Krishna J Patel[1],

Madhavi B Desai

Ai-Driven Advances and Challenges in Deepfake Technology: A Comprehensive Review

JES Journal of Electrical Systems

Abstract: **-** Deepfake technology has completely changed the production of synthetic media by allowing the alteration of photos, movies, and audio recordings. It is driven by machine learning and artificial intelligence algorithms. Deepfake provide a lot of amusement and artistic freedom, but they also pose serious problems, especially when it comes to disinformation and digital manipulation. This paper offers a thorough introduction to deepfake technology, covering all of its types, including voice synthesis, gesture control, and face-swapping. The research article delves into the fundamental workings of deepfake generation, emphasizing the part played by convolutional neural networks and generative adversarial networks in producing lifelike artificial content. The paper also investigates the methods and strategies used in detection, with a focus on the latest developments in deep neural network architectures, attention-based models, and hybrid approaches. This review article also focuses on availability of standard datasets and performance parameters for the evaluation of research models. With the aim to provide contribution to help researchers to create reliable and efficient deepfake detection systems that can stop the distribution of manipulated media and ensure the accuracy of digital content by tackling various issues, paper also focuses on key challenges and future work.

Keywords: Deepfake types; GAN; CNN; Deepfake detection; Deepfake generation.

1. INTRODUCTION

Deepfake are fake media which are created by replacing someone else's image in an already-existing photo or video. "Deepfake," a phrase that combines the words "deep learning" and "fake," is a method of changing the source character in a film. Deepfake has been used to produce fake news and other types of deception, even though technology can be used for valid reasons like producing lifelike visual effects in films and television series. Owing to its negative effects, digital face alteration has grown to be a serious problem in both social media and in our culture(Mohiuddin et al., 2023). In November 2023, a Deepfake video of Rashmika Mandanna, a well-known Bollywood actress, went viral on social media. The Bollywood actress's face was substituted for the face of a British-Indian influencer woman wearing a black exercise garment in the video (India Today. (2024, n.d.). It has been demonstrated that various machine learning and deep learning methods have been shown to be effective in detecting Deepfake. The model can be utilized to forecast the probability that a fresh image or video is a Deepfake once it has been trained.A variety of face-alteration techniques have recently been used by face-swap abusers to produce fake photos and videos. These images and videos are then exploited to make false news, support vile hoaxes, and fabricate legal evidence, all of which have had detrimental effects(Dong et al., 2023). The advancement of deep learning has sped up the creation of face forgery technology(Huang et al., 2023). Deepfake techniques have been largely successful due to recent advancements in generative models (Waseem et al., 2023). More significantly, Deepfake films are hard to spot with the same methods that are used to spot traditional forgeries (such copy-move and splicing). Thus, it's imperative to create effective methods for exposing DeepFake movies(Y. Yu et al., 2023). AI-powered software programs such as FaceApp and FakeApp have enabled realistic face alteration in images and movies. This alteration method allows people to alter their age, gender, haircut, front

[1]Research Scholar, Gujarat Technological University, Ahmedabad, Gujarat, India
Email id: kishu5372@gmail.com
2Associate Professor, Computer Science & Engineering Department, R N G Patel Institute of Technology, Bardoli, Gujarat, India
*Corresponding author Email id: desaimadhavi30@gmail.com

look, and other personal traits(Rana et al., 2022). Using a range of training techniques, convolutional neural network (CNN) models have become state-of-the-art (SOTA) techniques. They are employed in the fields of object recognition, picture generation, and image classification(Heo et al., 2023). Facial modifications are classified into many types of manipulations based on the degree of transformation, such as identity switching, face synthesis, attribute manipulation, and full expression shifting (Abdulreda & Obaid, 2022).

Figure 1 Original and Deepfake image

The Deepfake image of well-known Hollywood star Tom Cruise is displayed in Figure 1. We can observe from the graphic that it is difficult to recognize Deepfake modified photographs. As seen in Figure 2, scientists have developed Mona Lisa Deepfake, which makes the image come to life and produces a film in which the image speaks and begins to express itself.

(News, n.d.).

Figure 2 Mona Lisa Deepfake video by Samsung Lab

1.1 Research Question and Contribution

The following research questions serve as a guide and source of information for this survey, which aids researchers in their investigation of Deepfake through the use of different deep learning and machine learning techniques. **RQ1:** What kinds of Deepfake are there? Understanding the different kinds of Deepfake is necessary for the generation and detection process. This question contributes by outlining the many kinds of Deepfake. **RQ2:** What are the general method for creating Deepfake in images and videos? In order to answer this question, this research explores different techniques of deep learning methods such as Generative Adversarial Networks (GANs) and auto encoders which are widely used techniques for Deepfake generation.

RQ3: What is the typical procedure for identifying Deepfake? We provide a quick explanation of several Deepfake detection techniques and strategies in order to detect the deepfake. **RQ4:** Which standard datasets are used for model training? For dependable model training, standard datasets such as FaceForensics++, DeepFake Detection Challenge (DFDC), Celeb-DF, and UADFV are more useful dataset.

RQ5: What are the different assessment parameters that are applied to gauge the Deepfake detection model's accuracy? metrics such as the Receiver Operating Characteristic (ROC) curve, F1-score, accuracy, precision, recall, and Area under the Curve (AUC). **RQ6:** What are the difficulties and potential applications of Deepfake? Future study may need to focus on a number of issues in order to minimize potential harm; therefore, this paper discusses these issues in an effort to assist the researchers.

Using the aforementioned research questions as a guide, we first created an article search strategy. We were able to access the most pertinent material by using this strategy to filter out key concepts like "Deepfake detection," "different methods for Deepfake detection," "Generating Deepfake," "Types of Deepfake," "Machine Learning," "Deep Learning," and "Evaluation Metrics." These search terms were used to look through well-known databases, such as IEEE Explorer, Scopus, SCI, Web of Science, and Computer Vision. 58 references were ultimately determined to be worthy of a thorough examination and analysis to justify the research questions and provide contribution by doing so. (RQ1 through RQ6)

1.2 Organization of Paper

The structure of this review paper is as follows: In second section various Deepfake types are discussed. Section three discusses deepfake generation. The literature review on Deepfake generation is also covered in this part. The detection of Deepfake detection literature review is covered in section four of this study. In section five, standard Deepfake datasets are mentioned, which researchers can use to assess algorithms. The literature for deepfake detection uses a variety of evaluation factors in Section six. The challenges and upcoming efforts to detect Deepfake are discussed in Section seven for the researcher's future guidance.

2. DEEPFAKE TYPES

Deepfake is the term for artificial intelligence (AI) and machine learning algorithms that produce synthetic media, including audio recordings, videos, and images. Deepfake comes in several varieties depending on the manufacturing methods and modified material(Seow et al., 2022).

2.1 Entire Face Synthesis:

This method creates an entirely new set of facial photos that are unrelated to any known individual.

Figure 3 Entire Face Synthesis

Figure 3 illustrates how we can design an entirely new face by providing input for facial features such as a person's eyes, nose, and lips.

2.2 Identity Swap:

As the name indicates, one's identity changing is the practice of substituting the face of one person for another in a film.

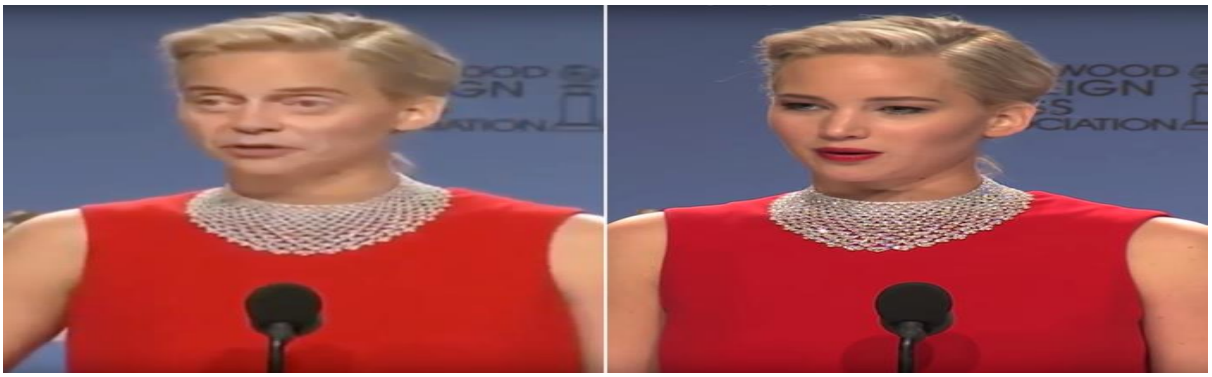

Figure 4 Identity Swap

Figure 4 illustrates how identity swapping was used to establish the well-known American actress Jennifer Lawrence's persona.

2.3 Characteristics Alteration:

Characteristics Alteration aims to change specific face features of characters in images or videos.

Figure 5 Attribute Manipulation

Figure 5 illustrates three distinct adjustments. Glasses are put on and taken off of two distinct people in section (a). Similar to component (b), opening and shutting the mouth is simple when attribute manipulation techniques are used. This method makes it simple to add and remove a man's beard in part (c).

2.4 Expression Swap

This technique makes it possible to modify facial expressions, gestures, and emotions to produce realistic expressions that weren't captured in the original film (Afchar et al., 2018). Figure 6 shows the Deepfake technology being used to generate an image (left) of former US Secretary of State Hillary Clinton (right).

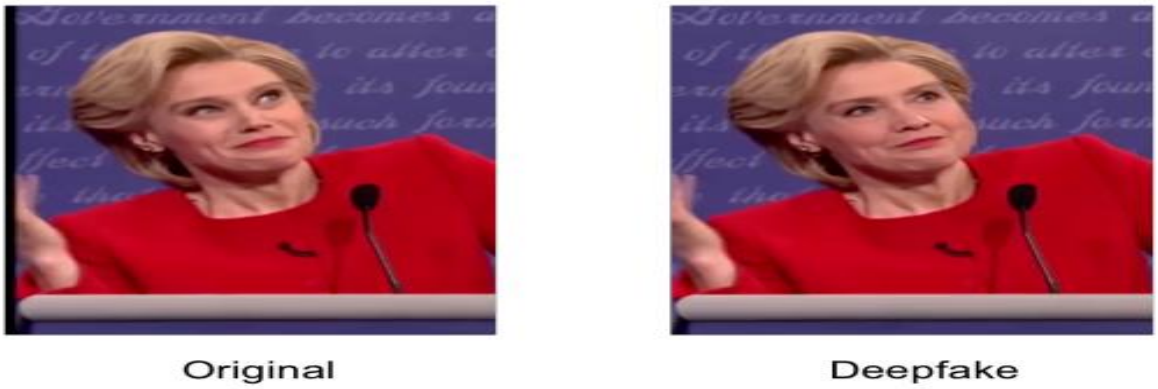

Figure 6 Expression Swap

2.5 Voice Synthesis

Voice synthesis deepfake include producing artificial voice recordings that mimic certain people's speech patterns, intonations, and tones.

2.6 Hybrid Deepfake

Hybrid Deepfake combines multiple Deepfake techniques to produce more complex and realistic synthetic media.

3. DEEPFAKE GENERATION

The term "Deepfake generation" describes the process of producing synthetic media, usually by means of deep learning algorithms, in order to alter or substitute pre-existing information, most frequently voices and faces in films.

3.1 LITERATURE REVIEW OF DEEPFAKE GENERATION

Deep learning models like auto encoders and generative adversarial network, which have been most current prevalent in the computer vision field for producing deep fakes (Nguyen et al., 2022). Implicit generative models such as GAN have made impressive strides in the creation of images. However, because there is an extra temporal dimension, creating videos is more difficult than creating images. In addition to maintaining a constant geometric structure and the appearance of diverse objects throughout the film at all frames, the created videos should also include physically plausible object motion(Kasaraneni, 2022). Two neural networks make up a GAN: a discriminator and a generator. The discriminator attempts to differentiate between real and fake images or videos, which are produced by the generator. Together, the two networks undergo training, which helps the generator become increasingly proficient at producing realistic-looking Deepfake over time(Rekha G, 2023)(Rana et al., 2022). GANs are made up of two parts: a decoder that uses a SoftMax function to categorize image as real or false, where as encoder that creates fake data. Several programs, including VGGFace and FakeApp, use autoencoder structures to produce incredibly lifelike deepfakes, making detection more difficult(Almars, 2021).GAN Based technique AttGAN presents a novel method of modifying facial attributes by combining GAN with encoder-decoder architecture. It guarantees accurate attribute alterations while maintaining other information by directly applying attribute classification limitations to generated images(He et al., 2019). Similar to A GAN architecture called StyleGAN is excellent at producing realistic and detailed visuals. It's a helpful tool for creating high-quality deepfakes because it can manipulate different facial features of the picture(Karras et al., 2021). RSGAN is capable of encoding facial and hair appearances into underlying latent space representations, which allows for the manipulation of the representations in the latent spaces to alter the appearances of the images(Natsume et al., 2018). Another GAN variation that focuses on style transfer between two domains is called CycleGAN. Face-swapping applications can benefit from its ability to transfer facial features from one person to another (Jay et al., 2017). In a similar vein, STGAN alters particular facial features in photos while maintaining the subject's identity. The model has demonstrated promising results in precisely regulating attribute changes and is able to operate with both labeled and unlabeled data(Liu et al., 2019). The goal of LipGAN is to generate realistic lip movements and speech synchronization for a given audio input. These methods highlight how Deepfake technology is developing overall and highlight the necessity for more study to overcome ethical and

technological obstacles(Philip, 2019).

4. DEEPFAKE DETECTION

Deepfake detection is the act of identifying and reporting digitally manipulated content, particularly images or videos created using deep learning techniques.

4.1 LITERATURE REVIEW OF DEEPFAKE DETECTION

Several techniques have been put up to identify edited videos. We classify various studies based on the methods used, and the resulting sections provide an overview of each study.

4.1.1 MACHINE LEARNING BASED METHODS

To detect modifications, the study gives visual cues that highlight the shape of the face, teeth, and eyes. They created a technique that contrasted head placements determined by utilizing all available visual cues with those determined by utilizing just the central area(Matern et al., 2019). Similar features can also be seen in Deepfake films, which can be found by estimating the 3D head position. The SVM classifier is used to calculate inconsistent head positions (Matern et al., 2019). The another research includes identifying authentic from phony video sequences by examining spatiotemporal texture dynamics. This suggest capturing textural dynamics in both spatial and temporal dimensions by employing Local Derivative Patterns on three Orthogonal Planes(Bonomi et al., 2021a). Authors in paper (Mitra et al., 2021) introduce an innovative machine learning-based approach tailored for the detection of Deepfake videos within the constraints of compressed social media platforms. They used CNN with Xception network method uniquely employs key video frame extraction, effectively reducing computational demands while maintaining a high level of accuracy. These models have trouble with high-dimensional data and may need a lot of feature engineering, which can cause problems with scalability and poor generalization performance.

4.1.2 DEEP LEARNING BASED METHODS

Deep learning has become the foundation of deepfake detection, with convolutional neural networks (CNNs) in particular playing a key role. XceptionNet and MobileNet are two methods that use the subtle analysis of visual abnormalities to achieve detection accuracy of up to 98% on benchmark datasets(Pan et al., 2020). In Table 2 methods are shortly explained which are based on biological signals, camera fingerprints, visual artifacts, temporal consistency, and general networks.(P. Yu et al., 2021)

Table 1 Existing Deepfake detection methods

Methods	***Description***
General network-based methods	This approach views detection as a frame-level classification task which is done by CNNs.
Temporal consistency-based methods	Deepfake films are discovered to have inconsistent frames between neighboring ones for this RNN is utilized.
Visual artifacts-based methods	Intrinsic image differences in the blending boundaries would result from the blending procedure during the generation phase. These artifacts are identified using CNN-based approaches.
Camera fingerprints-based methods	Devices leave various imprints in the photos they take because of unique generation processes. Faces and background images are recognized to originate from distinct devices concurrently. Thus, by utilizing these traces, the detecting work can be finished.
Biological signals-based methods	A biological signal is extracted to identify Deepfake videos based on this observation.

Table 2 shows the different existing Deepfake detection methods which is achieved by various architecture of deep learning algorithms.

Convolutional Neural Networks and autoencoders have been at the forefront of DeepFake detection methodologies. This attribute not only aids in the DeepFake generation process but also enhances the precision of detection systems. As such, AEs have become a critical element, mirroring their prevalent use in the fabrication of Deepfake and their widespread dissemination across the internet(Siegel et al., 2021).

4.1.2.1 General Network-based methods

Generally, network-based methods treat detection as a frame-level classification issue and handle it using convolutional neural networks (CNNs). To address premature convergence problems, for example, methods like Grey Wolf Optimization and Vortex Search have been used, which has feature selection(Mohiuddin et al., 2023). The exceptional performance of cascaded network topologies such as EfficientNetV2S and Vision Transformer (ViT) is demonstrated by additional research utilizing the FaceForensics++ and DFDC datasets (Thaseen Ikram et al., 2023). A cascaded network architecture combining Vision Transformer (ViT) and EfficientNetV2S is useful for other binary classification tasks in addition to Deepfake detection. Their approach shows excellent performance on DFDC and FaceForensics++(Deng et al., 2023).The paper (Wang et al., 2023) present a novel deep convolutional Transformer using convolutional pooling and re-attention techniques for facial feature learning both locally and globally in Deepfake detection,by highlighting the evaluation of the proposed approach on the FF++ dataset. A different promising strategy (Deng et al., 2022) used the EfficientNet-V2 network as the foundation for the researchers' Deepfake video identification system. The FaceForensics++ and FFIW10K datasets demonstrated this method's impressive accuracy. In paper (Kawa & Syga, 2020) author propose a set of handcrafted features, which don't rely on extensive computational power or data, and then employ a simple classifier to detect Deepfake

4.1.2.2 Temporal consistency-based methods

The study (Singh et al., 2020) proposed method that represents a significant advancement by harnessing spatio-temporal features, the authors employ a Time-Distributed CNN with LSTM for feature extraction, achieving an impressive accuracy rate on the challenging Deep Fake Detection Challenge dataset (DFDC).Using a fully temporal convolution network and a Transformer Network that investigates the long-term temporal coherence, the paper (Zheng et al., 2021) suggested a method to reduce the spatial convolution kernel size to 1 while maintaining the temporal convolution kernel size constant. This allows for the exploitation of temporal coherence in the detection of Deepfake. The paper (Gu et al., 2022) proposed a Region-Aware Temporal Filter module that divides the dynamic temporal kernel into basic, region-independent filters in order to provide temporal filters to distinct geographical locations. To help these regions adaptively learn about temporal incongruities, region-specific aggregation weights are also included. A number of short clips are taken from the input video in order to cover the long-term temporal dynamics. In paper (Bonomi et al., 2021b) introduce a dynamic texture analysis approach to identify fake faces within video sequences. This method leverages the temporal dynamics of facial features and expressions to distinguish genuine faces from Deepfake or manipulated ones.The study in (Xu et al., 2023) introduces Thumbnail Layout (TALL), a novel approach for Deepfake video detection that emphasizes spatial and temporal dependencies by reorganizing video clips into a predefined layout. Optimization of TALL-Swin is achieved through the utilization of cross-entropy loss, further enhancing its effectiveness in Deepfake detection task.

4.1.2.3 Visual artifacts-based methods

A new deep learning-based strategy was introduced (Y. Li & Lyu, 2018) based on the findings of discrepancy between faces and background. The mixing procedure produced face warping artifacts, which were then utilized to identify fake pictures. The methods in paper (Xia et al., 2022) focuses on enhancing texture differences in images through a Preprocessing module, which effectively filters out low-frequency information while preserving high-frequency texture details. Feature extraction is then performed using the MesoNet architecture. Additional advancements were made in paper (L. Li et al., 2020), which suggested a unique picture representation called face X-ray. This representation was used to determine if the input image could be divided into the background and the foreground face. In particular, face X-ray was defined as the blending boundary between the altered foreground face and the backdrop. In contrast to publication (Y. Li & Lyu, 2018), it introduced in picture blending and shown excellent performance across a range of datasets. With the exception of suggesting face X-rays, this study specifically uses positive samples to create the procedure of generating negative samples. However, this approach

is not resistant to fully synthesized images because of its overemphasis on the blending boundary.

4.1.2.4 Camera fingerprints-based methods

Particularly in source identification tasks, camera fingerprint a type of very faint energy noise play a significant role in forensic fields. Three processes have generally been involved in camera fingerprint-based methods: noise print, recent video noise pattern, and photo response non uniformity (PRNU) patterns. The varying light sensitivity of the pixels as a result of the silicon wafers non uniformity and imperfections during the sensor's manufacturing process give rise to PRNU. The PRNU pattern is thought of as a device fingerprint because of its stability and uniqueness, which makes it useful for a variety of forensic operations(Korus & Huang, 2017). The author originally suggested using PRNU to detect Deepfake films based on these findings(Koopman et al., 2018).

4.1.2.5 Biological signals-based methods

The method used in (Y. Li et al., 2018) was based on recognizing eye blinking, a biological signal that is difficult to interpret in Deepfake footage. Consequently, the lack of eye blinking indicates the presence of a Deepfake video. It need consideration of prior temporal knowledge to identify open and closed eye states. As an alternative paper (Hernandez-Ortega et al., 2021) offer a novel method for identifying Deepfake films that centers on using remote photo plethysmography (rPPG) to analyze heart rate data. It is possible to determine whether human blood is present beneath the tissues by watching video sequences and seeing minute changes in skin tone. In order to extract spatial and temporal characteristics from video frames and efficiently combine the two origins for improved fake video detection, the suggested detection system, named Deepfake ONPhys, takes into consideration prior temporal knowledge. A newly developed method called DeepRhythm (Qi et al., 2020) monitored the cardiac rhythms using a dual-spatial-temporal attention mechanism and showed good generalization across various datasets. In (Fernandes et al., 2019) author introduces an intriguing approach for Deepfake detection. The authors propose using Neural Ordinary Differential Equations (ODE) to predict heart rate variations in videos, exploiting the fact that Deepfake often lack subtle physiological cues.This novel technique demonstrates promise in detecting Deepfake, particularly when coupled with other forensic methods. To detect Deepfake ,the researcher in (Heo et al., 2023) presents a unique method for DeepFake detection that combines CNN features and patch embedding with a DeiT-based distillation token. The suggested vision transformer model outperforms the state-of-the-art EfficientNet, beating the SOTA's AUC of 0.972 with an AUC of 0.978 without the need for an ensemble method. The paper (Waseem et al., 2023) present technique to identify and locate Deepfake, the researcher presents a novel attention-based multi-tasking technique.. Their method utilizes a combination of spatial attention (SAM) and channel attention (CAM) techniques, incorporating a Residual U-Net with a spatial channel attention block for Facial Manipulation Localization (FML) and employing Fourier Transform (FFT) for Facial Manipulation Detection (FMD). A new technique in paper (Khormali & Yuan, 2021) authors introduced a novel DeepFake detection approach known as ADD (Attention-Based DeepFake Detection). This method leverages attention mechanisms to effectively target manipulated regions and enhancing detection accuracy. The paper (Preeti et al., 2022) focus is on the development of a Deepfake detection model utilizing generative adversarial networks (GANs). The research delves into the utilization of pre-trained GANs and deep convolution-based GAN models for creating Deepfake content, while also examining the methods employed for implementing Deepfake, manipulation, and detection techniques. In paper (Maras & Alexandrou, 2019) author address the critical issue of verifying the credibility of video evidence in a rapidly evolving digital landscape. They explore the implications of artificial intelligence in both the creation and detection of Deepfake videos. The study in paper (Trinh & Liu, 2021) examines the fairness of AI models in Deepfake detection and finds notable differences in performance between demographic groups. Of particular note are race-based mistake rates, which among three widely used detectors can reach 10.7%. It draws attention to the disproportionate number of Caucasian subjects especially female Caucasians in datasets such as FaceForensics++.There is also another way to conceptualize the Deepfake detection method is as a binary classification process with labels for either original class or Deepfake. In order to distinguish between original and Deepfake content, Deepfake detection extracts features from the image or video(Ramadhani & Munir, 2020).Many deepfake detection techniques suffer from poor generalization when tested on datasets other than the training set. By dissecting deepfake-specific information from unimportant aspects developed a deep information decomposition (DID) approach that improves robustness in cross-dataset circumstances. (Yang et al., 2023).

5. DEEPFAKE DATASET

Prominent datasets such as DeepFake Detection Challenge (DFDC), FaceForensics++, and Celeb-DF have played a significant role in advancing research in this area(Thaseen Ikram et al., 2023). Here in table 3, we describe commonly used dataset for Deepfake detection.

Table 2 Deepfake Dataset

Dataset	*Original Videos*	*Deepfake Videos*	*Manipulation Methods*	*Total Videos*
FaceForensics++ (FaceForensics, n.d.)	977	1000	Deepfakes, Face2Face, FaceSwap, Neural Textures	1977
Celeb-DF(Github, n.d.)	590	5639	-	6229
UADFV(Kaggle, n.d.)	49	49	DNN	98
DF-TIMIT(Idiap, n.d.)	-	-	DF-TIMIT-LQ, DF-TIMIT-HQ	640

6. EVALUATION PARAMETERS

The evaluation of Deepfake detection models requires the use of robust performance metrics such as accuracy, precision, recall, F1-score, and Receiver Operating Characteristic (ROC) curves(Seow et al., 2022).Table 4 shows different Evaluation parameters with its formula.

Table 3 Evaluation Parameters

Evaluation Metric	*Formula*
Accuracy (ACC)	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1-Score	$\frac{2(Precision.Recall)}{Precision+Recall}$
AUC (Area Under the Curve)	$\int_0^1 TPR\ (FPR)d(FPR)$
False Positive Rate (FPR)	$\frac{FP}{FP+FN}$
False Negative Rate (FNR)	$\frac{FN}{FN+TP}$
ROC Curve	Plot of TPR against FPR at various threshold settings

7. CHALLENGES AND FUTURE WORK

The study draws attention to the difficulties and new issues that have been found when applying computational methods to detect Deepfake in images and videos.

i. It is essential to create detection algorithms that can generalize to a variety of Deepfake creation techniques. This guarantees that the algorithms continue to function even when new methods are developed.

ii. It is important to develop algorithms that are capable of real-time operation, enabling prompt identification of Deepfake in live broadcasts or during their creation and dissemination.

iii. Increasing the resilience of the model is crucial to combating various DeepFake generation techniques.

iv. Deepfake detection systems can be implemented on a larger range of devices, including those with constrained processing capabilities, by developing algorithms that use less run-time memory.

v. The model's effectiveness can be further enhanced by include various attention mechanism, as this allows it to capture a wider range of irregularities that could be signs of a Deepfake.

vi. It is critical to develop models that function effectively even when trained on small datasets, which is frequently the case because labeled Deepfake data is so hard to come by.

Addressing these challenges is key to building reliable and efficient systems that can keep pace with the rapidly advancing domain of digital content manipulation.

8. CONCLUSION

Deepfake technology poses a significant danger to humanity as a notable development in artificial intelligence. Misuse of technology can have negative effects, including dissemination of misleading information and a drop in public trust in the media, even though it presents new opportunities for amusement and artistic expression. The various forms of Deepfake, such as voice synthesis, identity swapping, emotion swapping, changing attributes, and full face synthesis, have been clarified by this survey. The survey also explores how Deepfake is created, mostly using various deep learning algorithms like auto encoders and GANs. These methods make it possible to mimic look and behavior, leading to the production of artificially created media that appears realistic but is fake. Although it can be difficult to detect Deepfake, development in deep learning and machine learning methods made it possible to develop techniques that can spot irregularities and inconsistencies in modified content. Standard datasets for detection model training, such as DFDC and FaceForensics++ essential. Moreover, measures like accuracy, precision, ROC, AUC etc. are used to evalaute how accurate Deepfake detection models are. These metrics evaluate the model's ability to differentiate between original and fake media. Future research into increasingly complex detection approaches, diversified datasets, real-time processing models, and algorithms that can generalize across various Deepfake production strategies will be necessary to meet the problems presented by Deepfake. By addressing these issues, researchers can lessen the possible harm such as deception, fraud, and invasions of privacy caused by the widespread usage of Deepfake.

CONFLICT OF INTEREST: There is no conflict of interest.

REFERENCES

1. Abdulreda, A. S., & Obaid, A. J. (2022). A landscape view of deepfake techniques and detection methods. *International Journal of Nonlinear Analysis and Applications*, *13*(1), 745–755. https://doi.org/10.22075/IJNAA.2022.5580

2. Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: A compact facial video forgery detection network. *10th IEEE International Workshop on Information Forensics and Security, WIFS 2018*, 1–7. https://doi.org/10.1109/WIFS.2018.8630761

3. Almars, A. M. (2021). Deepfakes Detection Techniques Using Deep Learning: A Survey. *Journal of Computer and Communications*, *09*(05), 20–35. https://doi.org/10.4236/jcc.2021.95003

4. Bonomi, M., Pasquini, C., & Boato, G. (2021a). Dynamic texture analysis for detecting fake faces in video sequences. *Journal of Visual Communication and Image Representation*, *79*(July), 103239. https://doi.org/10.1016/j.jvcir.2021.103239

5. Bonomi, M., Pasquini, C., & Boato, G. (2021b). Dynamic texture analysis for detecting fake faces in video sequences. *Journal of Visual Communication and Image Representation*, *79*, 1–11. https://doi.org/10.1016/j.jvcir.2021.103239

6. Deng, L., Suo, H., & Li, D. (2022). Deepfake Video Detection Based on EfficientNet-V2 Network. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/3441549

7. Deng, L., Wang, J., & Liu, Z. (2023). Cascaded Network Based on EfficientNet and Transformer for Deepfake Video Detection. *Neural Processing Letters*. https://doi.org/10.1007/s11063-023-11249-6

8. Dong, S., Wang, J., Ji, R., Liang, J., Fan, H., & Ge, Z. (2023). *Implicit Identity Leakage: The Stumbling Block to Improving Deepfake Detection Generalization*. *1*, 3994–4004. https://doi.org/10.1109/cvpr52729.2023.00389

9. FaceForensics. (n.d.). *https://kaldir.vc.in.tum.de/faceforensics_benchmark/*. https://kaldir.vc.in.tum.de/faceforensics_benchmark/

10. Fernandes, S., Raj, S., Ortiz, E., Vintila, I., Salter, M., Urosevic, G., & Jha, S. (2019). Predicting heart rate variations of deepfake videos using neural ODE. *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*, 1721–1729. https://doi.org/10.1109/ICCVW.2019.00213

11. Github. (n.d.). *Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics*. https://github.com/yuezunli/celeb-deepfakeforensics

12. Gu, Z., Yao, T., Chen, Y., Yi, R., Ding, S., & Ma, L. (2022). Region-Aware Temporal Inconsistency Learning for DeepFake Video Detection. *IJCAI International Joint Conference on Artificial Intelligence*, 920–926. https://doi.org/10.24963/ijcai.2022/129

13. He, Z., Zuo, W., Kan, M., Shan, S., & Chen, X. (2019). AttGAN: Facial Attribute Editing by only Changing What You Want. *IEEE Transactions on Image Processing*, *28*(11), 5464–5478. https://doi.org/10.1109/TIP.2019.2916751

14. Heo, Y. J., Yeo, W. H., & Kim, B. G. (2023). DeepFake detection algorithm based on improved vision transformer. *Applied Intelligence*, *53*(7), 7512–7527. https://doi.org/10.1007/s10489-022-03867-9

15. Hernandez-Ortega, J., Tolosana, R., Fierrez, J., & Morales, A. (2021). DeepFakesON-Phys: Deepfakes detection based on heart rate estimation. *CEUR Workshop Proceedings*, *2808*.

16. Huang, B., Wang, Z., Yang, J., Ai, J., Zou, Q., Wang, Q., & Ye, D. (2023). *Implicit Identity Driven Deepfake Face Swapping Detection*. 4490–4499. https://doi.org/10.1109/cvpr52729.2023.00436

17. Idiap. (n.d.). *Deepfaketimit*. https://www.idiap.ch/en/scientific-research/data/deepfaketimit

18. India Today. (2024, J. 20). (n.d.). *Rashmika Mandanna deepfake video accused arrested: Andhra engineer wanted to boost followers. India Today.* https://www.indiatoday.in/india/story/rashmika-mandanna-deepfake-video-accused-arrested-andra-engineer-wanted-to-boost-followers-2491386-2024-01-20

19. Jay, F., Renou, J.-P., Voinnet, O., & Navarro, L. (2017). Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks Jun-Yan. *Proceedings of the IEEE International Conference on Computer Vision*, 183–202. http://link.springer.com/10.1007/978-1-60327-005-2_13

20. Kaggle. (n.d.). *UADFV-dataset*. https://www.kaggle.com/datasets/ahmadawad732/uadfv-dataset-new

21. Karras, T., Laine, S., & Aila, T. (2021). A Style-Based Generator Architecture for Generative Adversarial Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(12), 4217–4228. https://doi.org/10.1109/TPAMI.2020.2970919

22. Kasaraneni, S. H. (2022). *Autoencoding Video Latents for Adversarial Video Generation*. http://arxiv.org/abs/2201.06888

23. Kawa, P., & Syga, P. (2020). *A Note on Deepfake Detection with Low-Resources*. http://arxiv.org/abs/2006.05183

24. Khormali, A., & Yuan, J. S. (2021). Add: Attention-based deepfake detection approach. *Big Data and Cognitive Computing*, *5*(4). https://doi.org/10.3390/bdcc5040049

25. Koopman, M., Rodriguez, A. M., Macarulla Rodriguez, A., & Geradts, Z. (2018). *Detection of Deepfake Video Manipulation Encyclopedia View project Detection of Deepfake Video Manipulation*. *August*, 27–31. https://www.researchgate.net/publication/329814168

26. Korus, P., & Huang, J. (2017). Multi-Scale Analysis Strategies in PRNU-Based Tampering Localization. *IEEE Transactions on Information Forensics and Security*, *12*(4), 809–824. https://doi.org/10.1109/TIFS.2016.2636089

27. Li, L., Bao, J., Zhang, T., Yang, H., Chen, D., Wen, F., & Guo, B. (2020). Face X-ray for more general face forgery detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 5000–5009. https://doi.org/10.1109/CVPR42600.2020.00505

28. Li, Y., Chang, M. C., & Lyu, S. (2018). In Ictu Oculi: Exposing AI created fake videos by detecting eye blinking. *10th IEEE International Workshop on Information Forensics and Security, WIFS 2018*, 1–7. https://doi.org/10.1109/WIFS.2018.8630787

29. Li, Y., & Lyu, S. (2018). *Exposing DeepFake Videos By Detecting Face Warping Artifacts*. http://arxiv.org/abs/1811.00656

30. Liu, M., Ding, Y., Xia, M., Liu, X., Ding, E., Zuo, W., & Wen, S. (2019). STGAN: A unified selective transfer network for arbitrary image attribute editing. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 3668–3677. https://doi.org/10.1109/CVPR.2019.00379

31. Maras, M. H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *International Journal of Evidence and Proof*, *23*(3), 255–262. https://doi.org/10.1177/1365712718807226

32. Matern, F., Riess, C., & Stamminger, M. (2019). Exploiting visual artifacts to expose deepfakes and face manipulations. *Proceedings - 2019 IEEE Winter Conference on Applications of Computer Vision Workshops, WACVW 2019*, 83–92. https://doi.org/10.1109/WACVW.2019.00020

33. Mitra, A., Mohanty, S. P., Corcoran, P., & Kougianos, E. (2021). A Machine Learning Based Approach for Deepfake Detection in Social Media Through Key Video Frame Extraction. *SN Computer Science*, *2*(2), 1–18. https://doi.org/10.1007/s42979-021-00495-x

34. Mohiuddin, S., Sheikh, K. H., Malakar, S., Velásquez, J. D., & Sarkar, R. (2023). A hierarchical feature selection strategy for deepfake video detection. *Neural Computing and Applications*, *35*(13), 9363–9380. https://doi.org/10.1007/s00521-023-08201-z

35. Natsume, R., Yatagawa, T., & Morishima, S. (2018). RSGAN: Face swapping and editing using face and hair representation in latent spaces. *ACM SIGGRAPH 2018 Posters, SIGGRAPH 2018*. https://doi.org/10.1145/3230744.3230818

36. News, B. (n.d.). *Mona Lisa "brought to life" with deepfake AI*. https://www.bbc.com/news/technology-48395521

37. Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., Pham, Q. V., & Nguyen, C. M. (2022). Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding*, *223*. https://doi.org/10.1016/j.cviu.2022.103525

38. Pan, D., Sun, L., Wang, R., Zhang, X., & Sinnott, R. O. (2020). Deepfake Detection through Deep Learning. *Proceedings - 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, BDCAT 2020*, 134–143. https://doi.org/10.1109/BDCAT50828.2020.00001

39. Passos, L. A., Jodas, D., da Costa, K. A. P., Júnior, L. A. S., Rodrigues, D., Del Ser, J., Camacho, D., & Papa, J. P. (2022). *A Review of Deep Learning-based Approaches for Deepfake Content Detection*. http://arxiv.org/abs/2202.06095

40. Philip, J. (2019). *Towards Automatic Face-to-Face Translation*. 1428–1436.

41. Preeti, Kumar, M., & Sharma, H. K. (2022). A GAN-Based Model of Deepfake Detection in Social Media. *Procedia Computer Science*, *218*, 2153–2162. https://doi.org/10.1016/j.procs.2023.01.191

42. Qi, H., Guo, Q., Juefei-Xu, F., Xie, X., Ma, L., Feng, W., Liu, Y., & Zhao, J. (2020). DeepRhythm: Exposing DeepFakes with Attentional Visual Heartbeat Rhythms. *MM 2020 - Proceedings of the 28th ACM International Conference on Multimedia*, 1318–1327. https://doi.org/10.1145/3394171.3413707

43. Ramadhani, K. N., & Munir, R. (2020). A Comparative Study of Deepfake Video Detection Method. *2020 3rd International Conference on Information and Communications Technology, ICOIACT 2020*, 394–399. https://doi.org/10.1109/ICOIACT50329.2020.9331963

44. Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake Detection: A Systematic Literature Review. *IEEE Access*, *10*, 25494–25513. https://doi.org/10.1109/ACCESS.2022.3154404

45. Rekha G, P. S. (2023). Deepfake: Creation and Detection using Deep Learning. *International Journal for Research in Applied Science and Engineering Technology*, *11*(5), 4513–4518. https://doi.org/10.22214/ijraset.2023.52674

46. Seow, J. W., Lim, M. K., Phan, R. C. W., & Liu, J. K. (2022). A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities. *Neurocomputing*, *513*, 351–371. https://doi.org/10.1016/j.neucom.2022.09.135

47. Siegel, D., Kraetzer, C., Seidlitz, S., & Dittmann, J. (2021). Media forensics considerations on deepfake detection with hand-crafted features. *Journal of Imaging*, *7*(7). https://doi.org/10.3390/jimaging7070108

48. Singh, A., Saimbhi, A. S., Singh, N., & Mittal, M. (2020). DeepFake Video Detection: A Time-Distributed Approach. *SN Computer Science*, *1*(4), 1–8. https://doi.org/10.1007/s42979-020-00225-9

49. Thaseen Ikram, S., V, P., Chambial, S., Sood, D., & V, A. (2023). A Performance Enhancement of Deepfake Video Detection through the use of a Hybrid CNN Deep Learning Model. *International Journal of Electrical and Computer Engineering Systems*, *14*(2), 169–178. https://doi.org/10.32985/ijeces.14.2.6

50. Trinh, L., & Liu, Y. (2021). An Examination of Fairness of AI Models for Deepfake Detection. *IJCAI International Joint Conference on Artificial Intelligence*, 567–574. https://doi.org/10.24963/ijcai.2021/79

51. Wang, T., Cheng, H., Chow, K. P., & Nie, L. (2023). Deep Convolutional Pooling Transformer for Deepfake Detection. *ACM Transactions on Multimedia Computing, Communications and Applications*, *19*(6). https://doi.org/10.1145/3588574

52. Waseem, S., Abu-Bakar, S. A. R. S., Omar, Z., Ahmed, B. A., Baloch, S., & Hafeezallah, A. (2023). Multi-attention-based approach for deepfake face and expression swap detection and localization. *Eurasip Journal on Image and Video Processing*, *2023*(1). https://doi.org/10.1186/s13640-023-00614-z

53. Xia, Z., Qiao, T., Xu, M., Wu, X., Han, L., & Chen, Y. (2022). Deepfake Video Detection Based on MesoNet with Preprocessing Module. *Symmetry*, *14*(5), 1–14. https://doi.org/10.3390/sym14050939

54. Xu, Y., Liang, J., Jia, G., Yang, Z., Zhang, Y., & He, R. (2023). *TALL: Thumbnail Layout for Deepfake Video Detection*. http://arxiv.org/abs/2307.07494

55. Yang, S., Hu, S., Zhu, B., Fu, Y., Lyu, S., Wu, X., & Wang, X. (2023). *Improving Cross-dataset Deepfake Detection with Deep Information Decomposition*. 1–10. http://arxiv.org/abs/2310.00359

56. Yu, P., Xia, Z., Fei, J., & Lu, Y. (2021). A Survey on Deepfake Video Detection. *IET Biometrics*, *10*(6), 607–624. https://doi.org/10.1049/bme2.12031

57. Yu, Y., Ni, R., Zhao, Y., Yang, S., Xia, F., Jiang, N., & Zhao, G. (2023). MSVT: Multiple Spatiotemporal Views Transformer for DeepFake Video Detection. *IEEE Transactions on Circuits and Systems for Video Technology*, *PP*(c), 1. https://doi.org/10.1109/TCSVT.2023.3281448

58. Zheng, Y., Bao, J., Chen, D., Zeng, M., & Wen, F. (2021). Exploring Temporal Coherence for More General Video Face Forgery Detection. *Proceedings of the IEEE International Conference on Computer Vision*, 15024–15034. https://doi.org/10.1109/ICCV48922.2021.01477