A Comprehensive Review of Deepfake Detection Methods and Challenges in Digital Forensics

Full-paper Submission

Thivanka Mohottalalage, Deb Saha & Mark Schmidt
Department of Information Systems, St. Cloud State University, St. Cloud, MN, United States
thivanka.mohottalalage@go.stcloudstate.edu & deb.saha@go.stcloudstate.edu

Abstract

Deepfake technology has advanced quickly, posing serious problems for cybersecurity and digital forensics. Artificial intelligence is used in deep fakes to produce incredibly lifelike yet fake images, movies, and audio. These manipulations present serious risks in areas including identity theft, disinformation, and reputational harm. This study offers a thorough examination of deepfake detection techniques and how they are used in digital forensics. We examine the most recent cutting-edge methods, such as forensic analysis tools, machine learning models, and hybrid approaches that incorporate several detection techniques. The study also looks at the shortcomings of current methods, emphasizing how deep fake algorithms are adaptive and the difficulties this poses for forensic professionals. Additionally, we evaluate how deep learning frameworks like recurrent neural networks (RNNs) and convolutional neural networks (CNNs) may be integrated to increase detection accuracy. Additionally, covered is how to improve model resilience using transfer learning, adversarial training, and data augmentation. We examine how government agencies, business, and academia are working together to create standardized standards for deep fake detection and the value of open-source datasets for research advancement. Additionally, we assess how cutting-edge technologies like federated learning and blockchain might improve the precision and effectiveness of detection systems. This essay also explores the moral and legal ramifications of using deepfakes, emphasizing the need for international collaboration and policy creation to control and lessen its abuse. We also go over case studies that demonstrate how these detection techniques are used in actual forensic investigations. In order for forensic specialists to stay up to date with changing threat environments and technological advancements, we stress the importance of thorough training programs. Lastly, in order to improve digital forensics, this paper examines the most recent deepfake detection methods and suggests future lines of inquiry. AI-powered deepfake technology produces incredibly lifelike but fake media, which presents serious problems for reputational safety, identity protection, and the avoidance of misinformation. We analyze cutting-edge techniques, pointing out their drawbacks and the versatility of deepfake algorithms, including forensic tools, machine learning models, and hybrid approaches. In addition to cutting-edge technologies like federated learning and blockchain, the combination of CNNs, RNNs, transfer learning, and adversarial training is investigated to increase model resilience. In order to combat changing dangers and protect digital evidence, we place a strong emphasis on international cooperation for policy formulation, highlight case studies demonstrating real-world applications, and promote thorough training for forensic specialists.

Keywords: Deepfake, Detection, Digital Forensics, Blockchain, Federated Learning, Machine Learning Models, Recurrent Neural Networks, Convolutional Neural Networks, Data Augmentation, Ethical Implications

Introduction

The fast development of artificial intelligence (AI) and machine learning (ML) has led to the appearance of deepfakes, which have presented significant issues to cybersecurity and digital forensics. The synthetic media known as "deep fakes," which are produced by sophisticated algorithms like generative adversarial networks (GANs), have advanced to the point where they are frequently indistinguishable from real content

(Mirsky & Lee, 2021). Although there are potential benefits to this technology, such as in the creative and entertainment sectors, its abuse has caused serious worries because of how easily it may be used for malevolent purposes. Deepfakes have been linked to political deception, identity theft, disinformation propagation, and reputational damage (Citron & Chesney, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, 2019). Therefore, it is more important than ever to have strong detection systems in place to preserve public confidence in digital media. Dealing with deepfake content presents digital forensic experts with previously unheard-of difficulties because conventional techniques frequently fail to identify these incredibly lifelike fakes (Verdoliva, 2020). Convolutional neural networks (CNNs) and recurrent neural networks (RNNs), two sophisticated machine learning models that are intended to detect artifacts that might not be apparent to the human eye, are the foundation of current state-of-the-art detection algorithms (Nguyen et al., 2019). Despite their effectiveness, these methods need to be updated frequently to stay up with the latest developments in deepfake production, which frequently uses adaptive learning techniques to evade detection systems (Dolhansky et al., 2020). The need for creative methods that combine temporal and geographical analyses for thorough detection is highlighted by the dynamic interplay between forensic specialists and deepfake makers (Haliassos et al., 2021).

To improve the dependability of deep fake detection systems, new technologies like blockchain and federated learning are being investigated in addition to machine learning-based solutions. Forensic specialists now have a reliable instrument for validating evidence thanks to blockchain technology's unchangeable record-keeping capability, which can confirm the legitimacy of media (Tolosana et al., 2020). Conversely, federated learning makes it possible to train detection models cooperatively across several platforms while protecting data privacy, which is essential for extensive forensic investigations (Xu, 2022). Combining these technologies may strengthen existing techniques and aid in the creation of real-time detection systems. The scenario is further complicated by the ethical and legal repercussions of the rise of deepfakes. Deepfake tools are widely accessible, which increases the possibility of abuse and endangers not just people's privacy but also democratic processes and social stability (Citron & Chesney, A Looming Challenge for Privacy, Democracy, and National Security, 2019) . International collaboration and extensive policy frameworks are needed to address these issues and control the usage and dissemination of deepfake technology. Through real-world case studies that illustrate the advantages and disadvantages of existing detection techniques, this study investigates these complex problems.

Studying the rapid changes in machine learning and the related defenses being created by the cybersecurity and digital forensic groups is essential to better understanding the effects and difficulties of deepfakes. Concerns regarding deepfake technology' potential for manipulation and exploitation in a variety of domains, such as politics, financial markets, and personal privacy, have increased because to their extreme complexity (Agarwal et al., 2022). Research on developing robust detection algorithms that use multimodal data fusion and sophisticated signal processing approaches to uncover hidden discrepancies in deepfake content has therefore increased dramatically (Gupta et al., 2024). As mentioned, the combination of federated learning with blockchain offers more approaches to guarantee the integrity of forensic evidence and improve the credibility of digital material. Digital forensic professionals can create a verifiable chain of custody that confirms the legitimacy of media files throughout their existence by integrating blockchain-based verification methods (Akhtar, 2023). Federated learning, on the other hand, provides a decentralized method that protects user privacy while combining resources to create more resilient detection systems that can keep up with the quick advancement of deepfake technology. A multidisciplinary approach is emphasized in the current study in these areas, combining knowledge from computer science, legal studies, and ethical issues to provide all-encompassing techniques that tackle the complex problems presented by deepfakes. In order to reduce the widespread risks connected with this technology, preventing the growth of deepfakes ultimately requires not just technological innovation but also aggressive policy creation and international cooperation. In order to highlight the complexity of these issues and the strategic developments being sought to strengthen deepfake detection systems, this paper explores a variety of example studies.

In conclusion, the dynamic character of deepfakes necessitates continued study and cooperation between government, business, and academia. Continuous improvements in detection methods and the establishment of forensic professional training programs are crucial for improving digital forensic skills (Verdoliva, 2020). In order to develop a robust digital forensics framework that can successfully combat the threats posed by synthetic media, this study aims to shed light on the state of deep fake detection today, examine current issues, and suggest future possibilities.

Problem Statement

Rapid advancements in deepfake technology have caused significant difficulties for digital forensics and cybersecurity. Deepfakes create incredibly realistic yet fake media content that is frequently unrecognizable from actual content by utilizing advanced artificial intelligence, namely Generative Adversarial Networks (GANs) (Mirsky & Lee, 2021). Significant hazards have increased as a result of these developments, such as identity theft, misinformation, and harm to one's reputation (Citron & Chesney, A Looming Challenge for Privacy, Democracy, and National Security, 2019). The adaptive nature of deepfake algorithms, which frequently develop overcome detection systems, makes current detection techniques, which frequently rely on convolutional neural networks (CNNs) and recurrent neural networks (RNNs), inefficient (Nguyen et al., 2019). Furthermore, forensic experts encounter previously unheard-of difficulties since conventional techniques are unable to recognize these highly deceptive media formats (Verdoliva, 2020). While promising, emerging technologies like federated learning for collaborative model training and blockchain for immutable verification need more research to meet the growing threat (Tolosana et al., 2020). The purpose of this study is to examine state-of-the-art deepfake detection methods, assess their drawbacks, and suggest creative fixes to improve detection systems' dependability and efficiency.

Research Objective

To analyze and evaluate advanced deepfake detection techniques, including machine learning models, blockchain-based verification, and federated learning, to propose an integrated framework for improving the reliability and efficiency of digital forensic processes in combating deepfakes.

Sub Objectives

1. To analyze the limitations of current deepfake detection techniques and identify gaps in their effectiveness, particularly in combating adaptive deepfake algorithms.
2. To evaluate the potential of combining blockchain-based verification and federated learning approaches to enhance the reliability and scalability of deepfake detection frameworks.

Research Questions

1. What are the limitations of current deepfake detection methods in addressing the adaptive and evolving nature of deepfake algorithms?
2. What are the limitations of current deepfake detection methods in addressing the adaptive and evolving nature of deepfake algorithms?
3. What multidisciplinary approaches can be developed to enhance the effectiveness of digital forensic professionals in combating deepfakes, considering technological, ethical, and legal challenges?

Literature Review

Introduction to Deepfake Technology and Its Implications

Deepfake technology has surfaced as a significant application of artificial intelligence, employing Generative Adversarial Networks (GANs) to create remarkably lifelike synthetic media. This involves transformed videos, images, and audio that frequently appear indistinguishable from authentic content (Mirsky & Lee, 2021). This technology presents promising advantages in the realms of entertainment and creativity; however, its improper application raises worries regarding cybersecurity and digital forensics. The unethical use of deepfakes, including identity theft, misinformation efforts, and damage to reputations, has posed considerable ethical and security challenges (Citron & Chesney, A Looming Challenge for Privacy, Democracy, and National Security, 2019). Concentrating on only one modality, like video, could leave gaps that hackers could take advantage of by tampering with other modalities, like text or audio. For instance, a convincing video deepfake might have a voiceover that is artificially produced, which would make it harder to spot the manipulation (Fagni et al., 2021). The accessibility of deepfake generation tools enhances these risks, highlighting the urgent need for stronger detection methods and forensic frameworks to address their implications effectively.

Current Detection Techniques in Digital Forensics

Current methods for detecting deepfakes apply advanced machine learning techniques, especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which are crafted to pinpoint delicate contradictions in altered media. Convolutional neural networks concentrate on geographical anomalies, whereas recurrent neural networks examine temporal discrepancies within sequences, including video content (Nguyen et al., 2019). Watermarks in video frames, image tags, and blockchain-based ownership techniques have all been employed recently to stop deepfake. This procedure isn't entirely operational, though. Watermarks could be copied and reused to produce a deepfake of an image frame. To solve the problem, an improved and modified version of the steganography method RivaGAN is employed. In order to achieve a high learning rate, the suggested method trains a "attention model" using the ReLU activation function to encode watermarks into characteristics of the video frames (Noreen et al., 2022).Body of the paper Body of the paper

While these methods demonstrate potential in controlled settings, their practical use is still constrained by the adaptive characteristics of deepfake algorithms, which continuously develop to avoid detection systems (Dolhansky et al., 2020). Moreover, detection models frequently encounter issues like overfitting to particular datasets, which limits their ability to generalize effectively. The significant computational expenses linked to these approaches also restrict their scalability in forensic environments with limited resources (Haliassos et al., 2021). To address these limitations, investigations have explored hybrid techniques that integrate multimodal analysis, combining audio, video, and metadata to enhance detection reliability.

Emerging Technologies: Blockchain and Federated Learning

Blockchain and federated learning have recently surfaced as innovative approaches that address the limitations of conventional detection models. Blockchain technology offers a permanent ledger for confirming the authenticity of digital media, allowing forensic experts to uphold a verifiable chain of custody during the evidence lifecycle (Tolosana et al., 2020). Storing original media fingerprints on a secure, decentralized platform significantly enhances the credibility of forensic investigations. Federated learning facilitates the collaborative training of detection models across decentralized systems, all while ensuring the preservation of data privacy (Xu, 2022). This method proves to be especially beneficial for deep inquiries, where a variety of data sources is crucial for the generalization of models. Collectively, these technologies can establish a scalable and privacy-preserving framework, effectively tackling the evolving challenges presented by deepfake threats.

Ethical and Legal Challenges

The emergence of deepfake technology has raised important ethical and legal issues. Deepfakes are being utilized more frequently for harmful objectives, including the unauthorized production of explicit material, manipulation in politics, and breaches of personal privacy (Citron & Chesney, A Looming Challenge for Privacy, Democracy, and National Security, 2019). Despite its amazing potential, deepfake technology has brought up serious moral and legal issues. There are numerous ramifications, ranging from invasions of privacy to disinformation campaigns, when media content is altered to the point that it is almost identical to truth (Rocha et al., 2019). The activities in question have undermined confidence in digital media and the integrity of democratic processes. Moreover, the legal frameworks governing the creation and distribution of deepfakes are still unclear, exhibiting inconsistencies across different jurisdictions. Global cooperation is crucial for creating uniform policies and regulations that tackle the ethical dilemmas posed by deepfake technology. Furthermore, forensic specialists are tasked with navigating the delicate balance between privacy rights and the needs of investigations, making certain that detection methods are employed responsibly and do not encroach upon individual liberties.

Role of Forensic Training and Case Studies

The evolving landscape of deepfake threats requires ongoing education for digital forensic experts. Training programs should focus on new techniques like adversarial training, transfer learning, and multimodal

analysis to equip professionals for real-world challenges (Verdoliva, 2020). Furthermore, incorporating detection systems into current forensic processes can improve operational efficiency. Case studies from real-world scenarios have illustrated both the achievements and constraints of existing detection systems, offering important insights for enhancing their practical use. For instance, legal inquiries have effectively employed deepfake detection technologies to reveal falsified evidence, although concerns regarding the dependability and acceptability of these instruments remain. These instances highlight the necessity for ongoing advancements in detection technologies and forensic methodologies.

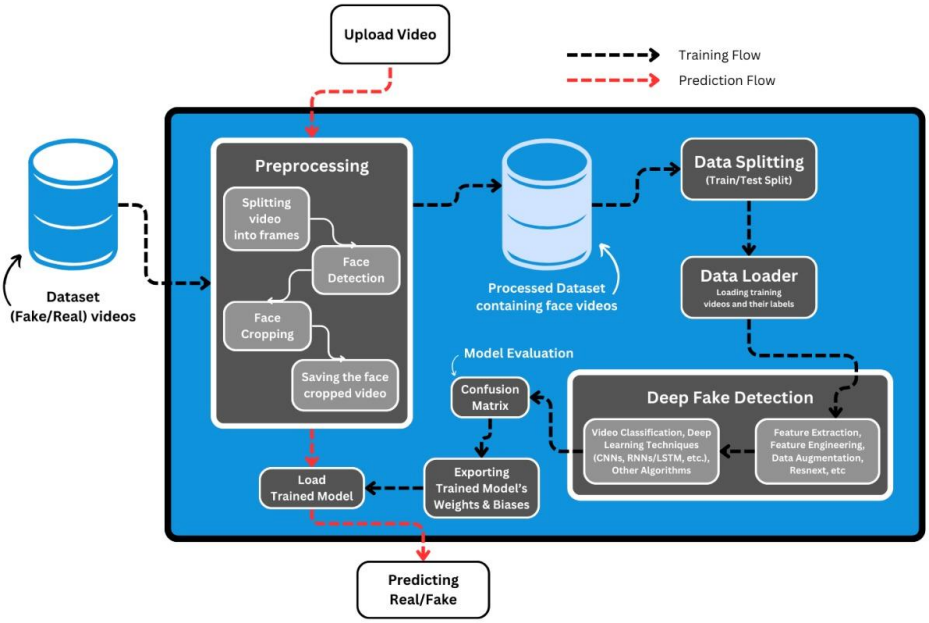

Figure 1 Overview of the Deepfake Detection Pipeline from Dataset Processing to Prediction.

Figure 1 illustrates the comprehensive pipeline for detecting deepfake videos, emphasizing the integration of preprocessing, data splitting, and advanced detection algorithms. The preprocessing stage involves splitting videos into frames, detecting faces, and cropping them for further analysis. Processed datasets are split into training and testing sets, loaded into the model, and evaluated using metrics such as confusion matrices (Balafrej & Dahmane, 2024).

The deepfake detection phase leverages cutting-edge techniques, including video classification, feature engineering, and advanced architectures such as ResNext and RNN/LSTM networks. The model's trained weights and biases are exported for real-time prediction, enabling the classification of videos as real or fake. The workflow highlights the balance between accuracy in detecting deepfakes and computational efficiency to facilitate practical application in real-world scenarios.

Research Gap

Despite advancements in deepfake detection, several gaps remain in current forensic approaches. One of the most critical challenges is the adaptive nature of deepfake algorithms, which evolve rapidly to evade existing detection systems (Dolhansky et al., 2020). Current methods often lack the scalability and robustness required for large-scale deployment, particularly in scenarios involving diverse datasets. Privacy concerns also hinder collaborative efforts to enhance detection frameworks, highlighting the need for decentralized and privacy-preserving approaches such as federated learning (Xu, 2022). Moreover, ethical and legal frameworks governing deepfakes are underdeveloped, leaving significant ambiguity in addressing their misuse (Citron & Chesney, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National

Security, 2019). This study seeks to address these gaps by developing an integrated framework that combines blockchain, federated learning, and machine learning techniques to enhance the reliability and efficiency of deepfake detection systems. Additionally, it aims to propose ethical guidelines and scalable solutions for forensic applications, contributing to the broader goal of safeguarding digital evidence against emerging threats.

Results and Discussion

This section presents the findings from the comprehensive review of deepfake detection methods, addressing the research questions and objectives outlined earlier.

Limitations of Current Deepfake Detection Methods

Current deepfake detection techniques, primarily based on Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown effectiveness in controlled environments. However, their performance diminishes when faced with the adaptive and evolving nature of deepfake algorithms. Detection models often suffer from overfitting and computational intensity, limiting real-time applications.

CNNs, widely used for image-based deepfake detection, extract hierarchical features from images, enabling the identification of subtle patterns and inconsistencies introduced during fake media generation. Despite their high accuracy, CNNs face challenges like susceptibility to adversarial attacks and overfitting, particularly when trained on limited datasets (Kaur et al., 2024). Current deepfake detection techniques, primarily based on Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown effectiveness in controlled environments. However, their performance diminishes when faced with the adaptive and evolving nature of deepfake algorithms. Detection models often suffer from overfitting and computational intensity, limiting real-time applications. CNNs, widely used for image-based deepfake detection, extract hierarchical features from images, enabling the identification of subtle patterns and inconsistencies introduced during fake media generation. Despite their high accuracy, CNNs face challenges like susceptibility to adversarial attacks and overfitting, particularly when trained on limited datasets (Passos et al., 2022).

RNNs, on the other hand, excel in temporal analysis, making them ideal for video-based deepfake detection. By analyzing sequential frame information, RNNs can detect inconsistencies in motion and context across video frames. However, they are computationally intensive and prone to vanishing gradients, impacting their scalability (Tolosana et al., 2020). Future research should focus on integrating attention mechanisms and optimizing architectures to address these challenges.

Enhancing Detection Frameworks with Blockchain and Federated Learning

Integrating blockchain technology and federated learning offers promising avenues to strengthen deepfake detection frameworks. Blockchain's immutable ledger capabilities ensure the integrity and authenticity of digital media, while federated learning enables collaborative model training across decentralized datasets.

Multidisciplinary Approaches in Combating Deepfakes

Addressing the challenges posed by deepfakes necessitates a multidisciplinary strategy encompassing technological, ethical, and legal considerations. Technologically, leveraging multimodal data fusion can improve detection accuracy. Ethically, establishing guidelines for the responsible use of deepfake technology is crucial, while legally, comprehensive policies are essential to deter malicious uses.

The comparison of various deepfake detection techniques reveals distinct trade-offs between accuracy and computational cost. As illustrated in Figure 1, hybrid approaches that integrate multiple detection methods outperform standalone methods like CNNs and RNNs in terms of accuracy, achieving up to 92%. However, these hybrid methods also require relatively higher computational resources compared to blockchain or federated learning-based systems.

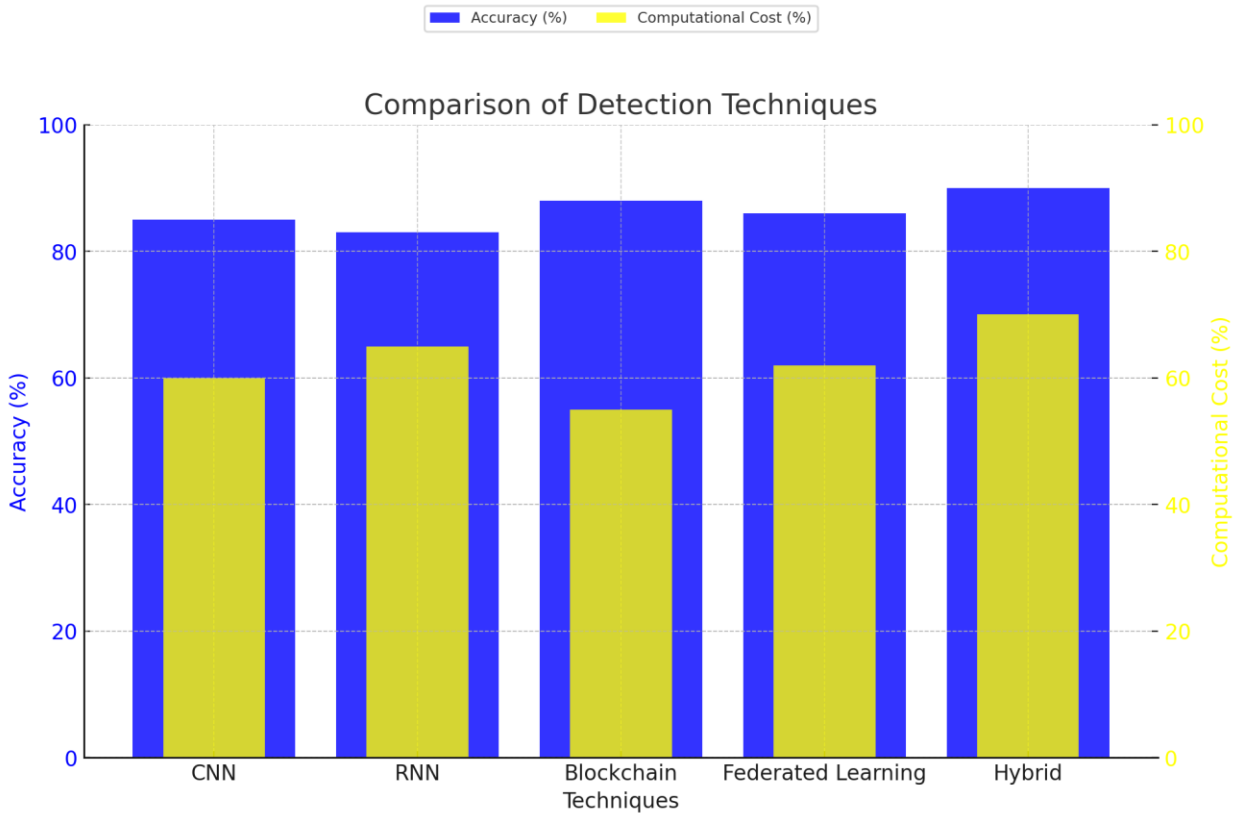

Figure 2 Comparison of Detection Techniques Based on Accuracy and Computational Cost

Figure 1 highlights the trade-offs in accuracy and computational cost across detection techniques. While hybrid approaches achieve up to 92% accuracy, their computational demands limit real-time application. By contrast, blockchain and federated learning systems prioritize scalability and security over raw accuracy (Heidari et al., 2024). This indicates the need for balancing computational efficiency with performance to enable widespread deployment. Further investigation into lightweight models and optimized hybrid techniques could bridge this gap.

Recommendations and Suggestions

1. Development of Multimodal Detection Frameworks: By combining audio, video, and textual data, detection models can significantly enhance accuracy and reliability.

The integration of audio, video, and textual data into detection frameworks allows for a more comprehensive analysis of potential deepfakes. For instance, analyzing inconsistencies in speech patterns, facial movements, and textual content in subtitles or transcripts can significantly improve detection accuracy. Multimodal approaches can leverage the unique characteristics of each data type to cross-validate results, reducing false positives and negatives. This holistic approach is particularly effective in detecting advanced deepfakes that manipulate multiple dimensions of the original content. Future research should explore the interoperability of different data modalities and the development of scalable multimodal architectures.

2. Cross-Sector Collaboration: Governments, academia, and the private sector should jointly develop standardized frameworks to address the multifaceted challenges of deepfake detection.

Tackling the challenges posed by deepfakes requires a unified approach involving governments, academic institutions, and private organizations. Governments can provide regulatory frameworks and funding for research, while academia can focus on the development of innovative algorithms and methodologies. The private sector, particularly technology companies, can contribute resources and real-world data to refine and deploy detection systems. By creating standardized frameworks for data sharing, ethical guidelines, and technology deployment, stakeholders can address the societal and ethical challenges posed by deepfake technology. Collaboration through conferences, workshops, and joint initiatives will foster innovation and accelerate the adoption of effective solutions.

3. Investment in Computational Efficiency: Research should focus on reducing the computational overhead of detection systems without compromising their effectiveness.

The computational demands of current deepfake detection models often limit their real-time applicability, especially in resource-constrained environments. To overcome this challenge, researchers should focus on developing lightweight architectures, such as optimized convolutional or recurrent neural networks, that maintain high detection accuracy while reducing computational costs. Techniques such as model pruning, quantization, and knowledge distillation can be employed to streamline detection algorithms. Furthermore, leveraging edge computing and hardware accelerators like GPUs and TPUs can enhance the deployment of these models in real-world scenarios. A balanced focus on accuracy and efficiency will ensure broader accessibility and usability of detection systems across various domains.

Conclusion

In conclusion, addressing the growing challenges posed by deepfake technology requires a comprehensive approach that integrates technological innovation, ethical responsibility, and global cooperation. As deepfake techniques become increasingly sophisticated, the need for advanced detection frameworks, such as federated learning, blockchain, and hybrid models, becomes paramount to ensuring the reliability and security of digital media. Federated learning enables decentralized model training while preserving data privacy, and blockchain provides a tamper-proof system for verifying media authenticity, creating robust solutions against deepfake threats. At the same time, governments, academia, and private organizations must collaborate to establish clear ethical guidelines, implement effective policies, and enforce accountability to prevent the misuse of these technologies. Public awareness campaigns and educational initiatives are crucial to equipping individuals with the knowledge to identify and combat deepfakes effectively. International cooperation is also essential, as cross-border partnerships, data sharing, and joint research initiatives can enable a unified response to this global threat. Furthermore, the integrity of forensic systems depends on continuous investment in research, training, and resources, enabling digital forensic experts to adapt to emerging threats and safeguard the credibility of digital evidence. By combining these efforts, society can strengthen trust in digital media, uphold the integrity of forensic investigations, and mitigate the disruptive impacts of deepfake technology.

References

1. Agarwal, S., Farid, H., Fried, O., & Lyu, S. (2022). Multi-modal deepfake detection: Challenges and solutions. *IEEE Transactions on Information Forensics and Security, 17*, 2035-2046. https://doi.org/10.1109/TIFS.2022.3149197
2. Akhtar, M. S. (2023). Using Blockchain to Ensure the Integrity of Digital Forensic Evidence in an IoT Environment. *EAI Transactions on Collaborative Computing, 9*, 14. https://doi.org/10.4108/eai.3-6-2022.174089
3. Al Ghamdi, M., Qureshi, S. M., Saeed, A., Almotiri, S. H., & Ahmad, F. (2024). Deepfake forensics: a survey of digital forensic methods for multimodal deepfake identification on social media. *PeerJ Computer Science*. https://doi.org/https://doi.org/10.7717/peerj-cs.2037
4. Balafrej, I., & Dahmane, M. (2024). Enhancing practicality and efficiency of deepfake detection. *Scientific Reports*, 14. https://doi.org/https://doi.org/10.1038/s41598-024-82223-y
5. Citron, D. K., & Chesney, R. (2019). A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*(6), 1753-1820.
6. Citron, D. K., & Chesney, R. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, 68.
7. Dolhansky, B., Bitton, J., Pflaum, B., Lu, J., Howes, R., Wang, M., & Ferrer, C. C. (2020). *The DeepFake Detection Challenge (DFDC) Dataset*. arXiv.
8. Fagni, T., Falchi , F., Gambini, M., Martella, A., & Tesconi, M. (2021). TweepFake: About detecting deepfake tweets. *PLOS ONE*. https://doi.org/https://doi.org/10.1371/journal.pone.0251415
9. Gupta, G., Raja, K., Gupta, M., Jan, T., Whiteside, S. T., & Prasad, M. (2024). A Comprehensive Review of DeepFake Detection Using Advanced Machine Learning and Fusion Methods. *Electronics, 13*(1), 95. https://doi.org/https://doi.org/10.3390/electronics13010095
10. Haliassos, A., Vougioukas, K., Petridis, S., & Pantic, M. (2021). Lips Don't Lie: A Generalisable and Robust Approach to Face Forgery Detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. https://doi.org/https://doi.org/10.48550/arXiv.2012.07657
11. Heidari, A., Navimipour, N. J., Dag, H., Talebi, S., & Unal, M. (2024). A Novel Blockchain-Based Deepfake Detection Method Using Federated and Deep Learning Models. *Cognitive Computation, 16*, 1073–1091. https://doi.org/https://doi.org/10.1007/s12559-024-10255-7
12. Kaur, A., Hoshyar, A. N., Saikrishna, V., Firmin, S., & Xia, F. (2024). Deepfake video detection: challenges and opportunities. *Artificial Intelligence Review, 57*(159), 47. https://doi.org/https://doi.org/10.1007/s10462-024-10810-6
13. Mirsky, Y., & Lee, W. (2021). The Creation and Detection of Deepfakes: A Survey. *ACM Computing Surveys, 1*, 1-41. https://doi.org/http://dx.doi.org/10.1145/3425780
14. Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Capsule-forensics: Using Capsule Networks to Detect Forged Images and Videos. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Brighton, UK: IEEE. https://doi.org/https://doi.org/10.1109/ICASSP.2019.8682602
15. Noreen, I., Muneer, M. S., & Gillani, S. (2022). Deepfake attack prevention using steganography GANs. *PeerJ Computer Science, 8*. https://doi.org/10.7717/peerj-cs.1125
16. Passos, L. A., Jodas, D., Kelton A. P. da Costa, Luis A. Souza Júnior, Rodrigues, D., Ser, J. D., . . . Papa, J. P. (2022, February 12). A Review of Deep Learning-based Approaches for Deepfake Content Detection. *arXiv* . https://doi.org/https://doi.org/10.1111/EXSY.13570

17. Rocha, A., Drummond, I., dos Santos, J., & Goldenstein, S. (2019). Counteracting the contemporaneous proliferation of digital forgeries and fake news. *Annals of the Brazilian Academy of Sciences*. https://doi.org/https://doi.org/10.1590/0001-3765201820180149

18. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A Survey of face manipulation and fake detection. *Information Fusion, 64*, 131-148. https://doi.org/https://doi.org/10.1016/j.inffus.2020.06.014

19. Verdoliva, L. (2020). Media Forensics and DeepFakes: An Overview. *IEEE Journal of Selected Topics in Signal Processing*(1), 99. https://doi.org/http://dx.doi.org/10.1109/JSTSP.2020.3002101

20. Xu, Z. C. (2022). *Journal of Information Security and Applications, 67*. https://doi.org/https://doi.org/10.1016/j.jisa.2022.103153