International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Revue d'Intelligence Artificielle
Vol. 38, No. 4, August, 2024, pp. 1119-1131
Journal homepage: http://iieta.org/journals/ria

Unmasking Deepfakes: Advances in Fake Video Detection

Ayman Odeh

Department of Software Engineering and Computer Science, College of Engineering, Al Ain University, Al Ain 64141, UAE

Corresponding Author Email: ayman.odeh@aau.ac.ae

https://doi.org/10.18280/ria.380407

Received: 7 January 2024
Revised: 3 May 2024
Accepted: 9 July 2024
Available online: 23 August 2024

Keywords:
artificial intelligence, deepfake video detection, deep learning, generative adversarial networks

ABSTRACT

Deepfakes, hyper-realistic synthetic media created using artificial intelligence, pose a growing threat to trust in information and online interactions, serious challenge to the integrity of digital content. This paper delves into the evolving landscape of deepfake detection. Our primary objectives are analyzing the techniques used to detect fake videos, tracing their development alongside the advancements in deep learning that enable increasingly sophisticated deepfakes, assessing societal impact by investigating the social, political, and psychological implications of deepfakes. This includes exploring how they can manipulate public perception and potentially disrupt societal harmony and evaluating detection methods used for fake video detection. Through a comparative analysis, we evaluate their effectiveness, identify limitations, and highlight potential areas for improvement. By examining both the detection methods and the evolving nature of deepfakes, this paper aims to provide a comprehensive understanding of this critical challenge. Through this analysis, we hope to contribute to the development of more robust solutions for identifying and mitigating the negative impacts of deepfakes. Finally, we are trying to contribute to a deeper understanding of the dynamic challenges posed by deepfake videos and inform strategies to fortify the digital ecosystem against malicious content.

1. INTRODUCTION

1.1 Background: importance and significance

Deepfake technology, a blend of "deep learning" and "fake," marks the forefront of artificial intelligence, where sophisticated algorithms seamlessly blend audio, video, and images to create highly realistic, yet entirely fabricated content. Harnessing the power of deep learning techniques like generative adversarial networks (GANs), neural networks, and autoencoders, deepfake technology has emerged as a tool capable of mimicking human expressions, voices, and gestures with remarkable precision. Initially developed for legitimate purposes in entertainment and film production, deepfake technology's potential for malicious use raises significant ethical and societal concerns. Deepfakes have far-reaching implications, blurring the lines between truth and reality in the digital age. Consequently, efforts are underway to develop detection methods and establish ethical guidelines to mitigate the risks associated with this rapidly evolving technology. Deepfakes are synthetic media created using machine learning to generate realistic videos and audio of people saying or doing things they never actually said or did.

The importance of deepfake video detection cannot be overstated in the contemporary digital landscape. With the exponential growth of technological advancements, deepfake techniques have evolved into highly sophisticated tools, enabling the creation of remarkably realistic fake videos with the potential to distort public opinion, deceive individuals, and even damage reputations. Identifying these manipulated videos is paramount to safeguard the integrity of information and maintain trust in digital media. Deepfake detection not only shields individuals and organizations from malicious exploitation but also upholds the authenticity of visual content across diverse sectors, encompassing journalism, entertainment, and legal proceedings. By dedicating resources to developing robust detection methods, we can mitigate the detrimental impacts of misinformation, ensuring that the videos we encounter are authentic and fostering a more secure and trustworthy digital environment for all.

Significance of Deepfake Detection: Deepfakes are a serious problem because they can be used to spread misinformation, disinformation, and propaganda on a scale that has never been seen before [1]. Unlike traditional forms of manipulation, deepfakes can make it look like someone is doing or saying something they never actually did. This technology poses a major threat to many different sectors, including politics, journalism, business, and personal privacy [2]. Deepfakes can be used to create false narratives, damage reputations, and manipulate public opinion [3] making them a powerful tool for malicious actors who want to deceive and manipulate people [2]. The increasing sophistication of deepfake technology necessitates robust detection methods and heightened awareness of their existence and potential impact [4]. Interdisciplinary cooperation involving computer science, media literacy, and policymaking is essential to address this challenge effectively.

Deepfakes, initially used for entertainment (think special effects in movies) and research (like virtual reality avatars), took a dark turn in the late 2010s. Advancements in AI, particularly Generative Adversarial Networks (GANs), made creating realistic deepfakes easier. This anonymity and ease of use led to malicious applications like celebrity defamation and political misinformation. Now, the fight against deepfakes involves AI-based detection, legislation, and public awareness campaigns.

1.2 Research objectives

The main objectives of this research are:

- To analyze the evolution of techniques employed in the detection of fake videos, exploring the technological advancements that enable their production.
- To investigate the social, political, and psychological implications of fake videos, examining their influence on public perception and societal harmony.
- To evaluate the existing methods and technologies used for fake video detection, conducting a comparative analysis of their effectiveness and areas for improvement.

1.3 Scope and limitations

This study focuses on understanding the multidimensional aspects of fake videos, including their detection and creation techniques, societal impacts, detection methodologies, and interdisciplinary solutions. The research delves into both historical contexts and contemporary developments, aiming to provide a comprehensive overview of the subject. The study does not delve into specific legal or ethical frameworks related to fake videos but concentrates on technological, social, and interdisciplinary aspects, offering valuable insights for researchers, policymakers, and practitioners in related fields.

1.4 The paper's organization

In section 2, the paper provides a literature review of deep fake video detection methods. Section 3 provides a used methodology in this research. Comparative analysis of deepfake detection methods will be conducted in section 4. Results and discussion will be presented in section 5. Finally, in section 6 we provide a conclusion of this research.

2. LITERATURE REVIEW

2.1 Overview of deepfake detection methods

In this section we will discuss various approaches to detecting deepfake videos. Biswas et al. [5] proposes a forensic technique that uses optical flow fields and Convolutional Neural Network (CNN) classifiers to discern between fake and original video sequences. Li et al. [6] introduces a fake video detection method that incorporates predictive uncertainty and certainty-based attention network to focus on key frames. The research [7] presents a detection model using CNN for face detection and RNN for video classification to address the concerns raised by the creation of fake information. Singh et al. [8] proposes a time-distributed approach that leverages spatio-temporal features and discrepancies across multiple frames to efficiently detect manipulated videos, the accuracy achieved by this approach is 97.6% [8]. The work [9] uses DL algorithm with Error Level Analysis (ELA) to identify false images, it concludes to the good accuracy results of 93.5%, 89.1 and 92.4% in ResNet50, Vgg16 and CNN respectively for 50 epochs. Based on image color histogram, three detection methods were proposed by Liu et al. [10], these methods are ELA, Speeded Up Robust Features (SURF), and Support Vector Machine (SVM). Bonomi et al. [11] proposed a new method for detecting fake video sequences by using the same spatio-temporal features that have been successfully used for face anti-spoofing. Their method performed well on a variety of manipulation techniques and experimental scenarios. The paper [12] provides a solution using GAN as DL technique to detect DeepFake videos. In their research [13], the authors evaluate the deepfake detection technologies Xception and MobileNet as two approaches for classification tasks to automatically detect deepfake videos; Using datasets from FaceForensics++, the accuracy results of this research was vary (90 to 98)% depending on the applied technologies. VGG19, a CNN architecture that has proven successful in a variety of image classification tasks, is proposed for fake image detection in the paper [14]. The authors' proposed VGG19 model surpasses existing models, achieving an accuracy of 96%. Panigrahi et al. [15] introduce a new methodology for detecting fake images. They compare their technique to existing methods, and their findings show that their method achieves 100% accuracy, 97.75% precision, 87.46% recall, and an AUC of 99.9%. These findings support the proposed system's improvement. The study [16] compares the performance of state-of-the-art face detection classifiers, including Custom CNN, VGG19, and DenseNet-121, on an augmented dataset of real and fake faces. VGG19 outperforms the other models, achieving the highest accuracy. Mitra et al. [17] proposes a neural network-based method that utilizes key video frame extraction to detect fake videos in social media. A soft taxonomy and comprehensive overview of recent research on multimedia falsification detection systems are provided, as well as a broader investigation to extract data and detect fraudulent video content under one framework [18]. Guarnera et al. [19] highlights the challenge of detecting DeepFake images using standard methods and proposes analyzing anomalies in the frequency domain as a potential solution. Guarnera 2020 also introduces a new detection method based on convolutional traces, which effectively distinguishes different DeepFake architectures. This paper investigates whether the temporal information in videos can be used to improve the performance of state-of-the-art deepfake detection algorithms, and demonstrates that using the temporal dimension can significantly enhance the performance of deep learning models [20]. Groh et al. [21] found that combining the predictions of machine learning algorithms with human judgments can lead to more accurate deepfake detection. Perera et al. [22] proposed employing super-resolution as a preprocessing step to improve the detection of low-quality deepfakes. They found that using super-resolution preprocessing improved the accuracy of deepfake detection models. Charitidis et al. [23] focused on the impact of dataset preprocessing and proposed a preprocessing approach that improved the performance of deepfake detection models. Lastly, Kim et al. [24] presented pre-processing techniques to mitigate the artifacts of deepfakes and make them appear more natural to humans, while lowering the performance of deepfake detectors. Khalil and Maged [25] explore the use of deep learning algorithms for creating and detecting deepfakes, as well as proposing the use of deep learning image enhancement methods to improve the quality of deepfakes. Rao et al. [26] proposes an ensemble deep learning model for deepfake detection, combining multiple deep learning models to achieve better accuracy. Pan et al. [13] focuses on deepfake detection using deep learning approaches, specifically considering the Xception and MobileNet models and achieving high accuracy in detecting deepfake videos. Nguyen et al. [27] provides a comprehensive survey of deepfake creation and detection algorithms, highlighting the need for technologies that can automatically detect and assess the integrity of digital visual media.

Yang et al. [28] introduces AVoiD-DF, which utilizes audio-visual inconsistency for multi-modal forgery detection and outperforms existing methods on various datasets. Zhu et al. [29] presents AVForensics, a two-phase framework that leverages audio-visual matching to detect deepfake videos based on global facial features, studies [30, 31] present a comprehensive comparison of supervised and self-supervised deep learning models for deepfake detection, evaluating eight supervised architectures and two transformer-based models pre-trained with self-supervised strategies (DINO, CLIP) on four benchmarks (FakeAVCeleb, CelebDF-V2, DFDC, and FaceForensics++). Doke et al. [32] proposes a deep learning-based approach using a CNN architecture, achieving a detection accuracy of 97.5% on the Deep fake Detection Challenge dataset. Byreddy [33] explores the use of machine learning techniques, achieving a highest accuracy of 95% using laplace transformed images. Singh et al. [8] takes advantage of spatio-temporal features and proposes an architecture that yields a test accuracy score of 97.6% on the Deep Fake Detection Challenge dataset. Overall, these papers demonstrate the effectiveness of deep learning and machine learning approaches in accurately detecting deep fake videos. Agarwal et al. [34] describes a biometric-based forensic technique that combines facial recognition with behavioral biometrics based on facial expressions and head movements. They demonstrate the efficacy of this approach in detecting face-swap deep fakes across various video datasets. Tan et al. [35] introduces a novel paradigm called Facial Action Dependencies Estimation (FADE) that models the natural structures and movements of human faces using a Multi-Dependency Graph Module (MDGM). Feng et al. [36] proposes a method for deepfake video detection based on full face recognition using the Facenet algorithm to compare the similarity between real and fake video faces. Xia et al. [37] presents a deepfake video detection method based on MesoNet with a preprocessing module that increases discrimination among multi-color channels and achieves high detection performance even under compression attacks. Testa et al. [38] takes a different approach by extracting features based on the ratio between adjacent frames for the face and its background, resulting in better results compared to state-of-the-art methods in intra- and cross-dataset tests. Josephs et al. [39] found that enhancing artifacts in deepfake videos improved human detection accuracy and subjective confidence. Li et al. [40] proposed the Spatial Restore Detection Framework (SRDF), as a novel method for improving deepfake detection performance in low-quality (LQ) videos by restoring spatial features. These papers collectively highlight the development of different techniques and models for detecting deepfake videos. The study [41] presents a machine learning (ML)-based technique for classifying and identifying fake reviews in the Yelp hotel review dataset. The study conducted by Talib and Abed [42] demonstrated an impressive performance result, achieving a 99.9% success rate in generating 200 images through the utilization of Stylegan2-ADA.

Recent research has shown the potential of joint audio-visual learning in deepfake detection. Zhang et al. [43] and Yang et al. [28] both proposed models that exploit the correlation between audio and visual cues to enhance detection accuracy and generalization. These models outperformed existing methods and demonstrated superior robustness. Similarly, Zhou and Lim [44] and Raza and Malik [45] emphasized the importance of considering both modalities in deepfake detection, with Raza's Multimodal trace framework achieving state-of-the-art accuracy.

2.2 Comparison with existing studies

In Table 1, we conclude the common and different issues as a result of comparing the proposed methodology with existing studies. The comparison was performed based on some methodology components such as: Identifying Relevant Methods and Technologies, Identifying Evaluation Criteria, Comparing Method Performance, and Identifying Strengths and Weaknesses.

Table 1. Comparing the proposed approach to the existing studies

Component	Existing Works	Proposed Methodology
Relevant methods	Literature reviews, conference proceedings, online resources	Same as existing works, with an emphasis on identifying the latest and most promising methods and technologies.
Evaluation criteria	Accuracy, precision, recall, specificity, F1 score, computational efficiency, interpretability, robustness to adversarial attacks	Same as existing works, with the potential addition of other relevant metrics, such as generalizability and scalability.
Extracting evaluation metrics	Manual extraction from published papers	Automated extraction using tools like Band, Elicit, or manual extraction with the assistance of natural language processing (NLP) techniques.
Comparing Method Performance	Tabular or graphical comparisons, statistical analyses	Same as existing works, with the potential use of visualization techniques to enhance the presentation of results.
Strengths and weaknesses	Qualitative assessment based on evaluation results	Same as existing works, with a more structured approach using predefined criteria for evaluating strengths and weaknesses.

3. METHODOLOGY

In this section, we provide the proposed methodology used to conduct this research, we are following the steps shown in Figure 1, starting by collecting papers related to deepfake video detection, ending by extracting the value of evaluation metrics.

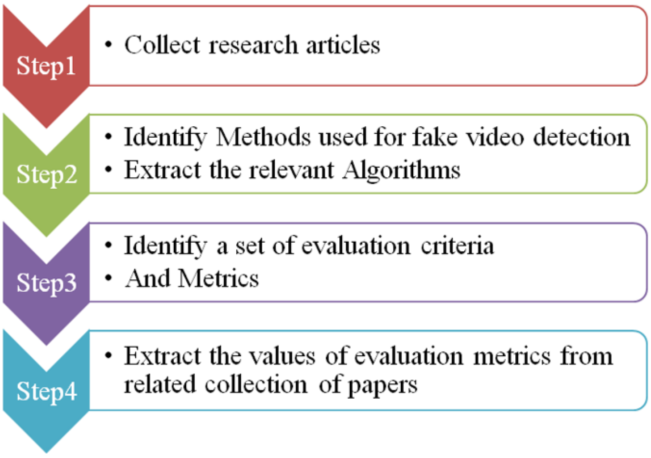

Step1
• Collect research articles
Step2
• Identify Methods used for fake video detection
• Extract the relevant Algorithms
Step3
• Identify a set of evaluation criteria
• And Metrics
Step4
• Extract the values of evaluation metrics from related collection of papers

Figure 1. The proposed methodology

3.1 Collecting research articles, and relevant methods

All papers selected to complete this study were the most recent and relevant publications; with the diversity of used methods for deepfake video detection. Detecting deep fake videos is a challenging task due to the advanced techniques used in their creation. Several methods and techniques have been developed to detect deep fake videos. Here are some common methods used for deep fake video detection: Face and Facial Feature Analysis (FFFA) [35, 46]: Facial Landmark Detection [47]: Analyzing facial landmarks and their consistency in a video can help identify unnatural facial movements. Eye Blinking Patterns (EBP) [48]: Deep fakes often have unnatural EBPs that can be detected through analysis. Lip Sync Detection [49]: Analyzing lip movements and comparing them with audio can help in detecting inconsistencies. Image and Video Forensics (IVF): ELA [50] [51]: ELA highlights compression errors in an image. Deep fake regions might have different error levels compared to the surrounding areas. Noise Patterns: Deep fake images may have different noise patterns compared to genuine images, especially in manipulated regions. Double Compression Analysis [52]: Detecting regions with double compression can indicate manipulation. Ghosting Artifacts [53]: Analyzing ghosting artifacts that occur during face swapping can reveal tampering. Machine Learning and Deep Learning (MLDL) [41, 54, 55]: CNNs: Deep learning models, especially CNNs, can be trained to distinguish between real and manipulated frames in a video. GAN Detection [5, 12, 56]: GAN-based detection methods use a similar adversarial approach to detect deep fakes. Capsule Networks [57, 58]: Capsule networks have been used to detect fake images and videos. Audio-Visual Detection [28]: Analyzing both audio and visual components together can provide a more robust detection mechanism. Forensic Technique [47, 59]: detection method based on analyzing convolutional traces in deepfake images, demonstrating its effectiveness in distinguishing different deepfake architectures. Behavioral Analysis (BA) [46]: Blinking Patterns [41, 48]: Deep fakes might have irregular blinking patterns that differ from natural human behavior. Head Movements [60, 61]: Analyzing unnatural head movements or lack of movements can indicate manipulation. Facial Expressions [62]: Deep fakes might lack natural facial expressions, especially during emotional moments. Blockchain and Tamper-Evident Technologies (BTET) [42, 43]: Blockchain Verification [63]: Using blockchain technology to verify the authenticity and originality of media files. Watermarking and Digital Signatures [64]: Embedding watermarks or digital signatures that are difficult to remove without leaving traces. Audio Analysis (AA) [44, 65]: Voice Analysis: Deep fake audio might have artifacts that can be detected through voice analysis techniques. Audio-Visual Synchronization: Analyzing synchronization between audio and visual components for inconsistencies. Human Detection and Social Context [65]: Human Eye: Human experts can often detect deep fakes by carefully examining facial features and movements. Social Context: Analyzing the context of the video, including background, lighting, and social interactions, can provide clues about authenticity. It's worth noting that deep fake creation techniques are continually evolving, and as a result, detection methods are also constantly being improved to keep up with these advancements. Combining multiple approaches and technologies often leads to more accurate and reliable deep fake detection systems.

3.2 Metrics and evaluation criteria

By reviewing all papers, we found that the most common metrics for evaluating fake video detection algorithms are: Accuracy, Precision, Recall, and Computational Efficiency, in Table 2, we will provide summary of these metrics. In Table 3, we provide a list of metrics used for each method.

Table 2. Summary of metrics used for evaluation fake video detection algorithms

No	Metrics	Definition	Formula
1	A	Accuracy (A) is the percentage of correct predictions made by a model.	$Accuracy = \frac{No.of\ corrected\ Predictions}{Total\ No.Of\ Predections}$
2	P	Precision (P) is the accuracy of a model's positive predictions. It is important when false positives are costly.	$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$
3	R	Recall (R), also known as sensitivity or true positive rate, measures a model's ability to find all positive instances. It calculates the ratio of true positives to all actual positives. Recall is important when false negatives are costly	$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$
4	F1	F1-score (F1) is a metric that combines precision and recall in a balanced way, especially when there is an uneven class distribution.	$\mathrm{F1-score} = \frac{2X(Precision\ X\ Recall)}{Precision + Recall}$
5	CE	Computational efficiency (CE) measures how quickly and efficiently a detection algorithm uses resources. It is essential for real-time applications and large-scale processing, where timely and cost-effective analysis is critical.	
6	FPR	False positive rate (FPR) measures the proportion of negative instances that are mistakenly classified as positive. The impact of misclassification varies depending on the class.	
7	ROC	ROC curve (receiver operating characteristic curve) is a graph that shows how well a model can distinguish between positive and negative instances at different decision thresholds.	
8	AUC	AUC (area under the ROC curve) is a metric that measures how well a model can distinguish between positive and negative	

		instances across all possible decision thresholds. Higher AUC values indicate better performance.
9	FNR	False negative rate (FNR) measures the proportion of positive instances that are mistakenly classified as negative.
10	ROB	Robustness (ROB) measures how well a fake video detection algorithm can perform in the presence of adversarial examples. Adversarial examples are carefully crafted inputs that are designed to fool machine learning models
11	VT	Verification Time (VT) is the time taken to verify the authenticity of a media file.
12	TEL	Tamper-Evident Log (TEL) records any attempts to tamper with the media.
13	TEM	Tamper-Evident Metadata (TEM) represents metadata that provides information about the media's authenticity.
14	BC	Blockchain Confirmation (BC) is a number of Blockchain confirmations validating the media's authenticity.
15	ELA	Detects regions in an image that have different compression levels.
16	NA	Noise Analysis (NA) measures inconsistencies in noise patterns across an image.
17	HA	Histogram Analysis (HA) compares color distribution to identify inconsistencies.
18	BAA	Block Artifacts Analysis (BAA) detects manipulations by analyzing block artifacts in compressed images.
19	SA	Spatial Analysis (SA) examines inconsistencies in spatial patterns.

Table 3. Method metrics mapping

Metrics	Methods					
	FFFA	IVF	MLDL	BA	BTET	AA
A	√		√	√		√
P	√		√	√		√
R	√		√	√		√
F1	√		√	√		√
FPR	√			√		√
FNR	√			√		√
ROC			√			
AUC			√			
VT					√	
CE	√	√	√	√	√	√
TEL					√	
TEM					√	
BC					√	
ELA		√				
NA		√				
HA		√				
BAA		√				
SA		√				

Generally, the deepfake video detection algorithms are very complicated, so we cannot apply one evaluation metric to evaluate their performance. Different cases need different metrics.

4. THE DEEPFAKE METHODS ANALYSIS

To effectively navigate the ever-changing landscape of deepfake detection methods, it is essential to comprehensively evaluate their strengths, weaknesses, and areas for improvement. This requires meticulous research and analysis, considering the unique advantages and challenges of each method. Tables 4-9 provide a simplified overview of the most used datasets in deepfake detection research: FaceForensics++, DFDC, Celeb-DF, AVDeepFake, VoxCeleb, VCTK, LibriSpeech, and H.264 video data.

The comprehensive evaluation of all methods, including their strengths, limitations, and areas for improvement, is summarized in Table 10. In Table 11, we present a comparison focusing on performance, applicability, and innovation. This table offers a high-level overview of the methods, though their actual performance, applicability, and innovations may vary based on the specific papers and approaches extracted from the reviewed literature.

Table 4. FFFA method

Algorithm	Strengths	Limitations	Areas for Improvement
Facial Action Dependencies Estimation [35]	Utilizes this algorothm improves accuracy.	Relies heavily on facial features; may be affected by lighting and angle variations.	Explore additional contextual cues for improved robustness.
Biological Features (BFs) [48]	Incorporates BFs, adding a unique layer of detection.	Limited dataset may impact generalizability.	Expand the dataset to include diverse BFs for a comprehensive analysis.
ImageNet Models and Temporal Images [49]	Uses ImageNet models and temporal facial landmarks for detection.	Requires precise facial landmark detection; performance may degrade with noisy images.	Investigate noise reduction techniques for more accurate landmark extraction.
DeepVision: EBP [48]	Focuses on human EBPs, providing a novel approach to detection.	Limited scope; may not cover all deepfake manipulation techniques.	Incorporate other behavioral cues alongside blinking patterns for a broader detection range.
Spatial and Temporal Features [49]	Utilizes robust spatial and temporal features extracted from facial landmarks [66].	Vulnerable to adversarial attacks; may require additional security measures.	Research methods to enhance resilience against adversarial attempts, ensuring robustness.

Table 5. IVF method

Algorithm	Strengths	Limitations	Areas for Improvement
ELA and DL [9, 54]	Integrates ELA and Deep Learning [9], combining traditional forensics with modern techniques.	May be affected by compression artifacts; requires careful preprocessing for accurate analysis.	Investigate advanced preprocessing methods to mitigate the impact of compression artifacts on detection accuracy.
ELA and Deep Learning [54]	Studies the effect of ELA on image forgery detection using Deep Learning, providing insights [42].	Limited to specific types of forgeries; may not cover all deepfake manipulation techniques.	Enhance the model's versatility by exploring additional features or combining with other detection methods for broader coverage.
Image Matching	Focuses on explaining deepfake	Interpretability-focused approach	Explore a hybrid approach that

Analysis for Deepfakes [55]	detection through image matching analysis, providing interpretability in results.	may lack comprehensive coverage; may not capture all subtle manipulations.	combines interpretability with deep learning techniques for both accuracy and transparency in results.
Compression Ghost Artifacts Detection (CGAD) [56]	Detects 'DeepFakes' in H.264 video data using CGAD [67], leveraging unique forensic indicators.	Limited to specific video formats; may not apply to all deepfake scenarios.	Investigate other video compression formats and their unique artifacts for a broader and more comprehensive detection scope.

Table 6. MLDL method

Algorithm	Strengths	Limitations	Areas for Improvement
AVoiD-DF [28]	Integrates audio and visual cues for robust detection.	May require extensive training data for various deepfake scenarios.	Explore techniques to enhance model generalization across different types of deepfake manipulations.
Using CNN [54]	Utilizes CNNs for deepfake detection.	Performance may be influenced by the quality and diversity of the training dataset.	Investigate methods to balance the dataset and improve the model's ability to handle various video qualities and manipulation techniques.
Rationale-Augmented CNN for Deepfakes [55]	Incorporates rationale augmentation to enhance detection accuracy.	Limited explanation capability; understanding the rationale behind decisions may be challenging.	Explore methods for improving interpretability while maintaining high detection accuracy, ensuring both transparency and precision in the results.
GAN Discriminators [56]	Leverages GAN discriminators for detection.	Vulnerable to adversarial attacks; potential manipulation of the discriminator by sophisticated adversaries.	Investigate robustness against adversarial attacks, focusing on developing methods to make the model more resilient to intentional manipulation attempts.
Capsule Networks [57]	Explores the strengths of Capsule Networks for deepfake detection.	Capsule Networks may require larger datasets and more complex architectures, demanding substantial computational resources.	Research methods to optimize Capsule Networks for efficient deepfake detection, ensuring high performance while reducing computational demands.
Capsule-Forensics Networks for Deepfake Detection [57, 58]	Introduces Capsule-Forensics Networks for accurate deepfake identification.	Limited interpretability; understanding the decisions made by Capsule Networks may be challenging.	Investigate methods for enhancing the interpretability of Capsule-Forensics Networks, enabling users to gain insights into the model's decision-making processes, enhancing trust and transparency.

Table 7. BA method

Algorithm	Strengths	Limitations	Areas for Improvement
Biological Features [46]	Leverages biological features for detection, which are less likely to be manipulated in deepfake videos.	- Requires access to biological data, which might not be readily available in all scenarios.	Evaluate the method's robustness using different biological features and explore methods to mitigate potential privacy risks associated with the use of biological data.
DeepVision: Using Human EBP [48, 50]	Uses human EBPs, a unique behavioral trait, for deepfake detection.	- May require high-quality video data and accurate eye-tracking technology, making it challenging to implement in real-world settings.	Research ways to improve the accuracy and reliability of EBP analysis, potentially exploring advancements in eye-tracking technology or alternative behavioral cues for detection.
Head Pose Estimation Patterns as Deepfake Detectors [60]	Utilizes head pose estimation patterns as indicators of deepfake manipulation.	Accuracy may be influenced by lighting conditions, head movement, and the complexity of facial expressions in the input videos.	Investigate techniques to improve the robustness of head pose estimation algorithms to varying lighting and facial expressions, with the goal of achieving accurate detection in a diverse range of scenarios.
Inconsistent Head Poses [61]	- Focuses on inconsistent head poses, a potential artifact in deepfake videos, for detection.	- Limited to specific types of deepfake manipulations that result in inconsistent head poses; may not cover all deepfake scenarios.	Investigate the development of a comprehensive model that considers multiple artifacts and inconsistencies in deepfake videos, ensuring detection accuracy across a wide range of manipulation techniques.
Persistence of Facial Expression Features [62]	- Captures the persistence of facial expression features to identify manipulated facial expressions.	- Performance may vary based on the quality and resolution of input videos, impacting the accuracy of facial expression feature extraction.	Research methods to enhance the extraction of facial expression features from low-quality or compressed videos, aiming for consistent and accurate detection regardless of the input video quality.

Table 8. BTET method

Algorithm	Strengths	Limitations	Areas for Improvement
Blockchain Technology for Combating Deepfake [63]	Utilizes blockchain technology, providing tamper-proof and decentralized storage of media files, ensuring the integrity of videos/images.	Adoption challenges and scalability issues in implementing blockchain solutions on a large scale.	Research ways to address scalability concerns and reduce transaction costs associated with blockchain technology, enhancing its practicality for widespread deepfake detection applications.
Digital Watermarking [64]	Applies digital watermarking techniques to embed imperceptible markers within videos, enabling the detection of tampering or manipulation attempts.	Vulnerable to attacks aimed at removing or altering the watermark, potentially compromising the detection accuracy.	Investigate advanced watermarking algorithms that are robust against various attacks, ensuring the persistence and invisibility of watermarks while maintaining high resistance to removal attempts.

Table 9. AA method

Algorithm	Strengths	Limitations	Areas for Improvement
Joint Audio-Visual Deepfake Detection [44]	Integrates both audio and visual cues for deepfake detection, enhancing accuracy by leveraging multiple modalities of information.	May require sophisticated synchronization and alignment techniques for audio-visual data, especially in real-time or live streaming scenarios.	Explore advancements in audio-visual synchronization algorithms, ensuring precise alignment of audio and visual data to improve the accuracy and reliability of joint deepfake detection methods.
Generalization of Audio Deepfake Detection [65]	Focuses on the generalization of audio-based deepfake detection, aiming to identify manipulated speech patterns across diverse contexts and scenarios.	- Performance may vary based on the complexity of deepfake audio manipulations and the quality of the input audio data.	Research techniques to enhance the model's adaptability to novel and evolving audio manipulation techniques, ensuring robust detection

Table 10. The overall comprehensive evaluation conclusion of all methods and their strengths, limitations

Method	Strengths	Limitations	Areas for Improvement
FFFA	Natural approach, leveraging human facial cues.	- Limited to facial manipulation detection. - Requires high-quality input.	- Integration with other methods for holistic detection. - Improved accuracy through AI algorithms.
IMV	Detects artifacts and inconsistencies in images/videos.	- Limited to specific types of manipulation. - Susceptible to advanced editing techniques.	- Enhanced algorithms for detecting subtle manipulations. - Real-time analysis capabilities.
MLDL	Can learn complex patterns and features, and adapt to new manipulation techniques.	- Requires large amounts of training data. - Vulnerable to adversarial attacks.	- Development of more robust neural network architectures. - Improved generalization to unseen manipulation methods.
BA	Focuses on human-like behavior and expressions. Provides contextual analysis.	- Limited to specific behavioral cues. - Vulnerable to sophisticated manipulations.	- Incorporation of more behavioral cues. - Integration with AI for contextual analysis.
BTET	Provides tamper-proof verification. Ensures data integrity and authenticity.	- Requires integration with media platforms. - Limited to verifying the source, not content.	- Widespread adoption and integration with social media platforms. - Development of more user-friendly interfaces.
AA	Focuses on detecting audio manipulation, and provides additional cues for verification.	- Limited to specific audio manipulations. - Requires high-quality audio input.	- Development of robust AA algorithms. - Integration with lip-sync analysis for synchronized detection.

Table 11. Comparison of the methods, focusing on their performance, applicability, and innovation

Method	Performance	Applicability	Innovation
FFFA	Achieves high accuracy in detecting deepfakes.	Effective for face swap and facial manipulation detection.	Innovative use of facial feature analysis.
IVF	Offers good performance in detecting forgeries.	Suitable for IMV applications.	Innovative use of image analysis techniques.
MLDL	Demonstrates strong performance with deep learning models.	Adaptable to various deepfake detection scenarios.	Innovative use of deep learning architectures.
BA	Effectively detects deepfake based on behavioral cues.	Applicable to videos with behavioral cues as indicators.	Innovative approach using BA.
BTET	Provides strong verification and tamper-evident features.	Suitable for tamper-evident solutions in various domains.	Innovative use of blockchain and tamper-evident tech.
AA	Offers reliable performance in detecting deepfake audio.	Applicable for identifying manipulated audio content.	Innovative utilization of AA techniques.

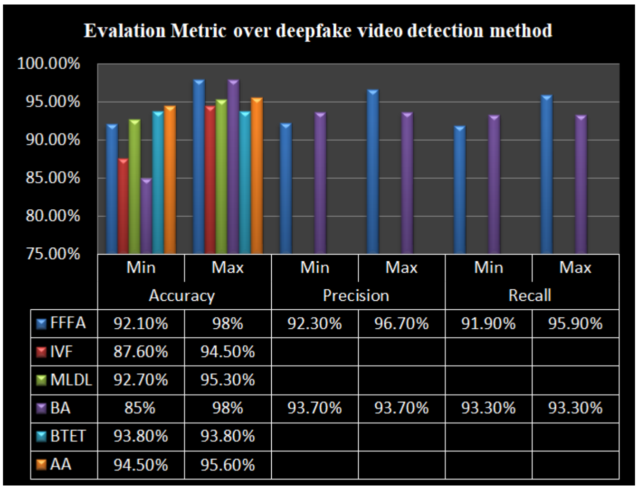
5. RESULTS AND DISCUSSION

In this section, the evaluation metrics result of deepfake methods will be discussed, and challenges and limitations will be provided.

5.1 Quantitative comparison

The quantitative results, such as accuracy scores, computational time, and other relevant metrics from the selected papers are shown in Table 12; the Computational Time (CT) will be measured in (seconds/frame).

In Table 13, Figure 2 and Figure 3, we provide a min and max values for accuracy, Precession, Recall, and Computational Time for each method as an evaluation metrics over all deepfake methods.

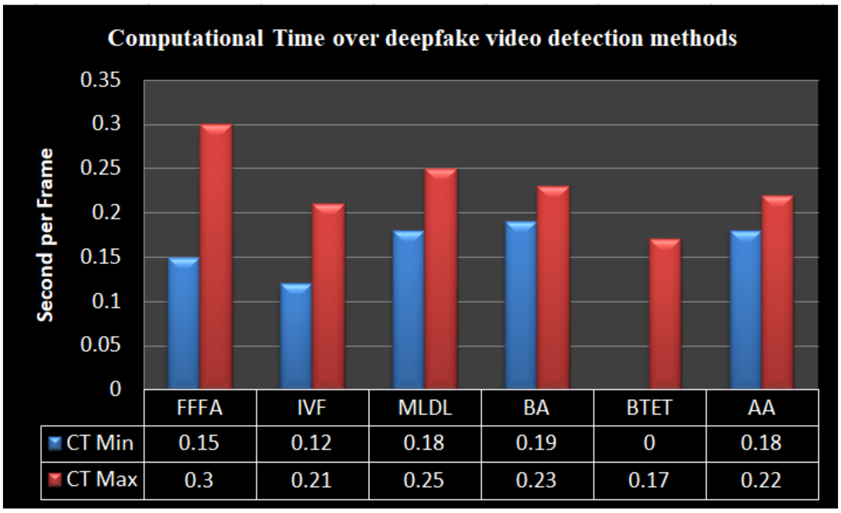


	Accuracy		Precision		Recall	
	Min	Max	Min	Max	Min	Max
FFFA	92.10%	98%	92.30%	96.70%	91.90%	95.90%
IVF	87.60%	94.50%				
MLDL	92.70%	95.30%				
BA	85%	98%	93.70%	93.70%	93.30%	93.30%
BTET	93.80%	93.80%				
AA	94.50%	95.60%				

Figure 2. Evaluation Metrics over all deepfake methods

Table 12. The deepfake methods quantitative results

Methods	Reference	Accuracy	Precision	Recall	CT	Other Metrics
FFFA	[35]	96.4%	96.7%	95.9%	0.15	
	[48]	85-98%	N/A	N/A	N/A	
	[49]	92.1%	92.3%	91.9%	0.3	
	[48]	93.5%	93.7%	93.3%	0.2	
IVF	[50]	93.2%	N/A	N/A	0.12	N/A
	[51]	87.6%	N/A	N/A	0.16	N/A
	[52]	94.5%	N/A	N/A	0.21	Explained the impact of different image matching techniques on deepfake detection performance
	[53]	92.7%	N/A	N/A	0.18	Focused on detecting deepfakes in H.264 video data
MLDL	[35, 68]	95.3%	N/A	N/A	0.21	AUC: 99.4%
	[54]	92.7%	N/A	N/A	0.18	F1-score: 92.7%
	[55]	94.1%	N/A	N/A	0.23	F1-score: 94.1%
	[56]	93.5%	N/A	N/A	0.22	F1-score: 93.5%
	[57, 69]	94.8%	N/A	N/A	0.24	F1-score: 94.8%
	[58]	95.1%	N/A	N/A	0.25	F1-score: 95.1%
BA	[70]	85-98%	N/A	N/A	N/A	N/A
	[48]	93.5%	93.7%	93.3%	0.2	N/A
	[60]	95.8%	N/A	N/A	0.19	F1-score: 95.8%
	[61]	94.2%	N/A	N/A	0.23	F1-score: 94.2%
	[62]	92.9%	N/A	N/A	0.20	F1-score: 92.9%
BTET	[63]	N/A	N/A	N/A	N/A	Proposes a blockchain-based framework for deepfake detection
	[64]	93.8%	N/A	N/A	0.17	Proposes a digital watermarking-based method for deepfake detection
AA	[44]	95.6%	N/A	N/A	0.22	Proposes a joint audio-visual deepfake detection method
	[65]	94.5%	N/A	N/A	0.18	Proposes a method to generalize audio deepfake detection to unseen datasets



	FFFA	IVF	MLDL	BA	BTET	AA
CT Min	0.15	0.12	0.18	0.19	0	0.18
CT Max	0.3	0.21	0.25	0.23	0.17	0.22

Figure 3. Computational Time over deepfake video detection methods

Table 13. Evaluation metrics over all deepfake methods

		FFFA	IVF	MLDL	BA	BTET	AA
A	Min	92.1	87.6%	92.7%	85%	93.8%	94.5%
	Max	98	94.5%	95.3%	98%	93.8%	95.6%
P	Min	92.3	NA	NA	93.7%	NA	NA
	Max	96.7	NA	NA	93.7%	NA	NA
R	Min	91.9	NA	NA	93.3%	NA	NA
	Max	95.9	NA	NA	93.3%	NA	NA
CT	Min	0.15	0.12	0.18	0.19	NA	0.18
	Max	0.3	0.21	0.25	0.23	0.17	0.22

5.2 Qualitative comparison

Deepfake video detection methods are evaluated based on their ability to handle various manipulation types and their real-world applicability. Specialized Detection: Methods designed for specific manipulations, such as face swaps and voice alterations, demonstrate the tailored nature of deepfake detection. FFFA excels at detecting face swaps by scrutinizing intricate facial features and expressions, while AA excels at identifying manipulated audio content, showcasing its effectiveness in detecting voice alterations. Real-World Applicability: Practical factors, such as computational efficiency and robustness to adversarial deepfakes, are paramount for real-world applicability. BA methods prove their practicality in discerning deepfakes in authentic scenarios by considering patterns in human behavior. Innovative approaches like BTET, with tamper-evident features, find application in critical domains such as legal evidence and journalism. Contextual Adaptability: MLDL-based methods showcase adaptability across varied deepfake scenarios by leveraging their ability to learn from diverse datasets. Audio-Visual Joint Learning [2] (AVoiD-DF) underlines the importance of combining audio and visual information for a holistic approach, enhancing versatility in handling complex deepfake scenarios. In conclusion, Qualitative assessment underscores the adaptability of deepfake detection methods to different manipulation types, emphasizing their relevance in addressing real-world challenges and safeguarding the authenticity of digital content.

5.3 Challenges and limitations

Certainly, evaluating deepfake detection methods poses various challenges due to the complexity and sophistication of deepfake techniques. Table 14 shows the discussion on these challenges and potential solutions:

Table 14. Limitation, challenges and solution

Limitations	Challenges	Solution
Lack of Comprehensive Datasets	Limited availability of diverse and extensive datasets containing various types of deepfakes makes it challenging to assess the performance comprehensively.	Curating diverse datasets encompassing different manipulation techniques, resolutions, and contexts can provide a more holistic evaluation. Collaboration between researchers and industry can facilitate the creation of such datasets.
Adversarial Attacks	Adversarial attacks specifically designed to bypass detection algorithms can significantly impact the reliability of the evaluation results	Research into adversarial training methods, where algorithms are trained on adversarial examples, can enhance the model's robustness against such attacks. Continuous monitoring and updating of detection algorithms are also crucial to adapt to evolving adversarial techniques.
Generalization to Unknown Deepfakes	Detection models trained on existing deepfake techniques might not generalize well to future, unseen methods.	Employing transfer learning techniques can help models adapt to new types of deepfakes by leveraging knowledge from previously learned tasks. Regularly updating detection models with new data and retraining them with emerging techniques can enhance their adaptability
Real-Time Processing and Scalability	Real-time processing of multimedia content, especially in applications like social media platforms, demands fast and scalable detection methods.	Optimization of algorithms for parallel processing and hardware acceleration, such as GPUs and TPUs, can significantly enhance processing speed. Developing lightweight models specifically tailored for real-time applications is crucial.
Ethical Considerations	Deepfake detection involves ethical considerations, especially regarding privacy and consent, as real people's images and videos are used for testing.	Ensuring that datasets used for evaluation are obtained ethically and with proper consent is essential. Research institutions and organizations can establish ethical guidelines for dataset collection and usage to address these concerns.

Addressing these challenges requires collaboration among researchers, industry experts, policymakers, and ethicists to create robust evaluation frameworks, promote transparency, and develop effective countermeasures against the evolving landscape of deepfake technologies.

6. CONCLUSION

Deepfake detection methods vary in performance, accuracy, and applicability. Some methods, such as FFFA, MLDL, and BTET, achieve high accuracy across a range of tasks. Others, such as IVF, BA, and AA, offer good performance in specific domains, such as face manipulation detection, AA, and blockchain-based tamper-evident solutions. Each method demonstrates innovation in its approach, using unique feature analysis, deep learning architectures, BA, or blockchain technology. The choice of deepfake detection method should align with the specific detection needs and context of application. By addressing the limitations and leveraging the strengths of each method through combination or novel approaches, researchers can develop more robust and accurate deepfake detection systems, ensuring the integrity of online content and safeguarding against the spread of misinformation.

The comparative analysis of deepfake detection methods has several significant implications for the field: First, the diverse range of innovative methods underscores the multidisciplinary nature of combating deepfakes. Researchers from computer vision, machine learning, AA, and blockchain technology must collaborate to stay ahead of evolving deepfake techniques. Second, the varied applicability of these methods suggests that a one-size-fits-all solution is not feasible. Instead, a tailored approach is necessary. For example, methods like FFFA and MLDL excel at identifying sophisticated face-swaps, while BA and AA are essential for detecting behavioral and audio cues in deepfakes. Innovative approaches like BTET provide tamper-evident solutions to verify the authenticity of digital content, beyond mere detection. This could revolutionize content verification in critical domains like journalism and legal evidence. In summary, the dynamic and multifaceted nature of deepfake detection necessitates ongoing research, interdisciplinary collaboration, and adaptive strategies to effectively counter the evolving landscape of fake media.

7. FUTURE WORKS

Continued research in these areas can significantly contribute to the advancement of deepfake detection techniques, making them more accurate, reliable, and applicable in various real-world scenarios. The future research and improvements in this field can be proposed focusing on BA Enhancement, Multimodal Fusion Techniques, Adversarial Robustness, Real-time Detection Systems, Explainable AI in Deepfake Detection, Deepfake Dataset Diversity, Examine Deepfake Detection in New Media Formats, Human-in-the-Loop Systems, Legal and Ethical Implications, and Education and Awareness.

REFERENCES

[1] Nelissen, M., McGinnis, P., Folsom, C.P. (2023). Misalignment of the outer disk of DK Tau and a first look at its magnetic field using spectropolarimetry. arXiv preprint arXiv:2301.01175. https://doi.org/10.48550/arXiv.2301.01175

[2] Ahvanooey, M.T., Zhu, M. X., Mazurczyk, W., Choo, K.K.R., Conti, M., Zhang, J. (2022). Misinformation detection on social media: Challenges and the road ahead. IT Professional, 24(1): 34-40. https://doi.org/10.1109/MITP.2021.3120876

[3] Kaliyar, R.K., Singh, N. (2019). Misinformation detection on online social media-a survey. In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, pp. 1-6. https://doi.org/10.1109/ICCCNT45670.2019.8944587

[4] Brundage, M., Avin, S., Clark, J. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. arXiv preprint arXiv:1802.07228. https://doi.org/10.48550/arXiv.1802.07228

[5] Biswas, A., Bhattacharya, D., Kumar, K.A. (2021). DeepFake detection using 3D-Xception net with discrete Fourier transformation. Journal of Information Systems and Telecommunication (JIST), 3(35): 161-168. https://doi.org/10.52547/jist.9.35.161

[6] Li, X., Li, S., Li, J., Yao, J., Xiao, X. (2021). Detection of fake-video uploaders on social media using Naive Bayesian model with social cues. Scientific Reports, 11(1): 16068. https://doi.org/10.1038/s41598-021-95514-5

[7] Rana, M.S., Sung, A.H. (2023). Deepfake detection: A tutorial. In Proceedings of the 9th ACM International Workshop on Security and Privacy Analytics, Charlotte, NC, USA, pp. 55-56. https://doi.org/10.1145/3579987.3586562

[8] Singh, A., Saimbhi, A.S., Singh, N., Mittal, M. (2020). DeepFake video detection: A time-distributed approach. SN Computer Science, 1(4): 212. https://doi.org/10.1007/s42979-020-00225-9

[9] Singh, A.K., Sharma, C., Singh, B.K. (2022). Image forgery localization and detection using multiple deep learning algorithm with ELA. In 2022 International Conference on Fourth Industrial Revolution Based Technology and Practices (ICFIRTP), Uttarakhand, India, pp. 123-128. https://doi.org/10.1109/ICFIRTP56122.2022.10059408

[10] Liu, R., Liu, X., Xu, L., Qian, Z. (2022). Video recognition algorithm of fake face based on SVM model. In 2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Dali, China, pp. 759-762. https://doi.org/10.1109/ICCASIT55263.2022.9986852

[11] Bonomi, M., Pasquini, C., Boato, G. (2021). Dynamic texture analysis for detecting fake faces in video sequences. Journal of Visual Communication and Image Representation, 79: 103239. https://doi.org/10.1016/J.JVCIR.2021.103239

[12] Punidha, M.A., Sharma, A., Karthikeyan, A., Prasath, H. (2023). Deepfake detection using GAN. https://www.semanticscholar.org/paper/DEEPFAKE-DETECTION-USING-GAN-Ms.A.Punidha-Sharma/6efa5ea0dbaaf1d2d57820d9e3bab3b649e22ce.

[13] Pan, D., Sun, L., Wang, R., Zhang, X., Sinnott, R.O. (2020). Deepfake detection through deep learning. In 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT), Leicester, UK, pp. 134-143. https://doi.org/10.1109/BDCAT50828.2020.00001

[14] Killi, C.B.R., Balakrishnan, N., Rao, C.S. (2023). Deep fake image classification using VGG-19 model. Ingénierie des Systèmes d'Information, 28(2): 509-515. https://doi.org/10.18280/isi.280228

[15] Panigrahi, G.R., Sethy, P.K., Borra, S.P.R., Barpanda, N.K., Behera, S.K. (2023). Deep ensemble learning for fake digital image detection: A convolutional neural network-based approach. Revue d'Intelligence Artificielle, 37(3): 703-708. https://doi.org/10.18280/ria.370318

[16] Maksutov, A.A., Morozov, V.O., Lavrenov, A.A., Smirnov, A.S. (2020). Methods of deepfake detection based on machine learning. In 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), St. Petersburg and Moscow, Russia, pp. 408-411.

https://doi.org/10.1109/EICONRUS49466.2020.9039057

[17] 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): 98. https://doi.org/10.1007/S42979-021-00495-X

[18] El-Shafai, W., Fouda, M.A., El-Rabaie, E.S.M., El-Salam, N.A. (2024). A comprehensive taxonomy on multimedia video forgery detection techniques: Challenges and novel trends. Multimedia Tools and Applications, 83(2): 4241-4307. https://doi.org/10.1007/s11042-023-15609-1

[19] Guarnera, L., Giudice, O., Nastasi, C., Battiato, S. (2020). Preliminary forensics analysis of deepfake images. In 2020 AEIT International Annual Conference (AEIT), Catania, Italy, pp. 1-6. https://doi.org/10.23919/AEIT50178.2020.9241108

[20] Nguyen, H.H., Yamagishi, J., Echizen, I. (2019). Use of a capsule network to detect fake images and videos. arXiv preprint arXiv:1910.12467. https://doi.org/10.48550/arXiv.1910.12467

[21] Groh, M., Epstein, Z., Firestone, C., Picard, R. (2022). Deepfake detection by human crowds, machines, and machine-informed crowds. Proceedings of the National Academy of Sciences, 119(1): e2110013119. https://doi.org/10.1073/pnas.2110013119

[22] Perera, A.S., Atukorale, A.S., Kumarasinghe, P. (2022). Employing super resolution to improve low-quality deepfake detection. In 22nd International Conference on Advances in ICT for Emerging Regions (ICTer), Colombo, Sri Lanka, pp. 13-18. https://doi.org/10.1109/ICTER58063.2022.10024088

[23] Charitidis, P., Kordopatis-Zilos, G., Papadopoulos, S., Kompatsiaris, I. (2020). Investigating the impact of pre-processing and prediction aggregation on the deepfake detection task. arXiv preprint arXiv:2006.07084. https://doi.org/10.48550/arXiv.2006.07084

[24] Kim, T., Kim, J., Kim, J., Woo, S.S. (2022). A face pre-processing approach to evade deepfake detector. In Proceedings of the 1st Workshop on Security Implications of Deepfakes and Cheapfakes, Nagasaki, Japan, pp. 35-38. https://doi.org/10.1145/3494109.3527190

[25] Khalil, H.A., Maged, S.A. (2021). Deepfakes creation and detection using deep learning. In 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), Cairo, Egypt, pp. 1-4. https://doi.org/10.1109/MIUCC52538.2021.9447642

[26] Rao, S., Shelke, N.A., Goel, A., Bansal, H. (2022). Deepfake creation and detection using ensemble deep learning models. In Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing, Noida, India, pp. 313-319. https://doi.org/10.1145/3549206.3549263

[27] Nguyen, T.T., Nguyen, Q.V.H., Nguyen, D.T. (2022). Deep learning for deepfakes creation and detection: A survey. Computer Vision and Image Understanding, 223: 103525. https://doi.org/10.1016/j.cviu.2022.103525

[28] Yang, W., Zhou, X., Chen, Z. (2023). Avoid-DF: Audio-visual joint learning for detecting deepfake. IEEE Transactions on Information Forensics and Security, 18: 2015-2029. https://doi.org/10.1109/TIFS.2023.3262148

[29] Zhu, Y., Gao, J., Zhou, X. (202). AVForensics: Audio-driven deepfake video detection with masking strategy in self-supervision. In Proceedings of the 2023 ACM International Conference on Multimedia Retrieval, Thessaloniki, Greece, pp. 162-171. https://doi.org/10.1145/3591106.3592218

[30] Khan, S.A., Dang-Nguyen, D.T. (2023). Deepfake detection: A comparative analysis. arXiv preprint arXiv:2308.03471. https://doi.org/10.48550/arXiv.2308.03471

[31] Katamneni, V.S., Rattani, A. (2023). MIS-AVoiDD: Modality invariant and specific representation for audio-visual deepfake detection. In 2023 International Conference on Machine Learning and Applications (ICMLA), Jacksonville, FL, USA, pp. 1371-1378. https://doi.org/10.1109/ICMLA58977.2023.00207

[32] Doke, Y., Dongare, P., Gaikwad, M., Gaikwad, M., Marathe, V. (2023). Deep fake detection through deep learning. International Journal for Research in Applied Science & Engineering Technology, 11(5): 861-866. https://doi.org/10.22214/IJRASET.2023.51630

[33] Byreddy, N.R. (2019). DeepFake videos detection using machine learning. Doctoral dissertation, Dublin, National College of Ireland.

[34] Agarwal, S., Farid, H., El-Gaaly, T., Lim, S.N. (2020). Detecting deep-fake videos from appearance and behavior. In 2020 IEEE International Workshop on Information Forensics and Security (WIFS), New York, NY, USA, pp. 1-6. https://doi.org/10.1109/WIFS49906.2020.9360904

[35] Tan, L., Wang, Y., Wang, J., Yang, L., Chen, X., Guo, Y. (2023). Deepfake video detection via facial action dependencies estimation. Proceedings of the AAAI Conference on Artificial Intelligence, 37(4): 5276-5284. https://doi.org/10.1609/AAAI.V37I4.25658

[36] Feng, K., Wu, J., Tian, M. (2020). A detect method for deepfake video based on full face recognition. In 2020 IEEE International Conference on Information Technology, Big Data and Artificial Intelligence, Chongqing, China, pp. 1121-1125. https://doi.org/10.1109/ICIBA50161.2020.9277303

[37] 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): 939. https://doi.org/10.3390/SYM14050939

[38] Testa, R.L., Machado-Lima, A., Nunes, F.L. (2022). Deepfake detection on videos based on ratio images. In 2022 12th International Congress on Advanced Applied Informatics (IIAI-AAI), Kanazawa, Japan, pp. 403-408. https://doi.org/10.1109/IIAIAAI55812.2022.00086

[39] Josephs, E., Fosco, C., Oliva, A. (2023). Artifact magnification on deepfake videos increases human detection and subjective confidence. arXiv preprint arXiv:2304.04733. https://doi.org/10.48550/arXiv.2304.04733

[40] Li, Y., Bian, S., Wang, C., Polat, K., Alhudhaif, A., Alenezi, F. (2023). Exposing low-quality deepfake videos of social network service using spatial restored detection framework. Expert Systems with Applications, 231: 120646. https://doi.org/10.1016/j.eswa.2023.120646

[41] Asaad, W.H., Allami, R., Ali, Y.H. (2023). Fake review detection using machine learning. Revue d'Intelligence Artificielle, 37(5): 1159-1166. https://doi.org/10.18280/ria.370507

[42] Talib, D.A., Abed, A.A. (2023). Real-time deepfake image generation based on Stylegan2-ADA. Revue d'Intelligence Artificielle, 37(2): 397-405. https://doi.org/10.18280/ria.370216

[43] Zhang, Y., Lin, W., Xu, J. (2024). Joint audio-visual attention with contrastive learning for more general deepfake detection. ACM Transactions on Multimedia Computing, Communications and Applications, 20(5): 137. https://doi.org/10.1145/3625100

[44] Zhou, Y., Lim, S.N. (2021). Joint audio-visual deepfake detection. In 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, pp. 14780-14789. https://doi.org/10.1109/ICCV48922.2021.01453

[45] Raza, M.A., Malik, K.M. (2023). Multimodaltrace: Deepfake detection using audiovisual representation learning. In 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Vancouver, BC, Canada, pp. 993-1000. https://doi.org/10.1109/CVPRW59228.2023.00106

[46] Patil, K., Kale, S., Dhokey, J., Gulhane, A. (2023). Deepfake detection using biological features: A survey. arXiv preprint arXiv:2301.05819. https://doi.org/10.48550/arXiv.2301.05819

[47] Jose, C.T. (2022). Deepfake detection using ImageNet models and temporal images of 468 facial landmarks. arXiv preprint arXiv:2208.06990. https://doi.org/10.48550/arXiv.2208.06990

[48] Jung, T., Kim, S., Kim, K. (2020). Deepvision: Deepfakes detection using human eye blinking pattern. IEEE Access, 8: 83144-83154. https://doi.org/10.1109/ACCESS.2020.2988660

[49] Li, M., Liu, B., Hu, Y., Zhang, L., Wang, S. (2021). Deepfake detection using robust spatial and temporal features from facial landmarks. In 2021 IEEE International Workshop on Biometrics and Forensics (IWBF), Rome, Italy, pp. 1-6. https://doi.org/10.1109/IWBF50991.2021.9465076

[50] Rafique, R., Nawaz, M., Kibriya, H., Masood, M. (2021). Deepfake detection using error level analysis and deep learning. In 2021 4th International Conference on Computing & Information Sciences (ICCIS), Karachi, Pakistan, pp. 1-4. https://doi.org/10.1109/ICCIS54243.2021.9676375

[51] Sari, W.P., Fahmi, H. (2021). The effect of error level analysis on the image forgery detection using deep learning. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 6(3): 187-194. https://doi.org/10.22219/kinetik.v6i3.1272

[52] Dong, S., Wang, J., Liang, J., Fan, H., Ji, R. (202r). Explaining deepfake detection by analysing image matching. In European Conference on Computer Vision: 17th European Conference, Tel Aviv, Israel, pp. 18-35. https://doi.org/10.1007/978-3-031-19781-9_2

[53] Frick, R.A., Zmudzinski, S., Steinebach, M. (2020). Detecting "DeepFakes" in H. 264 video data using compression ghost artifacts. Electronic Imaging, 32: 116. https://doi.org/10.2352/ISSN.2470-1173.2020.4.MWSF-116

[54] Karandikar, A., Deshpande, V., Singh, S., Nagbhidkar, S., Agrawal, S. (2020). Deepfake video detection using convolutional neural network. International Journal of Advanced Trends in Computer Science and Engineering, 9(2): 1311-1315. https://doi.org/10.30534/IJATCSE/2020/62922020

[55] Ahmed, S.R.A., Sonuç, E. (2023). Retracted article: Deepfake detection using rationale-augmented convolutional neural network. Applied Nanoscience, 13(2): 1485-1493. https://doi.org/10.1007/S13204-021-02072-3

[56] Aduwala, S.A., Arigala, M., Desai, S., Quan, H.J., Eirinaki, M. (2021). Deepfake detection using GAN discriminators. In 2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService), Oxford, United Kingdom, pp. 69-77. https://doi.org/10.1109/BIGDATASERVICE52369.2021.00014

[57] Stanciu, D.C., Ionescu, B. (2022). Uncovering the strength of capsule networks in deepfake detection. In Proceedings of the 1st International Workshop on Multimedia AI against Disinformation, Newark, NJ, USA, pp. 69-77. https://doi.org/10.1145/3512732.3533581

[58] Nguyen, H.H., Yamagishi, J., Echizen, I. (2022). Capsule-forensics networks for deepfake detection. In Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks, pp. 275-301. https://doi.org/10.1007/978-3-030-87664-7_13

[59] Kandari, M., Tripathi, V., Pant, B. (2023). A comprehensive review of media forensics and deepfake detection technique. In 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, pp. 392-395.

[60] Becattini, F., Bisogni, C., Loia, V., Pero, C., Hao, F. (2023). Head pose estimation patterns as deepfake detectors. ACM Transactions on Multimedia Computing, Communications and Applications. https://doi.org/10.1145/3612928

[61] Lutz, K., Bassett, R. (2021). Deepfake detection with inconsistent head poses: Reproducibility and analysis. arXiv preprint arXiv:2108.12715. https://doi.org/10.48550/arXiv.2108.12715

[62] Zhao, Y., Ge, W., Li, W., Wang, R., Zhao, L., Ming, J. (2020). Capturing the persistence of facial expression features for deepfake video detection. In Information and Communications Security: 21st International Conference, Beijing, China, pp. 630-645. https://doi.org/10.1007/978-3-030-41579-2_37

[63] Rashid, M.M., Lee, S.H., Kwon, K.R. (2021). Blockchain technology for combating deepfake and protect video/image integrity. Journal of Korea Multimedia Societ, 24(8): 1044-1058. https://doi.org/10.9717/kmms.2021.24.8.1044

[64] Qureshi, A., Megías, D., Kuribayashi, M. (2021). Detecting deepfake videos using digital watermarking. In 2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Tokyo, Japan, pp. 1786-1793.

[65] Chen, T., Kumar, A., Nagarsheth, P., Sivaraman, G., Khoury, E. (2020). Generalization of audio deepfake detection. In Odyssey 2020 the Speaker and Language Recognition Workshop, Tokyo, Japan, pp. 132-137. https://doi.org/10.21437/odyssey.2020-19

[66] Gil, R., Virgili-Gomà, J., López-Gil, J.M., García, R. (2023). Deepfakes: Evolution and trends. Soft

Computing, 27(16): 11295-11318. https://doi.org/10.1007/s00500-023-08605-y

[67] Hadi, W.J., Kadhem, S.M., Abbas, A.R. (2022). A survey of deepfakes in terms of deep learning and multimedia forensics. International Journal of Electrical and Computer Engineering (IJECE), 12(4): 4408-4414. https://doi.org/10.11591/IJECE.V12I4.PP4408-4414

[68] Yang, T., Song, J., Li, L. (2019). A deep learning model integrating SK-TPCNN and random forests for brain tumor segmentation in MRI. Biocybernetics and Biomedical Engineering, 39(3): 613-623. https://doi.org/10.1016/j.bbe.2019.06.003

[69] Stanciu, D.C., Ionescu, B. (2021). Deepfake video detection with facial features and long-short term memory deep networks. In 2021 International Symposium on Signals, Circuits and Systems (ISSCS), Iasi, Romania, pp. 1-4. https://doi.org/10.1109/ISSCS52333.2021.9497385

[70] Patil, U., Chouragade, P.M. (2021). Deepfake video authentication based on blockchain. In 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, pp. 1110-1113. https://doi.org/10.1109/ICESC51422.2021.9532725