JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Deepfake Technology and the Detection Methods: A Review

[1]**Dr. Shwetambari Borade, [2]Aabha Wagh, [3]Vaidehi Salvi, [4]Parth Savla, [5]Drashti Nagda**

[1]Professor, [2]B.E. Cyber Security Student, [3]B.E. Cyber Security Student, [4]B.E. Cyber Security Student, [5]B.E. Cyber Security Student,

[1]Cyber Security,
[1]Shah & Anchor Kutchhi Engineering College Mumbai, India

Abstract : In recent years, there has been a pronounced escalation in cyber criminal activities, closely paralleling the rapid advancements in AI technology. The emergence of deepfake technology represents a formidable challenge to the maintenance of digital information integrity. This investigation underscores the critical exigency for dependable techniques in deepfake detection, imperative for stemming the dissemination of manipulated media. To preclude the proliferation of erroneous information, bolster credibility, and uphold public trust, the implementation of detection systems is indispensable. However, their deployment necessitates a substantial allocation of computational resources, giving rise to ethical considerations concerning consent and privacy. This study underscores the paramount importance of striking a judicious equilibrium to ensure both effective and ethically upright deepfake detection in the digital milieu. It achieves this through a comprehensive scrutiny of the advantages and drawbacks, including the preservation of content integrity and the attendant resource requisites.

IndexTerms - Deepfake Technology, Deepfake Detection, Convolutional Neural Networks, Artificial Intelligence bots, Machine Learning Artificial Intelligence, Face Morphing Detection, Deep Learning.

I. Introduction

The exponential progress witnessed in the realms of Artificial Intelligence (AI) and machine learning has ushered in an era dominated by the advent of deepfake AI bots.[1] This development represents a substantial menace to various facets of society, including security and the foundation of trust within the digital domain. Consequently, this surge has given rise to a slew of issues such as deceit, dissemination of false information, and the potential for harm, thereby necessitating an immediate and concerted effort to devise robust mechanisms capable of identifying and mitigating the perils posed by these highly sophisticated AI-driven manipulations.[1]

The implementation of deepfake detection systems holds the promise of not only reinstating faith in digital media but also of ameliorating the dissemination of disinformation, safeguarding both personal and organizational reputations, and ensuring the principled deployment of AI technology. Moreover, they play a pivotal role in fortifying digital identities, bolstering cybersecurity measures, propelling the frontiers of AI and machine learning research, lending crucial support to regulatory endeavors, and endowing both individuals and institutions with the means to effectively navigate this digital landscape. As a cumulative effect, these systems contribute tangibly towards cultivating a more secure and resilient digital ecosystem.

By affording accessible tools and resources, these detection systems serve as an empowering force, endowing individuals, enterprises, and academic institutions with the capacity to adopt a proactive stance in countering the potential threats posed by deepfakes. In doing so, they engender a climate of preparedness, thereby fostering a collective resilience against the insidious influence of fabricated media.

II. Literature Survey

A thorough overview of deepfake technology and its detection techniques is given by the literature review, which also covers its historical evolution, uses, societal, legal, and ethical ramifications, as well as new trends. It looks at the literature on deepfake content detection, including multimodal analysis, machine learning algorithms, and forensic methods. The survey assesses these techniques' effectiveness and application and emphasizes the value of interdisciplinary cooperation in the creation of increasingly complex detection procedures. It seeks to add to the expanding corpus of knowledge and lay the groundwork for further initiatives in the battle against false information and digital credibility. It attempts to provide a comprehensive knowledge of its origins and ongoing efforts to create reliable detection methods by integrating existing research and detecting upcoming patterns. The survey's overall goal is to provide readers a thorough grasp of how Deepfake technology is developing

Table 1. survey of some papers that were significant for the research

SR. NO.	Paper Citation	Problem Statement/ Approach	Proposed Solution	Datasets Used/Mentioned	Conclusion
1)	[2] 2023 IEEE	Paper highlights how important it is to develop deepfake social media message detection algorithms that are reliable and accurate.	The research makes use of word embeddings and a deep learning model to classify tweets as either human- or bot-generated. Popular methods for extracting features: Techniques for word embedding with Tf and TF-IDF: There is use of FastText and its subwords..	The research makes use of word embeddings and a deep learning model to classify tweets as either human- or bot-generated. Popular methods for extracting features: Techniques for word embedding with Tf and TF-IDF: There is use of FastText and its subwords.	With a 0.93 accuracy score in correctly identifying deepfake text, the suggested method showed encouraging results by utilizing a combination of CNN and FastText approaches.
2)	[3] 2023 IEEE	The paper offers the concept of cyber vaccination as a means of providing resistance to deepfakes.	The proposed immune system is made up of a vaccinator to induce immunity and a neutraliser to recover face content.	FaceForensics++	Deepfakes that are mask-dependent are easier to handle, with the algorithm adapting to head orientations and occlusive objects. However, drawbacks include mismatched make-up colors and imprecise skin tones. The study also investigates the effect of posture alterations on immunity
3)	[4] 2023 IEEE	The study offers an improved deep-CNN architecture with intermediate accuracy and good generalizability for deepfake image detection.	The suggested model is a binary classification model that uses CNN and D-CNN convolutional layers to extract deep features from input pictures. The proposed architecture consists of both pooling and convolution layers, plus a flattened layer that transforms feature maps into a one-dimensional array.	Real photos from the CelebA and FFHQ picture libraries comprised the dataset. Every one has five thousand pictures. A total of 1000 images from the GDWCT, AttGAN, STARGAN, StyleGAN, and StyleGAN2 datasets are used for deepfake identification.	The accuracy of the suggested model is 99.17% for StyleGAN and 98.33% for AttGAN.
4)	[5] 2023 IEEE	The research addresses the issue of masks	The study's Deepfake Face Mask Dataset	Deepfake Face Mask Dataset (DFFMD)	The study results detect face-mask-Deepfakes with

		obscuring defining face features, making phony videos even more difficult to spot, emphasizing the importance of effective Deepfake detection techniques	(DFFMD) is built using a unique Inception-ResNet-v2 that includes batch normalization, feature-based residual connections, and preprocessing phases.		99.81% accuracy when compared to the state-of-the-art techniques, InceptionResNetV2 and VGG19.
5)	[6] 2023 IEEE	The study provides a method based on decentralized blockchain technology to protect image and video integrity from identity theft using a Deepfake Analyzer.	Blockchain, encryption, media filtration, convolutional neural networks, consensus algorithms, and SHA-256 hashing algorithms are all included in the proposed approach.	FFPMS, DFDC, publicly available dataset from GitHub	With the use of this technology, almost flawless films can be created and the faces of two people can be superimposed. The essay offers perceptive ideas about how blockchain technology might be applied to safeguard the integrity of images and videos and lessen the dissemination of false information in the media. By guarding against identity theft, it provides an additional layer of security to video and image fidelity.
6)	[7] 2023 IEEE	The article addresses the use of deep learning technology to make fake videos for both positive and negative reasons using generative adversarial networks (GAN).	An analysis is presented of the differences in performance between several deepfake video detection models in deep learning, including CNN models like ResNet, VGG16, and Efficientnet.	DFDC	The article discusses the use of Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN) for high-precision deepfake detection. When the accuracy of the model was checked, it scored 97.6%.
7)	[8] 2023 IEEE	The paper compares four deep-learning models that can help in deepfake detection.	The article covers the use of transfer learning with imagenet weights for feature extraction, but due to low accuracy, the author decided to train the entire architecture from scratch. The input image size of each model was changed to 224x224x3, and the default top layer	FaceForensics++, Deepfakes, Face2Face, FaceSwap, NeuralTextures	The researchers assessed the performance of four models trained on the FaceForensics++ dataset and discovered that various models perform differently on different types of deepfakes photos. Deepfakes has the best feature identification and

			was replaced with a fully linked layer. Precision can be used to establish the relevance of classification findings, whereas recall seeks to quantify the fraction of true positives detected properly for a given class. F1-score can be used to calculate the total of both measures.		accuracy, whereas XceptionNet has the least variation in accuracy.

A. Findings of Literature Survey

Innovations like dual-scale receptive fields, eye blink pattern analysis, meta-learning, hybrid face forensics frameworks, dual attention mechanisms, cyber vaccination concepts, steganalysis networks, blockchain-based defenses, and specialized datasets like DFFMD are among the best ways to solve the problem.

- Cyber vaccination against deepfakes is a novel approach that combines a neutralizer and vaccine to provide immunity and content recovery
  Offering a proactive defensive mechanism against deepfake threats is an advantage over current methods.
  It is suggested that in order to strengthen defenses against potential deepfake assaults, investigate the incorporation of cyber vaccine concepts into current security frameworks.[3]
- Detecting face-mask-Deepfakes with a high accuracy of 99.81% is an advantage over the current system.
  It is advised to incorporate DFFMD into current datasets in order to enhance the precision of deepfake identification, particularly in situations when face masks are used.[5]
- Leveraging blockchain technology for defending image and video integrity.
  Offers an extra degree of protection for integrity and security.[6]
- Applying deep learning for deepfake detection, with a focus on face detection architecture and dataset size.
  Benefit Beyond Current solutions is that it utilizes a semi-supervised GAN architecture for enhanced detection.
  To improve accuracy and efficiency, integrate the characteristics of deep learning in computer vision, as described in the studied literature, into the current detection models.[8]
- Deepfake picture detection using a dual-scale large receptive field network is an innovative approach.
  Advantage Over Existing solutions is that it achieves a 64.9% reduction in model size without losing performance on benchmark datasets.
  It is recommended that dual-scale receptive fields be incorporated into the current deepfake detection algorithms in order to optimize performance and minimize computing demands.[10]
- Using GANs and DeepVision for Deepfake detection analysis of human eye blinking patterns
  It is advised to incorporate eye blink pattern analysis into current detection frameworks in order to improve accuracy—particularly when dealing with video content.
- Using meta-learning to create a generalized model that can handle domains that have never been explored before is known as Meta Deepfake Detection (MDD).
  One advantage over current approaches is the ability to enhance model representations by applying varying weights to facial photographs from different domains.[12]
- Developinga convolutional neural network-based hybrid face forensics framework.
  It offers improved accuracy and durability at varying video compression rates.
  It is recommended that hybrid face forensics frameworks be incorporated into current systems in order to improve the identification of manipulation, particularly in a variety of compression settings.
- The Deepfake Face Mask Dataset (DFFMD) represents an innovative approach for detecting face-mask deepfakes. It is based on the Inception-ResNet-v2 model.
- Introducing a steganalysis network to identify landmarks and pixel-wise residual-noise traces in deepfake detection.
  Advantage Over Existing solutions is that it demonstrates efficiency and stability on various deepfake kinds.
  It is suggested that in order to improve pixel-wise detection capabilities, steganalysis approaches be incorporated into the current deepfake detection models.[15]

By incorporating these developments into current technologies, we may strengthen and improve our defenses against deepfake technology threats in a number of different fields

-DATASETS

FaceForensics++[9] is the most commonly used, discussed, or compared dataset discovered in this literature review along with DFDC[1][24] and Celeb-DF(V2,V1).[2][16]

III. Summary

Finally, this extensive literature analysis sheds light on the fundamental consequences of deepfake technology, both in terms of its manipulation potential and the critical requirement for effective detection approaches. Rapid growth of Deepfake technology has ushered in an era in which distinguishing reality from manufactured content has become increasingly difficult. This technology has the potential to disrupt several parts of society, including politics, media, and personal relationships, as it develops. Deepfake production techniques have become more advanced over time, incorporating not only facial modifications but also voice synthesis and even full-body movements, according to the review. The introduction of such multimodal deepfakes adds to the difficulty of detection, emphasizing the significance of holistic detection methodologies.

Despite tremendous development in machine learning-based detection systems, the cat-and-mouse game between makers and detectors continues. Although adversarial training and the creation of more diverse and vast datasets have yielded promising advances, the ongoing arms race needs continued innovation in this discipline. Furthermore, the paper underlines explainable rising significance of AI in delivering interpretable insights into detection processes, hence increasing confidence and responsibility.

Furthermore, the ethical and legal ramifications of deepfakes cannot be overstated. They endanger the integrity of information and trust in digital content by blurring the lines between truth and untruth. As a result, policy and regulatory frameworks must evolve to meet these issues, combining free expression with the need to counteract misinformation.

IV. Conclusion

These research papers collectively highlight the importance of deepfake detection and offer diverse approaches to address the issue. While some papers focus on text-based detection using deep learning models and word embeddings, others propose innovative methods such as cyber vaccination, deep-CNN architectures, and blockchain technology. Each approach demonstrates promising results, with high accuracy scores in identifying deepfakes in various forms, including text, images, and videos. The papers emphasize the significance of dataset creation, feature extraction, and model selection in achieving accurate deepfake detection. However, it is essential to consider the limitations and variations of each approach, as they may perform differently depending on the specific type of deepfake and dataset utilized. Overall, these papers contribute valuable insights and advancements in combating the spread of deepfakes by offering a range of effective techniques for detecting and identifying manipulated content.

V. References

[1] S. Yadav, S. Bommareddy and D. K. Vishwakarma, "Robust and Generalized DeepFake Detection," 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2022, pp. 1-6, doi: 10.1109/ICCCNT54827.2022.9984553.

[2] S. Sadiq, T. Aljrees and S. Ullah, "Deepfake Detection on Social Media: Leveraging Deep Learning and FastText Embeddings for Identifying Machine-Generated Tweets," in IEEE Access, vol. 11, pp. 95008-95021, 2023, doi: 10.1109/ACCESS.2023.3308515.

[3] C. -C. Chang, H. H. Nguyen, J. Yamagishi and I. Echizen, "Cyber Vaccine for Deepfake Immunity," in IEEE Access, vol. 11, pp. 105027-105039, 2023, doi: 10.1109/ACCESS.2023.3311461.

[4] Y. Patel et al., "An Improved Dense CNN Architecture for Deepfake Image Detection," in IEEE Access, vol. 11, pp. 22081-22095, 2023, doi: 10.1109/ACCESS.2023.3251417.

[5] N. M. Alnaim, Z. M. Almutairi, M. S. Alsuwat, H. H. Alalawi, A. Alshobaili and F. S. Alenezi, "DFFMD: A Deepfake Face Mask Dataset for Infectious Disease Era With Deepfake Detection Algorithms," in IEEE Access, vol. 11, pp. 16711-16722, 2023, doi: 10.1109/ACCESS.2023.3246661.

[6] J. A. Costales, S. Shiromani and M. Devaraj, "The Impact of Blockchain Technology to Protect Image and Video Integrity from Identity Theft using Deepfake Analyzer," 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 2023, pp. 730-733, doi: 10.1109/ICIDCA56705.2023.10099668.

[7] A. Das, K. S. A. Viji and L. Sebastian, "A Survey on Deepfake Video Detection Techniques Using Deep Learning," 2022 Second International Conference on Next Generation Intelligent Systems (ICNGIS), Kottayam, India, 2022, pp. 1-4, doi: 10.1109/ICNGIS54955.2022.10079802.

[8] N. Khatri, V. Borar and R. Garg, "A Comparative Study: Deepfake Detection Using Deep-learning," 2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2023, pp. 1-5, doi: 10.1109/Confluence56041.2023.10048888.

[9] A. H. Khalifa, N. A. Zaher, A. S. Abdallah and M. W. Fakhr, "Convolutional Neural Network Based on Diverse Gabor Filters for Deepfake Recognition," in IEEE Access, vol. 10, pp. 22678-22686, 2022, doi: 10.1109/ACCESS.2022.3152029.

[10] Y. -X. Luo and J. -L. Chen, "Dual Attention Network Approaches to Face Forgery Video Detection," in IEEE Access, vol. 10, pp. 110754-110760, 2022, doi: 10.1109/ACCESS.2022.3215963.

[11] W. Shahid, Y. Li, D. Staples, G. Amin, S. Hakak and A. Ghorbani, "Are You a Cyborg, Bot or Human?—A Survey on Detecting Fake News Spreaders," in IEEE Access, vol. 10, pp. 27069-27083, 2022, doi: 10.1109/ACCESS.2022.3157724.

[12] V. -N. Tran, S. -G. Kwon, S. -H. Lee, H. -S. Le and K. -R. Kwon, "Generalization of Forgery Detection With Meta Deepfake Detection Model," in IEEE Access, vol. 11, pp. 535-546, 2023, doi: 10.1109/ACCESS.2022.3232290.

[13] A. Malik, M. Kuribayashi, S. M. Abdullahi and A. N. Khan, "DeepFake Detection for Human Face Images and Videos: A Survey," in IEEE Access, vol. 10, pp. 18757-18775, 2022, doi: 10.1109/ACCESS.2022.3151186.

[14] T. Dar, A. Javed, S. Bourouis, H. S. Hussein and H. Alshazly, "Efficient-SwishNet Based System for Facial Emotion Recognition," in IEEE Access, vol. 10, pp. 71311-71328, 2022, doi: 10.1109/ACCESS.2022.3188730.

[15] J. Kang, S. -K. Ji, S. Lee, D. Jang and J. -U. Hou, "Detection Enhancement for Various Deepfake Types Based on Residual Noise and Manipulation Traces," in IEEE Access, vol. 10, pp. 69031-69040, 2022, doi: 10.1109/ACCESS.2022.3185121.

[16] M. S. Rana, M. N. Nobi, B. Murali and A. H. Sung, "Deepfake Detection: A Systematic Literature Review," in IEEE Access, vol. 10, pp. 25494-25513, 2022, doi: 10.1109/ACCESS.2022.3154404.

[17] D. B. Frolov, D. D. Makhaev and V. V. Shishkarev, "Deepfakes and Information Security Issues," 2022 International Conference on Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS), Saint Petersburg, Russian Federation, 2022, pp. 147-150, doi: 10.1109/ITQMIS56172.2022.9976507.

[18] H. Ilyas, A. Irtaza, A. Javed and K. M. Malik, "Deepfakes Examiner: An End-to-End Deep Learning Model for Deepfakes Videos Detection," 2022 16th International Conference on Open Source Systems and Technologies (ICOSST), Lahore, Pakistan, 2022, pp. 1-6, doi: 10.1109/ICOSST57195.2022.10016871.

[19] J. John and B. V. Sherif, "Comparative Analysis on Different DeepFake Detection Methods and Semi Supervised GAN Architecture for DeepFake Detection," 2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Dharan, Nepal, 2022, pp. 516-521, doi: 10.1109/I-SMAC55078.2022.9987265.

[20] K. Jalui, A. Jagtap, S. Sharma, G. Mary, R. Fernandes and M. Kolhekar, "Synthetic Content Detection in Deepfake Video using Deep Learning," 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2022, pp. 01-05, doi: 10.1109/GCAT55367.2022.9972081.

[21] S. S. Chauhan, N. Jain, S. C. Pandey and A. Chabaque, "Deepfake Detection in Videos and Picture: Analysis of Deep Learning Models and Dataset," 2022 IEEE International Conference on Data Science and Information System (ICDSIS), Hassan, India, 2022, pp. 1-5, doi: 10.1109/ICDSIS55133.2022.9915885.

[22] V. Jolly, M. Telrandhe, A. Kasat, A. Shitole and K. Gawande, "CNN based Deep Learning model for Deepfake Detection," 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), Ravet, India, 2022, pp. 1-5, doi: 10.1109/ASIANCON55314.2022.9908862.

[23] A. Rahman et al., "Short And Low Resolution Deepfake Video Detection Using CNN," 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC), Hyderabad, India, 2022, pp. 259-264, doi: 10.1109/R10-HTC54060.2022.9929719.

[24] E. Kim and S. Cho, "Exposing Fake Faces Through Deep Neural Networks Combining Content and Trace Feature Extractors," in IEEE Access, vol. 9, pp. 123493-123503, 2021, doi: 10.1109/ACCESS.2021.3110859.

[25] M. F. B. Ahmed, M. S. U. Miah, A. Bhowmik and J. B. Sulaiman, "Awareness to Deepfake: A resistance mechanism to Deepfake," 2021 International Congress of Advanced Technology and Engineering (ICOTEN), Taiz, Yemen, 2021, pp. 1-5, doi: 10.1109/ICOTEN52080.2021.9493549.