FMDB *Transactions on Sustainable Computing Systems*

A Comprehensive Review on Deepfake Detection on Social Media Data

Battula Thirumaleshwari Devi[1], Rajasekaran. R[2,*],

[1,2] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
Malashwari.devi@vit.ac.in[1], vitrajkumar@gmail.com[2]

Abstract: Deepfake technology has become a growing concern for society in recent years, as it allows for manipulating text, images, video, and audio. Deepfake technology, which leverages deep learning and artificial intelligence to create realistic yet entirely synthetic media, has garnered significant attention and raised both promises and concerns. The use of deepfakes has been a cause for concern due to their potential to be used for malicious purposes, such as spreading false information, committing fraud, or impersonating individuals. Artificial intelligence plays a crucial role in the development of deepfake technology. Detecting deepfakes is urgently necessary to protect individuals, organizations, and society. Deepfake technology is a double-edged sword. While it offers significant potential benefits, its malicious use poses substantial social risks. Striking a balance between innovation and regulation while emphasizing responsible use and ethical considerations is essential to harness the positive aspects of deepfake technology while mitigating its potential harm.

Keywords: Deepfake; Audio Image; Video Text; Artificial Intelligence (AI); Deepfake Technology; Impersonating Individuals; Deep Learning-Based Facial Recognition; Generalization of The Model.

Received on: 19/10/2022, **Revised on:** 07/12/2022, **Accepted on:** 11/01/2023, **Published on:** 17/02/2023

Cite as: B. T. Devi, and R. Rajasekaran, "A Comprehensive Review on Deepfake Detection on Social Media Data," *FMDB Transactions on Sustainable Computing Systems*., vol. 1, no. 1, pp. 11–20, 2023.

1. Introduction

Deepfake technology, powered by advanced artificial intelligence (AI) techniques, has gained notoriety for its ability to create hyper-realistic manipulated videos and audio recordings, often blurring the line between reality and fabrication. These convincing forgeries have raised significant concerns regarding their potential to deceive, manipulate, and spread misinformation. As a response to this emerging challenge, deepfake detection using AI has emerged as a critical field of research and development. Deep learning is where the phrase "deep fake" originates from, which is a subset of machine learning that uses neural networks to handle enormous amounts of data [1]. Deepfakes have raised concerns about the potential for malicious usages, such as distributing false information, impersonating someone, or committing fraud [2]; [4]. Deepfake is a class of artificial intelligence techniques that create convincing synthetic media, such as text, image, video, and audio.

2. Review of Literature

Sinaga et al. [13] have improved video quality by detecting human biometric elements and using CNNs for categorizing and analyzing two-dimensional data. False photos are classified using features, deep learning, and a machine learning model. YouTube clips are chosen at random. Deep learning-based facial recognition CNN's identification of a real or fake label's facial similarity.

Khatri et al. [14] have compared four deep learning models, VGG16, MobileNetV2, XceptionNet, and InceptionV3, and trained them using the FaceForensics++ dataset. They found that the Deepfakes dataset yields the best feature identification and model

*Corresponding author.

correctness results, followed by Face2Face, FaceSwap, and NeuralTextures. NeuralTextures yield the worst results, while XceptionNet has the lowest degree of accuracy variation. Patel et al. [15] have introduced enhanced deep-CNN (D-CNN) architecture for deep-fake detection with good accuracy and generalizability. Training the model with several images has improved the generalizability by binary-cross entropy, and Adam optimizer boosts D-CNN model learning. Seven reconstruction challenge datasets of 5,000 deep-fakes and 10,000 actual pictures were examined. AttGAN [Facial Attribute Editing by Only Changing What You Want (AttGAN)] has a 98.33% accuracy. GDWCT: 99.33%, 95.33% in StyleGAN, 94.67% in StyleGAN2, and 99.17% in StarGAN, which can learn domain mappings from actual and deep-fake pictures, proving its experimental feasibility. Tran et al. [16] offered a face forgery detection approach called the Generalization Deepfake Detector. This method aims to boost the generalization of the model (GDD). It can rapidly solve new, previously unknown domains without needing a model.

Chintha et al. [17] have presented a digital forensic technique for audio spoof and visual deepfake detection, using bidirectional recurrent structures, entropy-based cost functions, and convolutional latent representations. These techniques extract semantically rich information from recordings, identify deepfake renditions' spatial and temporal characteristics, and establish new benchmarks on FaceForensics++, Celeb-DF, and ASVSpoof 2019 audio datasets. Khochare et al. [18] used a Fake or Real (FoR) dataset and feature-based and image-based methods to classify audio as false or genuine. The feature-based technique converts audio input into spectral properties, while the image-based technique transforms audio samples into Mel spectrograms and uses deep learning algorithms like Temporal Convolutional Network (TCN) and Spatial Transformer Network (STN), with 92 percent test accuracy. This solution proposed a model for audio deepfake classification with accuracy equivalent to VGG16, XceptionNet, etc. Chen et al. [19] have presented a neural network system for spoofing countermeasures, focusing on generalization and robust feature embeddings. The system is tested on the ASVspoof 2019 dataset, noisy version, and call centre spoofing attack dataset.

The EER is reduced to 1.26% with further enhancements. Jung et al. [20] have proposed a novel method to identify Deepfakes in generative adversarial network models using DeepVision. This method will examine voluntary blinking patterns based on physical health, cognitive activity, and information processing. Integrity verification uses a heuristic technique based on medical, biology, brain engineering research, machine learning, and statistical methods. DeepVision found Deepfakes in seven of eight video formats, with an 87.5% accuracy rate in this research. Ramachandran et al. [21] have evaluated deep facial recognition's ability to detect deepfakes using loss functions and deepfake creation methods. It outperforms two-class CNNs and ocular modality on Celeb-DF and FaceForensics++ datasets. A biometric face recognition technology eliminates false data for model training and improves generalizability to new deep fake-generating approaches in this research.

Liu et al. [22] have introduced a block shuffling learning strategy for counterfeit face detection, partitioning images into blocks and applying random shuffling for robustness. The proposed model is compatible with CNN models and offers excellent generalization in the face of frequent picture alterations in this research work. Rafique et al. [23] have proposed an ELA-based deepfake detection for recognizing authentic and fraudulent pictures. This research sends pre-processed datasets to CNN models Alex Net and Shuffle Net for image categorization, while KNN and SVM classifiers produce deep feature vectors. The proposed system achieves 88.2% accuracy and is lightweight, strong, and efficient.

Zhuo et al. [24] have presented a novel model named WISERNet (Wider Separate-then-reunion Network), a newly suggested deep-learning-based data-driven colour image step analyzer in false colourized picture detection. In this model, the detection performance of the proposed detector beats FCIDHIST and FCID-FE, and it has never explicitly used any information from any channels other than the standard red, green, and blue colour channels. This is a significant departure from previous research in this area. Zhao et al. [25] have proposed a multi-attentional deepfake detection network with three key components: multiple spatial attention heads, textural feature enhancement block, and attention maps. In this research, they have shown superior performance over vanilla binary classifiers.

Ismail et al. [26] have proposed a model named YOLO-InceptionResNetV2- XGBoost, an efficient design that detects video frame abnormalities and inconsistencies and assesses if a video is genuine or deepfake. This model YOLO detector performs well in object and face detection systems and can reduce background false positives. XGBoost is used for deep-learning-based feature extraction and COVID-19 and pneumonia chest X-ray diagnosis. Qureshi et al. [27] have presented a vocal impersonation-based deepfake detection method that embeds digital watermarks in a video's audio track using hybrid speech watermarking. Robust and fragile watermarks can be detected by independent software, and the integrated watermark's resistance to signal processing and video integrity assaults is simulated in this work. This method is one of the earliest digital watermarking techniques to identify bogus material.

Rahman et al. [28] have trained a CNN model to recognize false films in low-resolution and short-time video data. They have used Kaggle Deepfake Detection Challenge (DFDC) and Face Forensics++ datasets. In this work, their algorithm detects false movies with 94.93% accuracy for DFDC and 93.2% for FaceForensics++ datasets. They have compared their models against

state-of-the-art approaches using several performance criteria. Zhong et al. [29] have proposed a graph-based methodology that uses document factual structure to identify deepfake text. In this research, a proposed graph neural network learns sentence representations from a document's entity graph, enhancing RoBERTa-built strong foundation models and distinguishing between machine-generated and human-written factual structures. Patel et al. [30] have presented an enhanced deep-CNN architecture for detecting deepfakes with high accuracy and generalizability. In this work, the proposed architecture uses various image sources and techniques like binary-cross entropy and Adam optimizer to train the model. The accuracy ranges from 98.33% in AttGAN to 99.33% in GDWCT, StyleGAN2, and StarGAN obtained in the research work. Li et al. [31] have proposed an Artifacts-Disentangled Adversarial Learning (ADAL) system to accurately identify deepfakes by disentangling artifacts from irrelevant information.

The proposed method in this work constructs fresh samples using Multi-scale Feature Separator (MFS) and Artefacts Cycle Consistency Loss (ACCL) to estimate artifacts and offer visual proof. Ilyas et al. [32] have presented a unique AVFakeNet framework for detecting deepfakes in videos by concentrating simultaneously on both the audio and visual modalities of a video. It is known as a Dense Swin Transformer Net (DST-Net), and its components are as follows: an input block, a block for feature extraction, and an output block. In this research, the results of the experiments demonstrate that the developed framework is successful in effectively identifying deepfake videos. Yang et al. [33] have presented Audio-Visual Joint Learning for Detecting Deepfake (AVoiD-DF) to identify multi-modal forgeries using audio-visual inconsistencies. It fuses multi-modal information and collaboratively learns inherent correlations using Temporal-Spatial Encoder, Multi-Modal Joint-Decoder, and Cross-Modal Classifier. Experimental findings on DefakeAVMiT, FakeAVCeleb, and DFDC show that the AVoiD-DF surpasses several state-of-the-art in deepfake detection and generalization.

Deng et al. [34] have proposed a cascaded Network based on EfficientNet and Transformer to achieve deepfake detection tasks. In this research, EfficientNetV2S is used as a feature extractor, Transformer is used for classification, and SRA is used to improve the traditional attention mechanism. They have achieved state-of-the-art results with 92.16% and 96.75% accuracy and excellent visualization results on deepfake videos.

3. Categories of Deepfake

Deepfake has different types of categories based on the applications. The following four are the important categories of deepfake (fig.1).

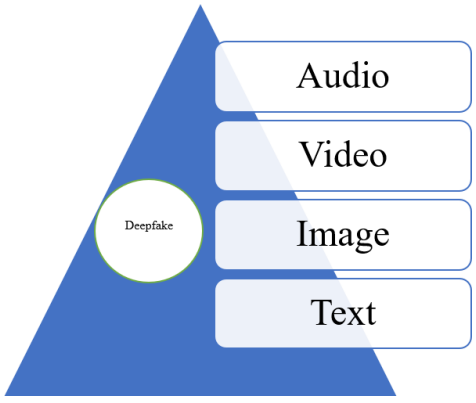

Figure 1: A graphical representation of categories of deepfake

3.1. Deepfake Text

Deepfake text uses artificial intelligence to generate fake news articles, social media posts, or text-based content [9]. The technology used to generate deepfake text is known as Natural Language Processing (NPL) [6]. NLP algorithms can be trained on a large text dataset to learn the language patterns and generate new text that mimics the style and tone of the training data. GPT-3, developed by Open AI, is one of the most advanced NLP models available, capable of generating highly convincing text [43]; [44]. Below Fig.2 is an example of deepfake text, which has been created by using AI.

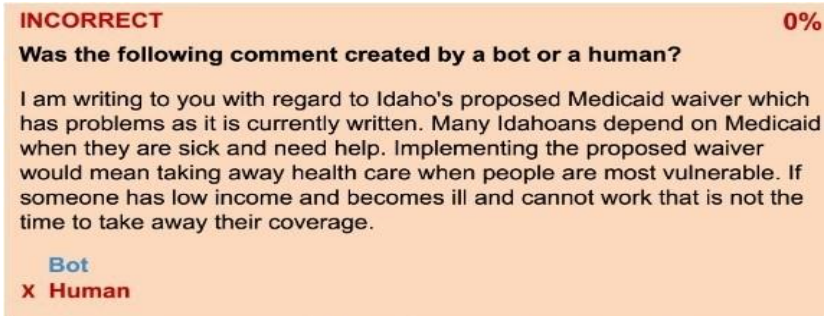

Figure 2: Illustration of deepfake text that was created by a bot yet seemed to come from a person, according to the results of a poll [51]

A recent study has shown that deep fake text can be used for various purposes, including generating fake news, spreading disinformation, and impersonating individuals online [6]. Researchers are also developing techniques to detect deepfake text, such as analyzing the writing style, checking for logical inconsistencies, or using AI to flag suspicious content [29].

3.2. Deepfake Image

Deepfake image technology involves manipulating images or photographs using AI algorithms to create realistic-looking but fake images [3]. The most used deepfake image technology is Generative Adversarial Networks (GANs), which consist of two neural networks; a generator creates fake images, while the discriminator.

Evaluate the authenticity of the images [11]. The two networks are trained together in a process known as adversarial training until the generator produces images that are distinguishable from formal ones [7]. Fig.3 is an example of a deepfake image created using the GAN algorithm.

Figure 3: Illustration of a deepfake Image using Artificial Intelligence Technology [50]

Deepfake images can be used for various purposes, such as creating fake profile pictures, generating fake images for social media, or manipulating images for social media or manipulating images for political propaganda [24]. However, researchers are also working on developing techniques to detect deepfake images, such as analyzing image metadata, checking for distortions in the image, or using AI to detect inconsistencies [30].

3.3. Deepfake Video

Deepfake video technology uses AI algorithms to manipulate footage or create entirely new videos not based on real footage [3]; [31]. The most commonly used deepfake video technology is based on GANs, which can be trained on motion patterns and create new videos that mimic the style and content of the training data [17].

Figure 4: An example of a deepfake video in which the president of Ukraine gives the order to surrender during a time of Russia-Ukraine war [52]

Deepfake videos can be used for various purposes, such as creating fake news footage, manipulating video evidence, or creating fake videos for entertainment [12]. Fig.4 is an example of a deepfake video that has been created using artificial intelligence technology; the video explains the president of Ukraine, Zelenskyy, gives the order to the soldiers of Ukraine to surrender during a Russia-Ukraine war. However, the increasing use of deepfake videos for malicious purposes has led to a growing concern about the potential impact of this technology [34]. Researchers are developing techniques to detect deepfake videos, such as analyzing facial movements, checking for inconsistencies in lighting and shadows, or using AI to identify unusual patterns of motion [35].

3.4. Deepfake Audio

Deepfake audio technology uses AI algorithms to create or manipulate audio recordings, such as speech or music [5]. The most commonly used deepfake audio technology is based on Generative Adversarial Networks (GANs), which can be trained on a large dataset of audio recordings to learn the patterns of sound and create new audio that mimics the style and content of the training data [18].

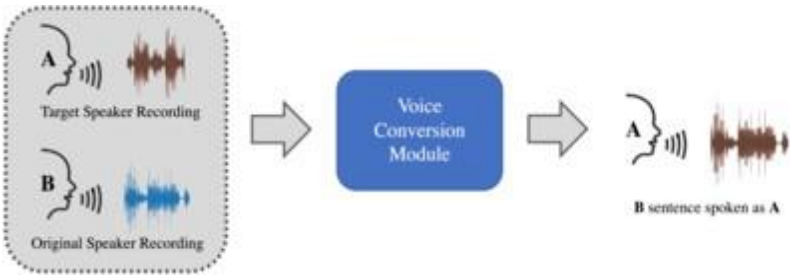

Figure 5: This is an example of a schematic of an Audio manipulation method [53]

Deepfake audio can be used for various purposes, such as creating fake audio evidence, impersonating someone's voice, or manipulating audio for entertainment [35]. In Fig.5 it explains how the deepfake audios are created by using a voice conversation module and created a deepfake audio. However, the increasing use of deepfake audio for malicious purposes has led to a growing concern about the potential impact of this technology [19]. Researchers are developing techniques to detect deepfake audio, such as analyzing the audio waveform.

4. Applications of Deepfake

There are different applications of deepfake technology. In this section, we have explained both beneficial and malicious applications of deepfake. They are as follows (fig.6):

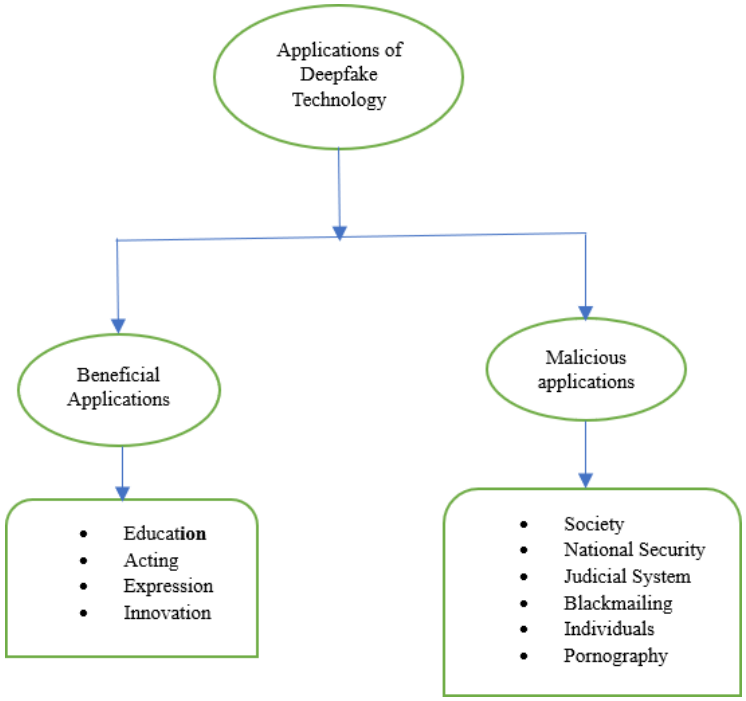

Figure 6: This diagram shows the Applications of Deepfake Technology

4.1. Beneficial Applications

Deepfake technology has a potential for misuse; despite that, it also has some useful applications in certain contexts. Some of the beneficial applications of deepfakes are explained in the following form.

4.2. Education

Deep-fake technology might be used in the classroom to give students a more engaging and hands-on learning experience than traditional methods like reading and listening to lectures. Concerns about fair use and intellectual property in altered content have been raised considering this technology [3]. Deepfakes, like augmented and virtual reality technology, are fascinating from a pedagogical perspective. Video editing software has come a long way, and today, it is simple and cheap to make instructional videos that utilize scenes from well-known shows or movies as a jumping-off point for debate. Learning about deep fakes has practical applications outside the classroom [36].

4.3. Acting

Deepfake technology has significantly advanced in recent years, impacting various fields, including the acting industry. Deepfakes are synthetic media, typically videos or images, created using artificial intelligence algorithms. These algorithms can manipulate or superimpose one person's likeness onto another, effectively allowing individuals to appear as someone else in a convincingly realistic manner [37]; [41].

4.4. Expression

Deepfake technology lets people with impairments such as amyotrophic lateral sclerosis (ALS) express themselves via virtual video and avatar experiences, allowing for virtual participation that may be unachievable in physically present situations [54]. It is also helpful in social communication since it allows speech to reach audiences whose native dialects are distinct from the speaker's natural language [8]; [37].

4.5. Innovation:

Innovations are beginning to embrace Deepfake as a digital transformation and automation tool. While Data Grid unveiled its artificial intelligence (AI) engine for the advertising and fashion industries, Reuters demonstrated an AI-generated sports news summary system [26]. When applied to older videos, deepfakes may be used to construct virtual trial rooms and mixed-reality settings and improve low-resolution photographs. Involving and satisfying consumers is a breeze with these methods [38].

5. Mischievous Applications

Unfortunately, deepfake technology can be misused for malicious purposes, causing harm and deception. Some of the naughty applications of deepfake include the following.

5.1. Society

Deepfake technology has significantly advanced in recent years, impacting various fields, including the acting industry. Deepfakes are synthetic media, typically videos or images, created using artificial intelligence algorithms. These algorithms can manipulate or superimpose the likeness of one person onto another, effectively allowing individuals to appear as someone else in a convincingly realistic manner [40].

5.2. National Security

Deepfakes may endanger national security if they are used deliberately by bad people. The countries are worried about misinformation tactics that might exacerbate political and social divides and endanger national security during the elections [36]. Inaccurate information might compromise national security in several ways, including endangering the safety of military personnel dealing with foreign civilian populations. Additionally, hostile foreign governments may employ Deepfakes to produce propaganda that portrays prominent international figures acting hostilely or aggressively [39].

5.3. Judicial System

Deepfake Technology is a new law area that raises concerns about how it may affect the court system. It may increase the number of cases the courts must handle and jeopardise the credibility of witnesses and litigants [37]. Additionally, Deepfakes may unintentionally or purposefully store media recordings that were formerly thought to be reliable, casting doubt in the minds of the jury and judge. The foundation of the legal system may be rocked because of this [39].

5.4. Blackmailing

Blackmailing is one of the applications of deepfake Artificial Intelligence technology. Blackmailers may find that Deepfake movies are a useful tool to abuse and get something of value from the people they are blackmailing. Although some individuals have nothing to lose, reversing the initial harm caused by Deepfakes could nevertheless drive them to give in to the threat and supply resources like money, knowledge, and so on [15].

5.5. Individuals

Deepfake may inflict physical and psychological suffering and physical pain on individuals. Fraudsters may utilize it to take significant assets, typically asking victims to supply personal information. Deepfake films in the workplace might represent anti-social behaviour, such as causing property damage or racist statements. Deepfake audio or video may also be used to prove harassment or sexual assault, damaging an individual's professional life and future goals [5]; [40].

5.6. Pornography

Deepfake porn videos for sexual or financial gain are common online. Deepfake videos use manipulated pornographic films to target celebrities or well-known members of society and to discriminate fake information using fictitious unreliable, and accounts witnesses [47]; [48]. Pornographic movies are the most popular kind of Exploitation, and they may lead to violent or humiliating behaviours [42].

6. Challenges and Limitations

Despite the advancement in deepfake detections using Artificial Intelligence, there are still several challenges and limitations [45]. Some of these challenges include:

- Limited Data: Deepfake detection requires large amounts of data for training and testing. However, limited data is available for deepfake detection, which can affect the accuracy of the algorithms [49].
- Adaptability: Deepfake algorithms are constantly evolving, and deepfake detection algorithms must adapt to these changes [10].
- Privacy Concerns: Deepfake Detection algorithms require access to personal data, which can raise privacy concerns [47].

7. Conclusion

Deepfake detection using artificial intelligence is a rapidly evolving field, and new techniques are constantly being developed. While traditional techniques have effectively detected deepfakes, deep learning techniques are more advanced and can analyze complex patterns and features. However, several challenges and limitations still need to be addressed. Nonetheless, the continued development of deepfake detection techniques is critical to maintaining the authenticity of information and presenting harm. Deep fake technology, powered by artificial intelligence (AI), is a double-edged sword with far-reaching implications for society. On one hand, it offers exciting possibilities for entertainment, creativity, and accessibility in various fields. On the other hand, it presents serious concerns and challenges related to misinformation, privacy, security, and ethical considerations.

Acknowledgement: We are grateful for everyone who helped me write this.

Data Availability Statement: The research in this study was conducted using internet benchmark data. The research presented here is brand new.

Funding Statement: No funding has been obtained to help prepare this manuscript and research work.

Conflicts of Interest Statement: The writers have made no disclosures of potential bias (s). This is brand new writing from the authors. The used information is cited and referenced appropriately.

Ethics and Consent Statement: Institutional review board permission and participant consent were obtained throughout data collection.

References

1. N. Khatri, V. Borar, and R. Garg, "A Comparative Study: Deepfake Detection Using Deep-learning," in 2023 13th International Conference on Cloud Computing, Confluence: IEEE, pp. 1–5, 2023.
2. N. Waqas, S. I. Safie, K. A. Kadir, S. Khan, and M. H. Kaka Khel, "DEEPFAKE Image Synthesis for Data Augmentation," IEEE Access, vol. 10, pp. 80847–80857, 2022.
3. R. L. Testa, A. Machado-Lima, and F. L. S. Nunes, "Deepfake detection on videos based on ratio images," in 2022 12th International Congress on Advanced Applied Informatics (IIAI-AAI), 2022.
4. H. Zhao, T. Wei, W. Zhou, W. Zhang, D. Chen, and N. Yu, "Multi-attentional Deepfake Detection," in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
5. J. Khochare, C. Joshi, B. Yenarkar, S. Suratkar, and F. Kazi, "A deep learning framework for audio deepfake detection," Arab. J. Sci. Eng., 2021.
6. S. G. Tesfagergish, R. Damaševičius, and J. Kapočiūtė-Dzikienė, "Deep fake recognition in tweets using text augmentation, word embeddings and deep learning," in Computational Science and Its Applications – ICCSA 2021, Cham: Springer International Publishing, pp. 523–538, 2021.
7. A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner, "FaceForensics++: Learning to detect manipulated facial images," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
8. Y. Patel et al., "An improved dense CNN architecture for deepfake image detection," IEEE Access, vol. 11, pp. 22081–22095, 2023.
9. S. J. Sohrawardi et al., "Defaking DeepFakes: Understanding journalists' needs for DeepFake detection," in Proceedings of the Computation+ Journalism 2020 Conference, 2020.
10. C. Li et al., "A continual deepfake detection benchmark: Dataset, methods, and essentials," in 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023.
11. K. A. Pantserev, "The malicious use of AI-based deepfake technology as the new threat to psychological security and political stability," in Advanced Sciences and Technologies for Security Applications, Cham: Springer International Publishing, pp. 37–55, 2020.

12. V. N. Tran, S. H. Lee, H. S. Le, B. S. Kim, and K. R. Kwon, "Learning Face Forgery Detection in Unseen Domain with Generalization Deepfake Detector," in 2023 IEEE International Conference on Consumer Electronics (ICCE), IEEE, pp. 1–06, 2023.
13. A. S. Sinaga, A. S. Sitio, and S. Dewi, "Identification Of Biometric Deepfakes Using Feature Learning Deep Learning," Jurnal Teknik Informatika (Jutif), vol. 3, no. 4, pp. 1125–1130, 2022.
14. N. Khatri, V. Borar, and R. Garg, "A Comparative Study: Deepfake Detection Using Deep-learning," in 2023 13th International Conference on Cloud Computing, Confluence: IEEE, pp. 1–5, 2023.
15. Y. Patel et al., "An improved dense CNN architecture for deepfake image detection," IEEE Access, vol. 11, pp. 22081–22095, 2023.
16. V. N. Tran, S. H. Lee, H. S. Le, B. S. Kim, and K. R. Kwon, "Learning Face Forgery Detection in Unseen Domain with Generalization Deepfake Detector," in 2023 IEEE International Conference on Consumer Electronics (ICCE), IEEE, pp. 1–06, 2023.
17. A. Chintha et al., "Recurrent convolutional structures for audio spoof and video deepfake detection," IEEE J. Sel. Top. Signal Process., vol. 14, no. 5, pp. 1024–1037, 2020.
18. J. Khochare, C. Joshi, B. Yenarkar, S. Suratkar, and F. Kazi, "A deep learning framework for audio deepfake detection," Arab. J. Sci. Eng., 2021.
19. T. Chen, A. Kumar, P. Nagarsheth, G. Sivaraman, and E. Khoury, "Generalization of audio deepfake detection," in The Speaker and Language Recognition Workshop (Odyssey 2020), 2020.
20. T. Jung, S. Kim, and K. Kim, "DeepVision: Deepfakes detection using human eye blinking pattern," IEEE Access, vol. 8, pp. 83144–83154, 2020.
21. S. Ramachandran, A. V. Nadimpalli, and A. Rattani, "An experimental evaluation on deepfake detection using deep face recognition," in 2021 International Carnahan Conference on Security Technology (ICCST), 2021.
22. S. Liu, Z. Lian, S. Gu, and L. Xiao, "Block shuffling learning for Deepfake Detection," arXiv [cs.CV], 2022.
23. R. Rafique, M. Nawaz, H. Kibriya, and M. Masood, "DeepFake detection using error level analysis and deep learning," in 2021 4th International Conference on Computing & Information Sciences (ICCIS), 2021.
24. L. Zhuo, S. Tan, J. Zeng, and B. Lit, "Fake colorized image detection with channel-wise convolution based deep-learning framework," Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, pp. 733–736, 2018.
25. H. Zhao, T. Wei, W. Zhou, W. Zhang, D. Chen, and N. Yu, "Multi-attentional Deepfake Detection," in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
26. A. Ismail, M. Elpeltagy, M. S Zaki, and K. Eldahshan, "A new deep learning-based methodology for video deepfake detection using XGBoost," Sensors (Basel), vol. 21, no. 16, p. 5413, 2021.
27. A. Qureshi, D. Megías, and M. Kuribayashi, "Detecting deepfake videos using digital watermarking," in 2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, IEEE, pp. 1786–1793, 2021.
28. A. Rahman et al., "Short and low resolution deepfake video detection using CNN," in 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC), 2022.
29. W. Zhong et al., "Neural deepfake detection with factual structure of text," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.
30. Y. Patel et al., "An improved dense CNN architecture for deepfake image detection," IEEE Access, vol. 11, pp. 22081–22095, 2023.
31. X. Li, R. Ni, P. Yang, Z. Fu, and Y. Zhao, "Artifacts-disentangled adversarial learning for deepfake detection," IEEE Trans. Circuits Syst. Video Technol., pp.Vol. 33, no.4, pp. 1658-1670, 2022.
32. H. Ilyas, A. Javed, and K. M. Malik, "AVFakeNet: A unified end-to-end Dense Swin Transformer deep learning model for audio- visual deepfakes detection," Applied Soft Computing, vol. 136, 2023.
33. W. Yang et al., "AVoiD-DF: Audio-visual joint learning for detecting deepfake," IEEE Trans. Inf. Forensics Secur., pp. 1–1, 2023.
34. L. Deng, J. Wang, and Z. Liu, "Cascaded Network Based on EfficientNet and Transformer for Deepfake Video Detection," Neural Processing Letters, pp. 1–20, 2023.
35. A. Chintha et al., "Recurrent convolutional structures for audio spoof and video deepfake detection," IEEE J. Sel. Top. Signal Process., vol. 14, no. 5, pp. 1024–1037, 2020.
36. M. Filimowicz, Ed., Deep Fakes: Algorithms and Society. Routledge. 2022.
37. C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, and C. Busch, Handbook of digital face manipulation and detection: from DeepFakes to morphing attacks. Springer Nature, 2022.
38. R. Chesney and D. K. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security," SSRN Electron. J., 2018.
39. L. Verdoliva and P. Bestagini, "Multimedia Forensics," in Proceedings of the 27th ACM International Conference on Multimedia, 2019.
40. C. M. Kerner, Detecting Deepfakes: Philosophical Implications of a Technological Threat. 2020.
41. L. Gaur, (Ed.). DeepFakes: Creation, Detection, and Impact. CRC Press, 2022.

42. M. Tanaka, S. Shiota, and H. Kiya, “A detection method of operated fake-images using robust hashing,” J. Imaging, vol. 7, no. 8, p. 134, 2021.
43. J. Zhang, B. Dong, and P. S. Yu, “Deep diffusive neural network based fake news detection from heterogeneous social networks,” in 2019 IEEE International Conference on Big Data (Big Data), 2019.
44. S. G. Tesfagergish, R. Damaševičius, and J. Kapočiūtė-Dzikienė, “Deep fake recognition in tweets using text augmentation, word embeddings and deep learning,” in Computational Science and Its Applications – ICCSA 2021, Cham: Springer International Publishing, pp. 523–538, 2021.
45. R. Rafique, R. Gantassi, R. Amin, J. Frnda, A. Mustapha, and A. H. Alshehri, “Deep fake detection and classification using error- level analysis and deep learning,” Scientific Reports, vol. 13, no. 1, 2023.
46. N. Bonettini, E. D. Cannas, S. Mandelli, L. Bondi, P. Bestagini, and S. Tubaro, “Video face manipulation detection through ensemble of CNNs,” in 2020 25th International Conference on Pattern Recognition (ICPR), 2021.
47. Y. Li and S. Lyu, “Exposing DeepFake videos by detecting face warping artifacts,” arXiv [cs.CV], 2018.
48. D. Wodajo and S. Atnafu, “Deepfake video detection using Convolutional Vision Transformer,” arXiv [cs.CV], 2021.
49. S. S. Prasad, O. Hadar, T. Vu, and I. Polian, “Human vs. Automatic detection of deepfake videos over noisy channels,” in 2022 IEEE International Conference on Multimedia and Expo (ICME), 2022.
50. G. Lee and M. Kim, “Deepfake detection using the rate of change between frames based on computer vision,” Sensors (Basel), vol. 21, no. 21, p. 7367, 2021.
51. H. Smith, Weaponised deep fakes: national security and democracy, 2020.
52. Arab News, “Zelenskyy deepfake video goes viral, reflecting troubling new wave of disinformation,” Arab news, Arabnews, Accessed by 18-Mar-2022.
53. Wikipedia contributors, “Audio deepfake,” Wikipedia, The Free Encyclopedia, 30-Aug-2022. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Audio_deepfake&oldid=1173001416.
54. D. Gong, Y. J. Kumar, O. S. Goh, Z. Ye, and W. Chi, “DeepfakeNet, an efficient deepfake detection method,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 6, 2021.