and realism provided by Deepfake technology that such a manipulation has become possible, this evolution was driven largely with the growth in machine learning and neural networks [[5,6,7].

Deepfake Challenges

Deepfake technology is rapidly changing the world but it poses some challenges as well. One critical issue is the perverse misuse and exploitation. As deepfakes get more and more believable it is harder to tell between what is real, what is fake. This not only threatens the lives of people whose image or video can be naughtily edited without permission, but also shakes the basis of audiovisual evidence in court cases. In addition, deep fakes can be used to spread misinformation and shape public opinion. The viral quality of these social media platforms has only added to the difficulty of stemming the spread of such misleading materials. Thus, society has to work on viable ways of detecting and disabling deepfakes without infringing privacy rights and preserving trust in digital media [1,8]. Deepfakes can cause a lot of damage if they are misused. Just think about a situation where a deepfake video is made depicting someone as a criminal, whom in reality, the person didn't commit the crime. This can destroy the person's reputation beyond repair and might even land them in legal issues. On top of that the misuse of deepfakes for political purposes is another factor such manipulations can disseminate false information and create distrust for our democratic processes.

II. DEEPFAKES AND AI METHODS

The issue of identifying deepfakes has been approached using a variety of artificial intelligence techniques. One of these methods is deep learning. Its foundation is a neural network with several hidden layers. The complicated nature of the data processed for the learning phases determines how many of these layers are needed [9]. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory (LSTM), variational autoencoders (VAEs), and transformers are the primary deep learning approaches used for solving this particular problem [10,11,12]. CNNs are neural networks that employ data from network theory. They are made up of several hidden layers, an input layer, and an output layer. But after reading the data, the hidden layers apply a mathematical convolution to the input data. Usually, this operation is a dot product or matrix multiplication. A pooling function that adjusts the output by taking into account neighboring outputs is used along with a non-linear activation function such as ReLU (Rectified Linear Unit) [13]. Learning from sequential data is made possible by RNNs. Because of its structure, which forms a cyclic directed graph connecting each neuron to itself, the network can store information from past inputs and use that information to modify the present output.

One similar to network that can manage longer dependencies between data is an LSTM neural network. They have links to earlier nodes, which makes it possible to learn entire data sequences. An input gate the gate to forget and an output gate make up their structure. The output gate chooses which data to send to the following cell the forget gate uses a sigmoid activation function to choose which data to keep and the input gate chooses which values enter the state [14,15].

Neural networks known as autoencoders reduce dimensionality while attempting to faithfully reproduce the input data to the output. They are made up of two components: a decoder aimed to recover the original data by reconstructing the information after it has been encoded. Variational autoencoders are autoencoders that encode data as a distribution over a latent space rather than as a single point. Neural networks with a sequence-to-sequence architecture or transformers, enable you to input a sequence of elements and change them into another sequence. By giving each element in a sequence a variable weight, the encoder uses this mechanism to identify which items are significant. The decoder then focuses on the elements that have a higher weight. The inputs contain positional encoding, which ensures that the model always understands the order of the inserted parts. This is another important distinction from conventional neural networks. These characteristics enable transformers to be a model that tackles certain neural network problems like parallel processing, gradient vanishing and long-range dependencies.

1. Deepfake Datasets

In order for effective training of deep fake detection models, the requisite number of available datasets as well as the data characteristics must be understood. The datasets study is listed below.

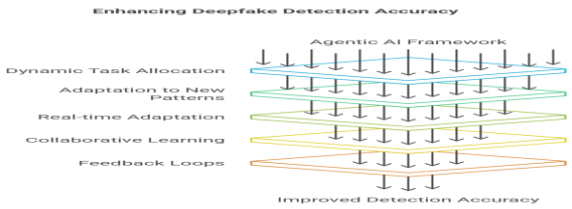
- Created by Facebook in 2019 for its Deepfake Detection Challenge (DFDC), this dataset was used to help improve the identification of video Deepfakes on the Kaggle platform through a major competition. In total the dataset includes more than 128k video face swap experiments using 8 different modifications of the faces involved [16].
- There are 1,000 manipulated videos within FaceForensics++ dataset that used four distinct techniques to create those videos (Deepfakes, Face2Face, FaceSwap, Neural Textures). Along with these manipulated videos, this dataset also contains some sample-deepfake models; This can be used for generating/augmenting dataset for future research activities [17].
- The University of Albany in New York developed UADFV A Dataset Used to Attempt to Detect Deepfakes by Looking for Physical Signals such as facial expressions, movements of lips and eye blink rate [18].

- TIMIT dataset (Deepfake TIMIT) has been constructed so that all of the facial data is based on an original deepfake generation algorithm using autoencoders along with generative adversarial techniques [19].
- The small-scale deepfake forgery video dataset (SDFVD) is designed specifically for research purposes and contains 53 real-based videos and 53 deep faked videos randomly classified based on environments featured in each video. Having a balanced number of both types of videos satisfies machine learning model training requirements, thus aiding researchers in verifying and validating their own deepfake detection algorithms. [20].
- The Celeb-DF (V2) Dataset contains true-life videos and deepfake videos of the equivalent quality as those widely shared on the internet today. It is an enormous expansion over its predecessor Celeb-DF (V1) which only contained 795 deepfake video files. There are 590 unique videos collected from YouTube containing various aged, ethnically diverse, and gender diverse subjects and 5639 deepfake versions of these 590 original videos [21].
- WildDeepfake is a database of 7,314 face sequences obtained from 707 deepfake videos freely available on the internet. Although this is a relatively low number of examples, WildDeepfake fulfills an important need by providing a true-to-life dataset to challenge the baseline deepfake detection algorithms that see drastically lower results when used to detect real-world deepfakes. Through ADD Networks, we present a new method of deepfake detection using two types of attention mechanisms (2D and 3D) to focus on the salient features of real versus fake faces more efficiently [22].
- DFGC dataset is a multi-phase, adversarial competition focused on improving AI security through innovative detection and creating ultra-realistic Deepfakes using identified video and image datasets such as Celeb-DF based on ID similarity or image quality also bypassing detection methods already in place [23].

2. Research Objectives

This paper provides a detailed description of the goals for this study whose objective is to create an innovative Hybrid InceptionV3-LSTM framework for deepfake video detection. The framework aims to respond to the rapidly evolving techniques that are used in the creation of deepfake videos with a focus on improving detection accuracy, assessing and evaluating scalability and analyzing robustness of detection systems. The study objectives listed in this document are expected to significantly advance the field of artificial intelligence and cybersecurity. When taken as an entire they can help create a complete solution.

- Design a hybrid InceptionV3-LSTM framework to detect deepfakes in real time Processing deepfakes in real time needs a hybrid InceptionV3-LSTM model. We build it to adapt quickly to new deepfake techniques. The system uses image and temporal data together. It watches video streams as they come in. This setup keeps up with evolving fake methods. Real-time detection happens without delay.
- **Enhancing Detection Accuracy:** Improving detection accuracy is the main focus because without it this new model would not be very effective. This goal would revolve around using the model's capabilities to allow it to be task-oriented and flexible to entirely new patterns of deepfake making. The main tactics to reach this goal are



shown in fig. 2

- **Evaluate the Scalability of the Model:** To enable the Hybrid InceptionV3-LSTM model to be applied in real world applications it is necessary that the model has sufficient scalability. It is also important that the Hybrid InceptionV3-LSTM can process large amounts of video data as well as connect to multiple detection models. Therefore, one of the primary factors to consider when evaluating the Hybrid InceptionV3-LSTM are the components show in fig. 3

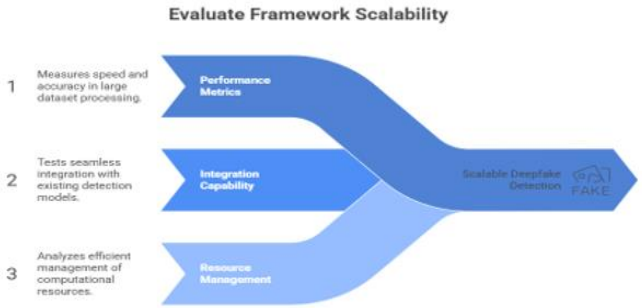


Fig.3 Component Evaluate Model Scalability

- **Model Evaluation:** The hybrid inceptionv3-lstm model's capability to evaluate and mitigate an array of deep fake attacks is also a significant factor of this project. As such, this objective will examine whether or not the hybrid

inceptionv3-lstm model can adapt to unknown data as well as withstand a variety of different deep fake methods. Key areas of analysis will include such as show in fig.4

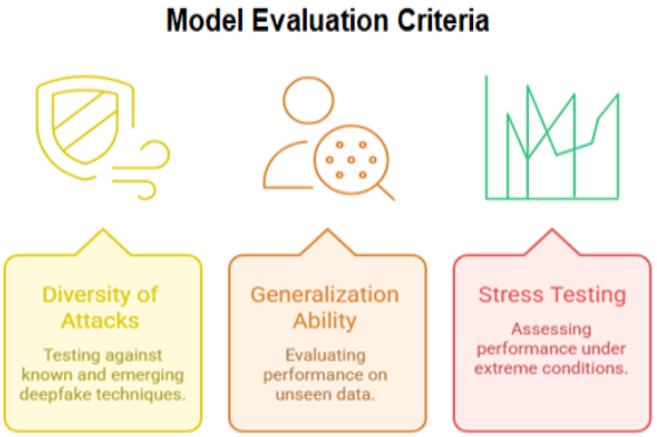


Fig 4 Key areas of Framework Evaluation

3. Research Questions

- RQ1: How can a strong framework be designed to effectively detect deepfake videos in a dynamic and evolving threat landscape?
- RQ2: What are the key factors that influence the accuracy and robustness of an Hybrid InceptionV3-LSTM framework for deepfake video detection?
- RQ3: Can the Hybrid InceptionV3-LSTM effectively handle diverse deepfake generation techniques and generalize to unseen data?

These research questions are addressed by the proposed Hybrid InceptionV3-LSTM for deepfake video detection which aims to provide a more adaptive and robust solution for detecting evolving deepfakes.

Hypothesis

- H1: The Hybrid InceptionV3-LSTM will outperform traditional deepfake detection methods in terms of accuracy and F1-score.
- H2: The framework's adaptive learning mechanism will enable it to detect new, unseen deepfake techniques with high accuracy.
- H3: The multi-agent collaboration will improve the framework's robustness against adversarial attacks and data quality issues.

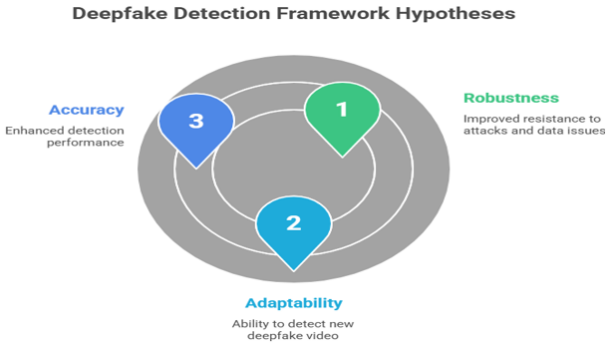


Fig.5. Hypotheses deepfake detection framework

III. RELATED WORK

This study outlines many current technologies to detect deepfakes such as CNNs (Convolutional Neural Networks) and cross-modal network frameworks. it describes several of the limitations with the current state-of-the-art for detection of abuse instances (i.e., particularly concerning scalability, generalizability and ethical governance) for detection. The authors suggest that using detection methods similar to; adversarial training, multi-modal fusions, etc. could be used to improve overall predictive performance. However, there are many concerns about how well the current detection methods will generalize from one context to another. Explainability of detection methods is identified by this study as a major shortcoming [24].

This paper will explore how AI generated deepfakes can be created using many different types of deep learning networks such as CNN's and RNN's and hybrid networks. It will also discuss the strengths weaknesses and possible direction that detection could take in terms of being lightweight, explainable and ethical. As a result of this rapid development in Synthetic Media it is essential to have leading methods identified and categorized into areas of Convolutional Networks, frequency of the audio (spectrogram) and visuals and multimodal frameworks based on Consistency. This paper illustrates how detection strategies will evolve using new techniques in which there is an increased focus on developing methodologies where adversarial robustness (and other types), privacy as well as multi-modal detection. This also outlines some of the current limitations/challenges which are currently hindering AI-based methodologies (deep learning/anomaly detection) from evolving to handle the increasing number of new deep-fake attacks. Furthermore, the authors highlight that the necessity of AI regulation has been demonstrated; however, at this point in time no such regulatory standards exist for deepfake detection [25].

The study emphasizes a need for adequate detection tools for Deepfakes in order to prevent misuses of these fake videos or pictures created using Artificial Intelligence which can nearly perfectly mimic real video/pictures. The paper emphasizes the need of open and transparent AI as well as universally accepted standards (as well as systems) to address the ongoing problems with the detection of Deepfakes. The researchers identify many weaknesses within current detection methods. In addition, they emphasize the importance of enhancing those detection methods. They provide examples of where future research could include integrating multiple data types and utilizing advanced learning methodologies to enhance detection. Finally, the researchers indicate that an effective method of detecting Deepfakes will help counter threats related to Hyper-Realistic Media Created Using AI. Additionally, the researchers emphasize the need for open and transparent AI as well as universally accepted standards (as well as systems) to adequately counteract the challenges presented by Deepfakes. Furthermore, the researchers emphasize the shortfalls currently associated with detection methods. In addition, the researchers emphasize that those detection methods should be enhanced. There is an unresolved issue regarding the adaptability of detection systems to new challenges posed by evolving deepfake technologies, highlighting the importance of developing flexible solutions [26].

This research establishes a significant increase in success rates for identifying deepfake content when using multimodal detection frameworks as a result of exploiting the complementary locations and inconsistencies across different modalities. It also identified that utilizing spatial-temporal consistency, cross-modal feature-alignments, and semantic correlations was necessary to successfully determine if an item was a deepfake or not. The authors indicate that implementations of this technology have significant implications for two critical areas: political communications authentication and digital forensics for the courts. Furthermore, they conclude that the proposed model has the potential to be a considerable benefit when mitigating against highly complex types of forgery attacks, and therefore improve digital security. Therefore, the research details scalable pathways to create intelligent countermeasures for advanced deepfake threats [27]. Recent detection systems are extensively analyzed, and few major architectural patterns are recognized. The one using a combination of ResNet and LSTM achieved a high accuracy rate of 95.2% on the transformed datasets and yet was able to outperform other popular models such as EfficientNet and RNN in terms of detection capabilities [58]

It was found through this initial survey that both the general public and researchers have an interest in how synthetic facial images can be created and how the best deepfake detection systems can be developed. As a result, it is essentially a review paper that covers all previously developed methods of detecting deepfakes presents advanced methods for detecting deepfakes and provides information regarding the databases utilized in several journals. A systematic approach was taken within this study according to the prisma method to identify studies published by reputable authors, to record the results reported in those articles and to utilize a systematic method to obtain and evaluate each article. Multiple algorithms were presented or used including DeepFace, FaceNet, DeepDream, YOLO, CNN-based Extreme Gradient Boosting (CNN, XGBoost), InceptionResNetV2, lightweight 3-D CNN, LSTM, RNN+ and Markov Chains, and XLNET_FT and DeepFakeTIMIT. In addition, the study made use of multiple databases which include FaceForensics, Celeb, DF, UADFV and HVoice. In addition, some of the potential areas of future research are described in the study as being the lack of sufficient amounts of unlabelled data; Temporal Aggregation; Media Manipulation Techniques; Deficiencies in Databases for Training DeepFake Detection Models; Lack of labelled training data for developing deepfake detection models. There is clearly no viable solution at present to effectively detect and prevent attacks utilizing deepfakes. Therefore, it has been determined that a need exists to develop a new generation of AI models capable of effectively fusing signal processing and neural networks; Improve upon the current spatial and temporal analysis; Develop more efficient means of extracting faces; Develop more robust visual cues; Develop more advanced models capable of tracing extracted faces [28].

The research introduces a 3-layer deepfake video detection framework that uses RGB, GAN, and facial region features to spot traces of forgery and surpass the current state, of, the, art works in the thorough experimental evaluations. The article does not mention a hybrid InceptionV3-LSTM, based framework for deepfake video detection at the related work section. Instead, it presents a complete multilayer detection framework with RGB features, GAN features and facial region intra frame inconsistency extraction layers [29].

The study explored deepfake detection framework that leverages XceptionNet a network that uses depth wise separable convolutions to analyze images at a very detailed level. The framework is capable of handling single images as well as videos by recognizing, extracting, and aligning the faces, which facilitates direct comparison at the frame level. In fact, each frame of a video is first examined with XceptionNet, then facial features over time are extracted through a combination of LSTM and CNN to identify the nuances of facial expressions and movements. Based on the results, the

model is able to measure very accurately its precision recall and F1, scores on both the individual frames and the videos. Deepfake detection framework proposed has the capacity to raise the level of media forensics by equipping the forensic practitioners with faultless instruments to tell apart the real from the fake, thus leading to the improvement of the digital media integrity. An advanced technique, XceptionNet, that writes a thorough analysis of the image can detect the deepfakes more accurately, which in turn can be highly beneficial in the scenarios of law and security where the authenticity is of greatest value. The combined use of temporal feature extraction methods like LSTM and CNN makes possible capturing emotions and new variations of movements, which usage could be extended to disciplines like psychology and behavioral studies. By testing the framework on different datasets, you can see its fitness for use in the cutting, edge areas of the future thus its great potential for the world, wide anti, misinformation fight [30].

This paper introduces a new neural network, based approach to deepfake video detection, with an emphasis on extracting only the most important frames in a video to reduce the computational load. The method proposed here takes advantage of a CNN architecture combined with a classifier network and a distinct algorithm to spot deepfake videos precisely while the computational requirements stay very low. The authors demonstrate processing only key video frames results in excellent deepfake video detection performances even when based on single, keyframe processing. Regarding the social and economic implications of false videos, the method is straightforward and effective in video content authentication. After an evaluation on a complete dataset of 600 videos downloaded from various websites, the model is able to detect deepfakes from different databases with an overall accuracy of 97.3%. The paper draws attention to the fact that deepfakes have become one of the major threats to society and politics since they can easily deceive people and at the same time, be turned into weapons. It goes into the details of the complexity of deepfake detection that is so high that new ways of detection need to be invented that are more efficient. The article stresses the urgent need for fool, proof detection technology since deepfakes can be used to create serious mistrust in video evidence amongst people, which can trigger various social and economic issues. The paper also highlights the effectiveness of only working on a few key video frames to achieve better detection results, which means that the conventional methods are unlikely to cope with the massive production of deepfake materials [31].

The study suggests the existence of an impressive system that can successfully identify Deepfake content and AI-created images or video. It is created with the purpose of addressing the problems of digital media integrity and security in the most "head-on" way. It utilizes the "most advanced" machine learning techniques such as "CNNs" and "GAN discriminators" in the identification of the "smallest" traces of synthetic content. To increase detection accuracy, a multi, modal analysis approach is used, which incorporates visual, spatial, and frequency, domain features. The system is capable of generalization over various manipulation techniques as it is trained on a heterogeneous dataset of both publicly available and custom, generated deepfake and AI, generated images. The experimental results demonstrate that the suggested system excels in terms of precision and recall rates, thus it significantly helps in the fight against digital misinformation. Used publicly available datasets custom-generated deepfake and AI-generated images. The original poster Deepfake and AI, Generated Image Detection System take the lead in strengthening the truthfulness and reliability of digital media by figuring out fake and synthetic content, thus greatly contributing to the fight against misinformation and helping retain people's trust in what they see. The system's strong and unshakable detection abilities are just the right tools that security agencies can use to block criminals from stealing identities and cheating people, hence, preventive measures for individuals and businesses being victimized through the unlawful use of deepfake technology. Solution of research based on cutting edge machine learning and multi, modal analysis, can serve as a tailor, made tool for a wide range of industries e.g., journalism, law enforcement, which can then use it to verify the visual evidence [32].

The study suggested hybrid framework for deepfake detection integrates a lightweight model that, among other features, takes full advantage of spatial, temporal feature analysis and also adaptive adversarial noise estimation and reduction. The Xception backbone with temporal attention layer is able to find the discrepancies between the compressed video frames. The model has managed to reach almost the best of the best level success rate of 90% accuracy on Celeb, DF, v2 dataset thus exceeding current first, class solutions by 5.7%. Not only this, but the model has also proven very powerful and effective on real, world video distortion problems as well as different dataset environments. Given the model's high performance and flexibility, it is especially appropriate for social media platforms, where it can be used to battle the great number of synthetic media threats that keep on changing. The fast progress of AI technology has resulted in the production of extremely lifelike deepfake media content, making it very difficult for the current detection methods to find them. The large, scale dissemination of fake content seriously endangers the trust in the media, making it crucial to put in place reliable detection

mechanisms that can handle privacy and security matters. The issue of adaptive adversarial noise removal challenging detection models' effectiveness is that the models have to be robust enough against different types of noise in the video data. The model's effectiveness in addressing real-world video distortion issues and different dataset environments demonstrates the need for continuous upgrades in detecting video distortion [33,18].

This study suggests a hybrid deep learning framework that integrates ResNet50 and LSTM for the detection of deepfake videos, reaching 94.98% accuracy, 98.06% precision, and 97.06% F1 score on the Celeb, DF (v2) dataset, thus contributing to changes in digital media forensics and multimedia authentication systems. This paper proposes a hybrid deep learning model that integrates Convolutional Neural Networks (CNNs) for the extraction of spatial features and Long, Short, Term Memory (LSTM) networks for capturing the temporal relationships. With the help of the model, it is possible to tell apart real and AI, generated video quite well since the model is able to learn both spatial and sequential patterns of video frames. The model is trained and tested on the Celeb, DF (v2) dataset. Preprocessing involves the extraction of frames, their resizing, and normalization to better represent the features. The outcomes of the experiments demonstrate that the proposed model is capable of delivering very good results on the test set accuracy 94. 98%, precision, 98. 06%, recall, 96. 08%, F1 score, 97. 06%, AUC, ROC, 98. 32%Additionally, the confusion matrix and loss accuracy curves tests certify the model's capability in detecting AI, generated content. The present study helps to make deepfake detection methods more effective, thus multimedia authentication systems can be more secure and trustworthy [34].

The blockchain-based augmented intelligent framework proposed in the paper is a truly innovative solution to the problem of deepfakes and has the potential to provide a robust solution for the detection and mitigation of such media. The proposed system has the advantage of providing 97% accuracy with a false positive rate of 3%. This ensures the security and accuracy of digital visual media against the threat of forgery. There are many other ways to enhance the accuracy and effectiveness of the blockchain-based augmented intelligent system in detecting deepfakes. Looking into combining more AI models and approaches could lead to a higher level of security in the detection system against the advancement of deepfake techniques. Scientists might be interested in utilizing the blockchain-based augmented intelligent method in a broader range of digital media types besides just images e.g., audio and text [35].

The paper proposes a method that can detect deepfakes from the abnormalities of the production processes by using the integration of spatial and temporal features. A novel source attribution technique is proposed which makes use of metadata examination, network forensics, and linguistic features to identify potential generators of deepfakes. According to the authors, the Inception, ResNet, V2 + LSTM model outperformed other models with a detection accuracy of 92% which paved the way for new Deepfake detection technologies. Deepfake technology is advancing rapidly which poses a major threat as it undermines the authenticity of the digital content and hence leads to misinformation. Deepfake creators intentionally employ complex mechanisms to hide their alterations leading to the methods for identifying subtle anomalies in Deepfake production becoming increasingly sophisticated. The challenge in developing a highly effective method against contemporary deepfake technologies is reflected in the call for a comprehensive framework that incorporates different detection methods [36,15].

In this paper explored systematic review of the literature related to techniques for automatically detecting deepfakes focuses on a broad range of algorithms, frameworks, and tools that utilize CNNs (Convolutional Neural Networks) and multi-task cascaded cnns, among others, for accurate detection of deepfake videos across the media domain. As there is an increasing need for accurate assessment of the reliability of data, what is being presented as credible sources of information on the internet presents a significant challenge for individuals attempting to identify false and misleading content in a digital world saturated with misinformation. This situation has been further complicated by the fact that there are many different online techniques available for both detecting and generating deepfakes, which makes it extremely difficult to create viable countermeasures. However, the fact that deepfake detection requires state-of-the-art AI solutions highlights the reality that the incursion of highly advanced and rapidly evolving deepfake technology is too fast to keep pace with. In addition, incorporating a number of different modalities (e.g. audio, video, images) into detecting deepfakes will present challenges for constructing comprehensive frameworks that are capable of addressing the many different contexts in which deepfakes arise [37].

In this article going to develop and design a new hybrid model using existing ones (EfficientNetB7, CapsNet, Transformer Encoder) for the purpose of having better detection and interpretability. Existing models often do not provide users with a clear understanding of how they classified an image and gave it an output. Because of that, there can be a lack of trust and understanding in the output from a model. Many current

models do not consistently perform well when applied to the types of datasets being used within today's real-world applications. The underlying reason for this is that the data being used by most of these models may differ significantly from each other. The use of deepfakes can alter what we consider as reality and hence create misinformation and manipulation. Need to have solid detection frameworks in place to maintain the integrity of our media. Due to the increase in synthetic media, there exist many security vulnerabilities, which indicates the necessity of having strong detection systems to address the potential exploitations of deepfake technology [38,3,35].

The paper introduces DeepFake Bot, a system powered by AI that is capable of identifying manipulated videos with a very high level of accuracy. The organism employs Convolutional Neural Networks (CNNs) to conduct spatial analysis and Recurrent Neural Networks (RNNs) to check for temporal consistency. Some of the principal detection methods are examination of the eye, blinking timing, finding irregularities in the facial texture, and identifying anomalies of the movement. DeepFake Bot has been exposed to publicly available deepfake datasets which made it capable of handling various kinds of manipulations. Its performance metric in terms of accuracy was 92.4%, which is significantly higher than that of the other deepfake detection models while the capability for real, time processing was preserved. Earlier methods of detection are not able to efficiently identify deepfakes because of the complexities of Generative Adversarial Networks (GANs) that in turn create the need for next, generation detection systems. The fast progress of the deepfake technology is making it a herculean task to spot manipulated media, thus leading to great vulnerabilities in digital security and misinformation. As the number of deepfakes keeps on increasing, there is a need for an automated and scalable detection system that will guarantee the authenticity of the media. In fact, the most important trait of a detection system is that it should be able to generalize different deepfake techniques in order to remain accurate and reliable at identifying manipulated media [39].

This paper presents a framework that combines ethical AI and explainable AI (XAI) methods to enhance the deepfake detection system with transparency, trustworthiness, and accountability features which facilitate its deployment in the real, world systems for the fight against malicious activities and misinformation. A large number of deepfake detection models function as black boxes resulting in reluctance among users as their trust level is low when they cannot understand the model decisions. If the decisions made by deepfake detection systems are not traceable, then it is difficult to hold the system accountable and therefore, one faces a challenge when the system's error or bias needs to be corrected [40].

The paper argues for a strong AI, based authentication framework to be present in a digital identity system to be able to deal with deepfake threats efficiently. In fact, the integration of behavioral biometrics and AI powered verification layers to strengthen conventional authentication methods is highly encouraged. Authors have recognized federated learning and privacy-preserving technologies as methods that allow safeguarding personal data without causing the loss of system performance. Besides that, the author has flagged the importance of ethics and following regulations related to AI as very important factors when deploying AI in authentication systems. Through this research author can be informed about how to use stringent identity verification processes to counter the more advanced digital impersonation attacks [41].

Methodology utilized in this research study consisted of GANs, RNN's, and CNNs. Through the use of artificial intelligence methodology, specifically Generative Adversarial Networks (GANs), Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs), a deepfake detection system was devised that can operate in real time and has a success rate of 88%. Of all the algorithms evaluated during this research study, GANs had the best performance as compared to the other algorithms, and this demonstrates that GANs have potential for being reliable detectors of deepfakes in complex and fast-paced scenarios [42,12].

The paper introduces a deepfake video detection system based on advanced convolutional neural networks from the beginning to end. It also adds a temporal component to the processing for further identification of deepfakes. The method attained high metric of performance, including an accuracy of 94.7%, a precision of 93.2%, and a recall of 95.1%. The network is capable of cross, dataset generalization and hence it performs very well on different datasets like DFDC and CelebDFV2. It is smaller compared to the robustness of the contemporary real-time applications. Therefore, the reliability in the presence of inconsistent video quality and variable levels of compression. The ethical implications and the effects of the use of such detection systems on society are not considered in the paper. There is very little information on how the model works in real, life situations outside of laboratory test environments which make up less than 1% of the paper. It also does not discuss the model's scalability to large datasets or real, time scenarios, a feature that is very important for the adoption of the technology on a wide scale. The issue of whether the model can perform well on different datasets is a major one since the effectiveness

of the detection system in varied real, world situations depend on it [43].

The paper introduces variational autoencoder (VAE) to extract features from videos, this study expands its research on detecting deepfake video content, the use of Vision Transformers (ViTs) for extracting spatial and temporal dependencies in video can further enhance detection of Deepfake videos by allowing the model to analyze temporal aspects of video. To classify manipulated videos, this study develops a generator-discriminator model that uses concepts from Generative Adversarial Networks (GANs) for classification and MTCNN (Multi-task Cascaded Convolutional Networks) that aid in detecting and aligning faces within videos therefore increasing reliability for detection by overcoming problems like occlusions, variable lighting, and sophisticated manipulation methods in detecting Deepfake videos. In conclusion the authors present a novel framework for detecting Deepfake videos using a combination of VAE, ViTs, and GANs, achieving superior performance than previous works on benchmark datasets with lower false-positive rates compared to previous studies [43,36,16].

The authors put forward a more dependable evaluation system for gauging the performance of learning based deepfake detectors in real world scenarios, which not only gives an overall idea of the detector's performance under real world conditions but also quantitatively evaluates the detector's robustness against various processing operations. Deepfake detectors are not robust to real world processing operations. Even slight distortions can greatly hinder detection accuracy. Developed a trustworthy evaluation framework for performance of deepfake detectors. Created a data augmentation strategy for enhancing the robustness of the detector [44].

A hybrid deep learning convolutional network (CNN) model based on the VGG19 and InceptionV3 architecture will be proposed in this paper. The VGG19 neural network in the proposed model will be used for obtaining fine-grained texture-based features from video frame images, while the InceptionV3 network will be used to learn complex patterns across different scales. The authors have incorporated a Multi-Task Cascaded Convolutional Network (MTCNN) for face detection and alignment of the person depicted in the video frame images. The performance of the proposed hybrid deep learning model was evaluated on standard datasets and the results from the experiments were analyzed to determine how accurate the MTCNNs with respect to the performance characteristics (accuracy, precision, and stability) of the hybrid model and its ability to provide improved accuracy while detecting deep fake technology [36,45].

The study constructs an innovative framework based on blockchain technology for the detection and authentication of deepfake video content. Deep learning and decentralized verification techniques were combined to create this framework in order to achieve the highest level of accuracy. Frame-level features were created using a ResNeXt Convolutional Neural Network, which will improve the detection and verification of deepfake videos. In addition, a Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) was developed to analyze temporal information from multiple video frames as part of the analysis process. The results and metadata are stored immutably on a blockchain ledger; therefore, there is an audit trail with no possibility of alteration or tampering, making it an excellent method for video authentication. The experimental results show that the proposed blockchain-integrated framework out performs all other current methods of deepfake detection and verification, providing an additional layer of security for multimedia authentication [46].

The paper discusses the major security and misinformation risks that deep fake technology raises to society apart from other issues. It points out how tricky it is to detect deep fakes of very low quality and at the same time it admits that recognizing deep fakes of different qualities is a big challenge.

The release of the Celeb, DF dataset that features high, quality deep fake videos of celebrities has been a headache for detection methods set up so far. The paper describes a new Quality Agnostic Deep Fake Detection (QAD) system that is able to identify deep fakes of various qualities thereby significantly exceeding the performance of existing methods on several datasets. It is acknowledged that, among other things, deep fake detection methods are continuously being improved by the application of powerful models such as Stable Diffusion and the combination of Vision Transformers with Support Vector Machines for better interpretability [47,40].

This work investigates deepfake video detection from spatial and temporal perspectives, using a contrastive spatiotemporal distilling approach. A two - stage spatiotemporal video encoder is applied, and two distillation modules are proposed. Extensive experiments validate the robustness of the approach on highly compressed deepfake datasets. Future work may broaden the scope of the distillation framework to other video understanding tasks and enhance its interpretability [48].

IV. RESEARCH GAP

The development of robust and generalizable deepfake detection systems particularly those incorporating Hybrid InceptionV3-LSTM frameworks faces several critical research gaps despite significant advancements in the field. Deepfake technology, including generative technology such as GANs and diffusion models is becoming increasingly capable. Indeed, deepfakes are able to fool even the prime detection systems. Besides the absence of standard real-time end-to-end protocols for multiplying the new deepfake detection features is one of the major challenges in the evolution of deepfake technology which entails the involvement of several Model. To address this, multi-agent fusion has been proposed for deep agent architecture in deepfake detection. [49].

However, such systems are frequently negatively affected by modality mismatches, noise, manipulation and a standard method for organizing these agents is still very much under development. The integration of different detection techniques, such as CNNs, Transformers, and Bi-LSTMs, into comprehensive frameworks for improved accuracy and interpretability is an active area of research but the coordination among these diverse components remains a challenge.

Many deepfake detectors are initially trained on a limited number of datasets, each of which corresponds to a certain deepfake generation technique. As a result, their ability to detect deepfakes drastically drops when they face deepfakes generated by new or different generative adversarial networks (GANs) or diffusion models or when the content is different from that which the model was trained on. Most existing video level methods do not completely utilize spatiotemporal inconsistencies in low quality, highly compressed deepfake videos which results in poor generalization and robustness.

Deep learning, based detection methods are typically built around highly sophisticated architectures such as Convolutional Neural Networks (CNNs), Transformers, and Recurrent Neural Networks (RNNs). These detectors mainly function as "black boxes, " which means that it is hard to know how and why they draw certain conclusions regarding whether a particular piece of media is fake or real. This situation is not only detrimental to trust but also to debugging and deploying these methods in sensitive cases.

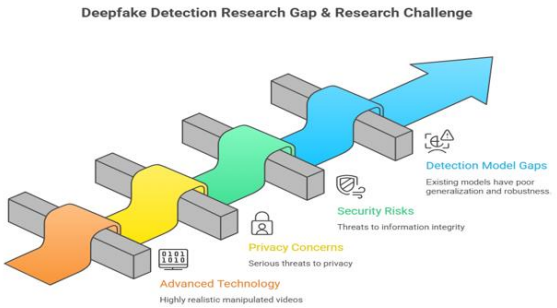


Fig.6 Deepfake Detection Research Gap & Challenge

V. METHODOLOGY

This section describes how the hybrid InceptionV3-LSTM framework for real-time deepfake detection for has been designed implemented and experimented. It covers the architectural elements data management methods agent particular features underagent communication and experimental setup. The methodology of the research on Hybrid InceptionV3-LSTM framework for real-time deepfake video detection combined a multi-phase deepfake detection frameworks analysis design and evaluation methods from a qualitative perspective. Basic steps of work follow as:

- Literature Review: Literature review has been conducted which covers various papers, market reports and cover deepfake video detection techniques as well as those which can recognize patterns in machine learning architecture obtained from reliable sources such as IEEE, Springer, ACM, Elsevier and Google Scholar academic papers
- Data Collection: Both videos have highly emotional content, including real videos and deepfakes, intended to show realistic scenarios. To ensure comprehensive training and evaluation the framework utilizes a combination of publicly available deepfake datasets encompassing various manipulation techniques and quality levels. The collection of datasets obtained from reliable sources Kaggle, ACM, and github.
- Feature Engineering and Preprocessing: Applied a standardized preprocessing method to both datasets to maintain consistency. Frames of both real and fake videos were extracted and organized into separate folders. To achieve reasonable temporal coverage while keeping computation affordable frames were sampled every 20-frame time step in data set. The data are cleansed, normalized and dimensionality reduction methods are also applied to prepare the data to the model and apply various data augmentation techniques.
- Model Development: The group is using several AI and machine learning models in their detection capability.

These include the use of supervised models such as decision trees, SVM, neural networks and unsupervised models like clustering anomaly detection and deep learning techniques.

- Validation and Evaluation: Model behavior on unseen data with the help of proper metrics such as precision, recall, F1-score, ROC-AUC, and false-positive rates. Also, cross-validation and live simulation can be utilized to check the stability and the capability to follow the changes of deepfake video and manipulation contents.
- Implementation and Integration: Work on prototype detection systems or modules specifically designed for deepfake video detection platforms with the need to consider real-time processing limited resources and user privacy. Also, the integration with the current detection system being investigated.

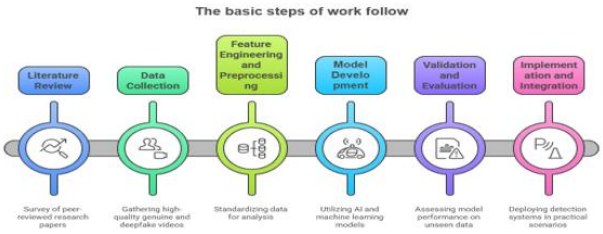


Fig. 5. The basic steps of work follow

Our method utilizes a InceptionV3 Convolutional Neural Networks to get frame level features. After that these frame features are fed to a Long Short, Term Memory (LSTM) based Recurrent Neural Network (RNN) for classifying the video as manipulated or not, i.e., distinguishing between deepfake (fake) and real videos. We decided to conduct experiments by testing the proposed method on a deepfake video dataset generated from various video websites. We are working hard to develop a deepfake detection model that performs better in real time scenarios. For that purpose, our training was based on a combination of existing datasets. Hence, our model gets an opportunity to learn features belonging to different kinds of videos. got the adequate number of videos datasets from the Small Scale Deepfake Forgery Video Dataset (SDFVD) [20], Deepfake detection challenge [16] and Celeb, DF datasets [21].

Analysis of Deepfake Video Parameters for Enhanced Accuracy

This section provides an analysis of various video parameters that can significantly influence recognition models. By identifying and understanding these parameters, we can improve the accuracy and reliability of deepfake video recognition technologies. The parameters discussed herein include physical attributes, expressions and environmental factors that affect deepfake video recognition. Parameters to be consider such as:

Blinking of Eyes

Blinking is a natural and involuntary action that can affect deepfake video recognition systems. It is not easy to get a clear shot of the eyes, even though the eyes are a vital identifier. The system should be designed in such a way that it can cope with this ever-changing feature by utilizing time data to monitor eye movement.

Teeth Enchantment

The appearance of teeth can also affect how a face is perceived. Even a smile can modify the appearance of a face to a great extent. Any system that analyzes faces should also consider the presence or absence of teeth as this can be a unique identifying feature.

Bigger Distance for Eyes

The distance between people eyes is not equal, and this distance can contribute significantly to face recognition in deep fake videos. Therefore, this aspect needs to be taken into consideration so that face recognition can be carried out in a more accurate manner.

Moustaches

Facial hair including moustaches can cover up parts of the face and alter the appearance of the actual shape. One should make sure the recognition system is capable of identifying people both with and without facial hair so that presence of facial hair does not have any impact on the recognition.

Double Edges, Eyes, Ears, Nose

The area surrounding important facial features like eyes, ears, and nose plays an important role in the detection of a face. The differences in these areas may be complicated but they are significant. For increasing the precision of facial recognition, sophisticated algorithms should be used to focus more on edge detection.

Iris Segmentation

Iris segmentation plays an important role in the accurate recognition of eyes. The patterns formed in the iris can be considered an identifier. Recognition systems should have powerful segmentation techniques to distinguish the iris from other structures in the eye.

Wrinkles on Face
Wrinkles show the actual age of the individual and vary considerably among different people. It is important for the recognition systems to identify and evaluate wrinkles as part of the structure of the face because they add additional context to the actual age and identity.

Inconsistent Head Pose
The difference in orientation of the heads can significantly the level of recognition accuracy. There is a need to have algorithms in place to recognize individuals despite differences in direction.

Face Angle
The angle from which a face is seen can change its shape. However, recognition should be made taking into account that for the same subject there can be different aspects by using 2D & 3D modeling systems that allow analysis under different approaches to guarantee that in whatever position is taken still should reflect a similar identification.

Skin Tone
Skin tone is an important factor in the recognition of faces. The model must be accommodating and have the ability to identify all kinds of skin tones. This is an important factor in the reduction of bias and the improvement of the efficiency of the recognition system.

Facial Expressions
Skin tones are very instrumental in the recognition of faces. A strong model must be able to recognize all tones of skin. This will reduce bias in the system, which will, in turn, make the recognition process more efficient. Facial expressions can greatly change the appearance of the deepfake video. The system must be trained to recognize all possible expressions, as they can interfere with the recognition process. Understanding the emotional state will improve the contextual understanding of the recognition deepfake video.

Lighting
Facial features may also appear differently depending on lighting conditions. Recognition systems ought to be robust under different lighting conditions and this can be achieved through histogram equalization or adaptive lighting weak.

Double Chins
The appearance of the face in the image created by the deepfake can be changed by the addition of a double chin, and this changes the appearance of the face. The recognition system must be flexible enough to accommodate the change and ensure that the system recognizes the image correctly, irrespective of the position and shape of the body.

Hairstyle
It is also important to note that hairstyles have an impact on the manipulation of videos. The recognition systems must be designed to recognize a variety of hairstyles and their impact on deepfake detection to ensure accuracy.

These parameters as explained in the above sections give us an idea of how complex deep fake video detection systems. Therefore, if we understand and incorporate these parameters, we can develop a more accurate and reliable recognition system as show in figure 7.

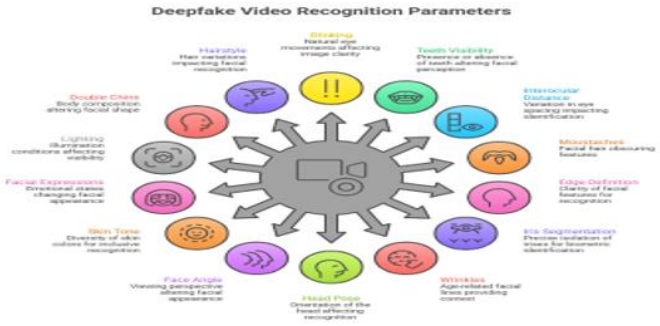


Fig.7 Deepfake Video Parameter Recognition

VI. PROPOSED MODEL ARCHITECTURE

The proposed framework for deepfake video detection is based on a multi-stage architecture that is both spatial and temporal in nature. Initially, the video is split into individual frames. Then, the faces are detected and normalized for better processing. Each of these frames is processed by a deep convolutional network called InceptionV3. This network is a high-dimensional spatial representation of the frame that includes information about texture, edges, and other important features. The output of the InceptionV3 network is sequentially processed by a Long Short-Term Memory (LSTM) network that includes information about temporal changes such as irregularities in blinking and expression. The framework includes four different artifact modules:

- A physiological module for modeling blinking patterns and iris segmentation consistency.
- A geometric module for analyzing inter-ocular distance, head pose, and facial angle deviations.

- A texture module for detecting wrinkles smoothing, teeth enhancement, moustache blending and double-edge artifacts
- An appearance and lighting module for evaluating skin tone variation, illumination coherence, hairstyle segmentation, pose diversity, and jawline contour anomalies. The outputs of these modules are then combined through an attention-based aggregation mechanism that assigns adaptive weights to each feature of the artifact depending on the reliability of the feature.

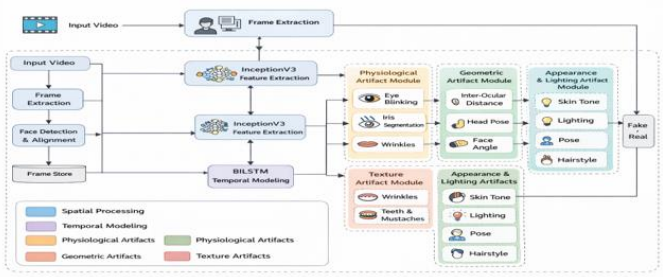


Fig. 8 Workflow of proposed model of deepfake videos detection

Experimental Setup

Dataset

The deepfake detection model proposed here has been tested on datasets which are publicly accessible and commonly used standards in the domain of deepfake forensics. Our major dataset for this research has been the Celeb-DF (V2) Dataset [21], Small-scale deepfake forgery video dataset (SDFVD) [20], DFGC dataset [23] that comprises high-quality real and deepfake videos. The dataset includes both authentic videos and synthetically manipulated videos generated using deepfake generation techniques. The dataset was divided into three subsets:

- Training Set: 70% of the videos
- Validation Set: 15% of the videos
- Testing Set: 15% of the videos

Data Preprocessing

To prepare the video data for model training, several preprocessing steps were applied.

- Frame Extraction: Each input video was first decomposed into frames using a frame sampling strategy. Frames were extracted at 20–25 frames per second (FPS) to reduce redundancy.
- Face Detection and Alignment: After frame extraction, faces were detected in each frame using a face detection model. The detected faces were then aligned to normalize orientation and scale. The cropped face regions were resized to 1280× 720 pixels which is the required input size for the feature extraction.

Feature Extraction

Spatial features were extracted using the deep convolutional neural network InceptionV3. The network was pre-trained on the ImageNet dataset and fine-tuned for the deepfake detection task.

Temporal Modeling

To capture temporal inconsistencies across frames the extracted feature vectors were feed into a Bidirectional Long Short-Term Memory (BiLSTM) network. The BiLSTM layer models temporal dependencies between consecutive frames and detects anomalies such as:

- Irregular eye blinking patterns
- Sudden facial motion inconsistencies
- Temporal distortion between adjacent frames

Artifact Detection Modules

The architecture incorporates multiple artifact detection modules to capture different categories of deepfake inconsistencies. This module analyzes biological and natural inconsistencies that frequently appear in deepfake videos. It includes:

- Eye blinking analysis
- Iris segmentation
- Facial wrinkle detection
- Skin tone variations
- Lighting inconsistencies
- Hairstyle inconsistencies
- Teeth
- Moustaches
- Wrinkles

Geometric Artifact Module

This module evaluates structural inconsistencies in facial geometry. The following geometric features are analyzed:

- Inter-ocular distance
- Head pose estimation
- Face angle variations

Implementation

The proposed deepfake video detection framework was implemented using the cloud-based development environment Google Colab, which provides GPU acceleration for training deep learning models. The entire implementation consists of several stages including dataset preprocessing, model training and evaluation. First, the dataset was downloaded and uploaded to Google Drive. The Google Drive storage was then installed

in the environment for fast dataset access. The necessary libraries including TensorFlow, OpenCV, NumPy and Keras were installed and imported. Next, video preprocessing was performed where each video was converted into frames using OpenCV. Before supplying to the feature extraction network the face images were resized to 1280 x 720 pixels. The spatial feature extraction stage utilized the pretrained convolutional neural network InceptionV3 is used to get the deep visual features from every frame. The extracted frame features were then arranged sequentially and passed to the Long Short-Term Memory (LSTM) network is utilized to model the temporal relations between the frames. Then the merged feature vector was forwarded to a fully connected classification layer with SoftMax activation function to distinguish the video as genuine or deepfake.

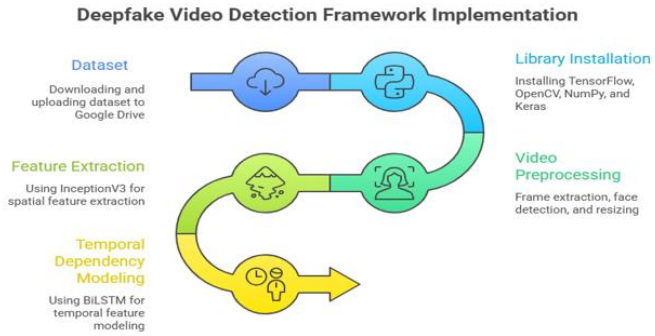


Fig. 9 Implementation Deepfake Video Detection Framework Compares to Existing Model

Our InceptionV3-LSTM framework, when compared to other deepfake detection methods that are currently available is undoubtedly the leader in the field. The drop in performance gap between our model and other models mainly depends on how efficiently our system allocate and divides the analytical workload. The function of InceptionV3 is very similar to a magnifying glass visually. Because this pre-trained model already has a strong track record in complex computer vision tasks it is highly reliable at scanning individual video frames to catch the tiny intricate visual artifacts left behind by manipulation tools. Identifying a highly sophisticated deepfake goes beyond simply examining the still images as it demands an understanding of the motion. That is exactly where the LSTM network plays a role by handling the video frames one after another. LSTM keeps track of the passage of time looking for the mishaps, subtle changes or unnatural flickering between frames. Most models fail to notice these brief temporal hiccups but in our system this time-based analysis is in fact what contributes to the high success rate.

Comparison table of existing deepfake detection model as

Method	Accuracy	Reference
CNN-based Detection	85.3%	[51]
XceptionNet	90.2%	[52]
Capsule Networks	91.8%	[53]
EfficientNet-based Model	93.4%	[54]
ResNext CNN and LSTM	95.5%	[55]
InceptionV3-LSTM	96%	

VII. CONCLUSION

In this paper we propose a new deep learning model for detecting deepfake videos.The model is an integration of InceptionV3 and LSTM technology. As synthetic media becomes increasingly sophisticated telling real videos from manipulated ones is more challenging. In fact, after carefully evaluating our model on well-known benchmark datasets including FaceForensics++, Celeb-DF and DFDC we achieved a highly impressive 96% accuracy rate, reflecting a very strong reliability of the model. The reason why this system performs so excellently lies in the manner that the two underlying techs operate InceptionV3 analyzing the visual features of single frames to identify inconsistencies, whereas LSTM network - at the continuity of the video revealing movements that do not fit the natural pattern. Our testing clearly showed that these two components are better together the combined system outperformed individual models by a significant margin. Our model can be used on different kinds of deepfake and is still potent enough to perform at the level of deployment in the real world like in digital forensics, content moderation, and media verification. Our plan to explore ways to make the system resistant to targeted adversarial attacks analysis alongside the video and build lighter and faster versions. We are also looking into using attention mechanisms tools that can highlight exactly where and when a video looks tampered with making the AI decisions much easier for humans to interpret.

REFERENCES

1. A. Kaur, A. Noori Hoshyar, V. Saikrishna, S. Firmin, and F. Xia, “Deepfake video detection: challenges and opportunities,” Artificial Intelligence Review, vol. 57, no. 6, Jun. 2024, doi: 10.1007/s10462-024-10810-6.
2. N. Mansoor and A. I. Iliev, “Explainable AI for Deepfake Detection,” Applied Sciences (Switzerland), vol. 15, no. 2, Jan. 2025, doi: 10.3390/app15020725.

3. G. Gupta, K. Raja, M. Gupta, T. Jan, S. T. Whiteside, and M. Prasad, "A Comprehensive Review of Deepfake Detection Using Advanced Machine Learning and Fusion Methods," Electronics (Switzerland), vol. 13, no. 1. Multidisciplinary Digital Publishing Institute (MDPI), Jan. 01, 2024. doi: 10.3390/electronics13010095.
4. M. Javed, Z. Zhang, F. H. Dahri, and A. A. Laghari, "Real-Time Deepfake Video Detection Using Eye Movement Analysis with a Hybrid Deep Learning Approach," Electronics (Switzerland), vol. 13, no. 15, Aug. 2024, doi: 10.3390/electronics13152947.
5. Z. Yan, Y. Zhang, X. Yuan, S. Lyu, and B. Wu, "DeepfakeBench: A Comprehensive Benchmark of Deepfake Detection." [Online]. Available: https://github.com/ondyari/FaceForensics
6. Z. Sun, Y. Han, Z. Hua, N. Ruan, and W. Jia, "Improving the Efficiency and Robustness of Deepfakes Detection through Precise Geometric Features," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 3608–3617, 2021, doi: 10.1109/CVPR46437.2021.00361.
7. N. Ur Rehman Ahmed, A. Badshah, H. Adeel, A. Tajammul, A. Daud, and T. Alsahfi, "Visual Deepfake Detection: Review of Techniques, Tools, Limitations, and Future Prospects," 2025, Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ACCESS.2024.3523288.
8. R. Chesney and D. K. Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," SSRN Electronic Journal, Aug. 2018, doi: 10.2139/ssrn.3213954.
9. Á. F. Gambín, A. Yazidi, A. Vasilakos, H. Haugerud, and Y. Djenouri, "Deepfakes: current and future trends," Artif Intell Rev, vol. 57, no. 3, Mar. 2024, doi: 10.1007/s10462-023-10679-x.
10. Ahmed, S.R., Sonuç, E., Ahmed, M.R., & Duru, A.D. (2022). Analysis Survey on Deepfake detection and Recognition with Convolutional Neural Networks. 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 1-7.
11. A. v Jamsheed, "Deep Fake Video Detection Using Recurrent Neural Networks," International Journal of Scientific Research in Research Paper. Computer Science and Engineering, vol. 9, no. 2, pp. 22–26, 2021, [Online]. Available: www.isroset.org
12. Chen, B., Li, T., & Ding, W. (2022). Detecting deepfake videos based on spatiotemporal attention and convolutional LSTM. Information Sciences, 601, 58-70.
13. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press. http://www.deeplearningbook.org
14. M. Soumya, "'Deepfake Detection Using ResNeXt and LSTM: A Hybrid Deep Learning Approach.'" [Online]. Available: www.rnsit.ac.in
15. T. S. Harikrishnan, S. Thomas, N. Simon, S. Johny, J. Ann John, and A. Professor, "A Comparative Study of Deepfake Detection Using ResNeXt50_32x4d + LSTM, EfficientNet + GRU, and Xception + Transformer Encoder." [Online]. Available: www.ijsdr.org
16. B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer. The deepfake detection challenge (dfdc) dataset, 2020.
17. A. R¨ossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner. Faceforensics++: Learning to detect manipulated facial images. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pages 1–11, 2019. doi: 10.1109/ICCV.2019.00009
18. Y. Li, M.-C. Chang, and S. Lyu. Inictu oculi: Exposing ai created fake videos by detecting eye blinking. In 2018 IEEE International Workshop on Information Forensics and Security (WIFS), pages 1–7, 2018. doi: 10.1109/WIFS.2018.8630787.
19. P. Korshunov and S. Marcel. Deepfakes: a new threat to face recognition? assessment and detection, 2018.
20. Kaman, Shilpa; Makandar, Dr. Aziz (2024), "SDFVD: Small-scale Deepfake Forgery Video Dataset", Mendeley Data, V1, doi: 10.17632/bcmkfgct2s.1
21. Li, Y., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-DF: A large-scale challenging dataset for deepfake forensics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 3207–3216). IEEE. https://doi.org/10.1109/CVPR42600.2020.00328
22. Zi, B., Chang, M., Chen, J., Ma, X., & Jiang, Y.-G. (2020). WildDeepfake: A challenging real-world dataset for deepfake detection. In Proceedings of the 28th ACM International Conference on Multimedia (pp. 2382–2390). Association for Computing Machinery.
23. B. Peng et al., "DFGC 2021: A Deepfake Game Competition," 2021 IEEE International Joint Conference on Biometrics (IJCB), Shenzhen, China, 2021, pp. 1-8, doi: 10.1109/IJCB52358.2021.9484387.
24. Gowsalya, S., & Devi, S. (2025). Deepfake Detection in the Era of Multimedia: Methods, Gaps, and Evolving Research Directions. 2797–2801. https://doi.org/10.38124/ijisrt/25jul1768
25. Eidan, S., & Eadan, S. (2025). Unmasking Deepfakes: A Systematic Review of Generation Techniques and Detection Strategies. Iraqi Journal of Intelligent Computing and Informatics (IJICI), 4(2), 134–154. https://doi.org/10.52940/ijici.v4i2.105

26. Kumar, N., & Sharma, C. (2025). Deepfake Detection Methods and Trends: A Comprehensive Analysis. 1–5. https://doi.org/10.1109/isac364032.2025.11156395
27. Goyal, H. O., Wajid, M. S., Wajid, M. A., Khanday, A. M. U. D., Neshat, M., & Gandomi, A. (2025). State-of-the-art AI-based Learning Approaches for Deepfake Generation and Detection, Analyzing Opportunities, Threading through Pros, Cons, and Future Prospects. https://doi.org/10.48550/arxiv.2501.01029
28. Thanga Raj M, Arunachalam M, Ramalakshmi R, Ramaraj K, Arunachalam M, Kaleeswari P. A review on the detection of deep fake and propaganda videos and images-based voice and facial manipulation using AI techniques. In: 2023 2nd International conference on automation, computing and renewable systems (ICACRS), Pudukkottai, India. 2023. p. 554–61. https://doi.org/10.1109/ICACRS58579.2023.10405046.
29. Rathoure, N., Pateriya, R. K., Bharot, N., & Verma, P. (2024). Combating deepfakes: a comprehensive multilayer deepfake video detection framework. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-024-20012-5
30. Kabilan, M. K., Logeswaran, S., Kubersrinivash, S., & Daniel, A. (2025). Advanced Temporal Analysis for Deepfake Detection using XceptionNet in Media Forensics. 1224–1228. https://doi.org/10.1109/iccmc65190.2025.11140668
31. Esty, F. T. T. (2025). Advancing Deepfake Detection: An Optimized Neural Network Model with Keyframe Analysis. 1–6. https://doi.org/10.1109/ncim65934.2025.11160287
32. Wadhwa, I., Choudhary, Y., Chauhan, P., Singh, A., Sathua, S., & Bains, S. S. (2025). DEEPFAKE AND AI-GENERATED IMAGE DETECTION SYSTEM. International Journal for Multidisciplinary Research, 7(2). https://doi.org/10.36948/ijfmr.2025.v07i02.42245
33. Dhol, S., Kanani, N., Koyani, D., Javiya, D., Thakar, H., & Thakkar, A. (2025). A Novel Hybrid Framework for Deepfake Detection. https://doi.org/10.21203/rs.3.rs-6273520/v1
34. Rout, J., & Mishra, M. (2025). Cascaded ResNet50-LSTM Architecture for Deepfake Video Identification. 1–6. https://doi.org/10.1109/icepe65965.2025.11139778
35. Priya, M., Murugesan, J., Bhuvaneswari, P., Rubigha, M., Lalithambikai, S., & Mohanraj, B. (2024). Preserving Visual Authenticity: Block Chain-Augmented AI Frameworks for Advanced Digital Deception Recognition and Mitigation. 707–713. https://doi.org/10.1109/icosec61587.2024.10722740
36. Ravikumar, C. P., Sowjanya, S., Batra, I., & Malik, A. (2025). A Holistic Approach to Detecting and Attributing Deepfake Media Using Advanced Computer Vision and Machine Learning Techniques. Advances in Information Security, Privacy, and Ethics Book Series, 311–328. https://doi.org/10.4018/979-8-3693-6135-1.ch012
37. B, S. A. (2025). A Survey on an Integrated Neural Framework for Real-Time Deepfake Detection across Multimedia to Mitigate Misleading Content. International Journal for Research in Applied Science and Engineering Technology. https://doi.org/10.22214/ijraset.2025.74208
38. Kadam, G., Tiwarekar, S., Sonawane, Y., Devadkar, K., & Sisodia, J. (2025). Hybrid Framework for Interpretable Deepfake Video Detection Using CapsNet and Transformer Encoders. https://doi.org/10.21203/rs.3.rs-7300434/v1
39. Bhonsle, O. D., Gupta, M., Gupta, M., & Waingankar, P. (2025). Detecting Deepfake Media with AI and ML. Deleted Journal, 3(02), 267–274. https://doi.org/10.47392/irjaeh.2025.0037
40. Lad, S. (2024). Applied Ethical and Explainable AI in Adversarial Deepfake Detection: From Theory to Real-World Systems. Deleted Journal, 6(1), 126–137. https://doi.org/10.60087/jaigs.v6i1.236
41. Marrivada, G. (2024). Advanced AI-Based Authentication Framework: Detecting and Mitigating Deepfake Threats in Digital Identity Systems. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 10(6), 1685–1694. https://doi.org/10.32628/cseit241061211
42. Fatima, A., & Ram, P. K. (2024). Gan-Enhanced Real-Time Detection of Deepfakes Videos. Journal of Artificial Intelligence and Copsule Networks, 6(4), 452–465. https://doi.org/10.36548/jaicn.2024.4.005
43. Kalyan, Ch. V., Kaaurthik, M. K. B., & Jagadiswary, D. (2025). Forward Thinking of Detecting Indiscernible Video Counterfeits. https://doi.org/10.21203/rs.3.rs-6929463/v1
44. Lu, Y., & Ebrahimi, T. (2023). Assessment Framework for Deepfake Detection in Real-world Situations. arXiv.Org, abs/2304.06125. https://doi.org/10.48550/arXiv.2304.06125
45. Shreyas, H. S., Choudhari, A., Arvind, T., & Ugarakhod, R. (2025). Hybrid Deep Learning-Based Deepfake Detection Using VGG19 and InceptionV3. 1672–1677. https://doi.org/10.1109/icdici66477.2025.11134949
46. Pandian, P. S., Manikandan, S., Ramya, M., Rajeswari, M., Hemalatha, J., Selvakumar, R. K., & Mothilal, S. (2025). Blockchain-Assisted Video Integrity Verification UsingResNeXt and LSTM-Based Deepfake Detection. International Journal of Basic and Applied Sciences, 14(4), 802–807. https://doi.org/10.14419/v12sar21

47. Ahumada, C. (2025). Progress In Deep Fake and Tampering: An In-Depth Analysis. Indian Scientific Journal of Research in Engineering and Management, 09(01), 1–9. https://doi.org/10.55041/ijsrem40707
48. Zhu, Y., Zhang, C., Gao, J., Sun, X., Rui, Z., & Zhou, X. (2024). High-compressed deepfake video detection with contrastive spatiotemporal distillation. Neurocomputing, 565, 126872. https://doi.org/10.1016/j.neucom.2023.126872
49. Zaman, S. B., Karim, W., Abian, A. I., Mohamed, R. E., Islam, M. R., Karim, A., & Azam, S. (2025). DeepAgent: A Dual Stream Multi Agent Fusion for Robust Multimodal Deepfake Detection (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2512.07351
50. Nguyen-Le, H.-H., Tran, V.-T., Nguyen, D.-T., & Le-Khac, N.-A. (2024). Passive Deepfake Detection Across Multi-modalities: A Comprehensive Survey (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2411.17911
51. I. Amerini, L. Galteri, R. Caldelli and A. Del Bimbo, "Deepfake Video Detection through Optical Flow Based CNN," 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South), 2019, pp. 1205-1207, doi: 10.1109/ICCVW.2019.00152.
52. A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Niessner, "FaceForensics++: Learning to Detect Manipulated Facial Images," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 1-11, doi: 10.1109/ICCV.2019.00009.
53. Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Capsule-forensics: Using capsule networks to detect forged images and videos. ICASSP 2019 – IEEE International Conference on Acoustics, Speech and Signal Processing.
54. Jain, N., Borade, S., Patel, B., Kumar, V., Godhrawala, M., Kolaskar, S., Nagare, Y., Shah, P., & Shah, J. (2023). Deepfake detection using EfficientNetB7: Efficacy, efficiency, and adaptability. International Journal of Intelligent Systems and Applications in Engineering
55. Pandian, P.., Manikandan, S., Ramya, M., Rajeswari, M., Hemalatha, J., Selvakumar, R.., & Mothilal, S. (2025). Blockchain-Assisted Video Integrity Verification UsingResNeXt and LSTM-Based Deepfake Detection. International Journal of Basic and Applied Sciences, 14(4), 802-807. https://doi.org/10.14419/v12sar21
56. R. Z. Khudhur and M. Mohammed, “A Spatio-Temporal Deep Learning Approach for Efficient Deepfake Video Detection,” ARO. The Scientific Journal of Koya University, Aug. 2025, doi: 10.14500/aro.12190
57. R. Z. Khudhur and M. Mohammed, “A Spatio-Temporal Deep Learning Approach for Efficient Deepfake Video Detection,” ARO. The Scientific Journal of Koya University, Aug. 2025, doi: 10.14500/aro.12190
58. M. Akbari et al., “Design and development of an efficient RLNet prediction model for deepfake video detection,” Frontiers in Big Data, Jul. 2025, doi: 10.3389/fdata.2025.1569147.