Received: 9 February 2022 | Revised: 17 August 2022 | Accepted: 5 October 2023

DOI: 10.1002/widm.1520

ADVANCED REVIEW

WIREs DATA MINING AND KNOWLEDGE DISCOVERY WILEY

Deepfake detection using deep learning methods: A systematic and comprehensive review

Arash Heidari[1] | **Nima Jafari Navimipour**[2,3] | **Hasan Dag**[4] | **Mehmet Unal**[5]

[1]Department of Software Engineering, Haliç University, Istanbul, Turkey

[2]Department of Computer Engineering, Kadir Has Universitesi, Istanbul, Turkey

[3]Future Technology Research Center, National Yunlin University of Science and Technology, Douliou, Taiwan

[4]Management Information Systems, Kadir Has Universitesi, Istanbul, Turkey

[5]Department of Computer Engineering, Nisantasi Universitesi, Istanbul, Turkey

Correspondence
Nima Jafari Navimipour, Department of Software Engineering, Haliç University, Istanbul, 34060, Turkey.
Email: nima.navimipour@khas.edu.tr

Edited by: Mehmed Kantardzic, Associate Editor and Witold Pedrycz, Editor-in-Chief

Abstract

Deep Learning (DL) has been effectively utilized in various complicated challenges in healthcare, industry, and academia for various purposes, including thyroid diagnosis, lung nodule recognition, computer vision, large data analytics, and human-level control. Nevertheless, developments in digital technology have been used to produce software that poses a threat to democracy, national security, and confidentiality. Deepfake is one of those DL-powered apps that has lately surfaced. So, deepfake systems can create fake images primarily by replacement of scenes or images, movies, and sounds that humans cannot tell apart from real ones. Various technologies have brought the capacity to change a synthetic speech, image, or video to our fingers. Furthermore, video and image frauds are now so convincing that it is hard to distinguish between false and authentic content with the naked eye. It might result in various issues and ranging from deceiving public opinion to using doctored evidence in a court. For such considerations, it is critical to have technologies that can assist us in discerning reality. This study gives a complete assessment of the literature on deepfake detection strategies using DL-based algorithms. We categorize deepfake detection methods in this work based on their applications, which include video detection, image detection, audio detection, and hybrid multimedia detection. The objective of this paper is to give the reader a better knowledge of (1) how deepfakes are generated and identified, (2) the latest developments and breakthroughs in this realm, (3) weaknesses of existing security methods, and (4) areas requiring more investigation and consideration. The results suggest that the Conventional Neural Networks (CNN) methodology is the most often employed DL method in publications. According to research, the majority of the articles are on the subject of video deepfake detection. The majority of the articles focused on enhancing only one parameter, with the accuracy parameter receiving the most attention.

This article is categorized under:

Technologies > Machine Learning

Algorithmic Development > Multimedia

Application Areas > Science and Technology

KEYWORDS
deep learning, deepfake, detection, neural networks, review

1 | INTRODUCTION

The widespread availability of low-cost digital devices, including smartphones, laptops, desktop computers, and digital cameras, has triggered the development of multimedia material (such as photographs and movies) on the Internet and wireless communication systems (Feng et al., 2021; Jia et al., 2022; D. Li, Ge, et al., 2020). Furthermore, during the last few decades, social media has enabled people to quickly communicate recorded multimedia content, resulting in significant growth in multimedia content output and accessibility (Lv et al., 2020). Increased the speed with which false and erroneous information can be manufactured and circulated, knowing the truth and trusting the information has become increasingly challenging, perhaps resulting in disastrous repercussions (Garg & Mago, 2021; A. Li, Spano, et al., 2020; Niu et al., 2022). So, a deepfake is material created by Deep Learning (DL) (Luo, Yuan, et al., 2022) that seems genuine in a human's eyes. The term deepfake is a mixture of the phrases DL and fake, and it generally refers to material created by a deep neural network (DNN), which is a subset of machine learning (ML) (S. Li, Liu, et al., 2020; Lv, Li, et al., 2021; Y. Wang et al., 2021; Zhong et al., 2021). Everyone can manipulate video and image items (R. Liu et al., 2021). This has been available for several years due to many user-friendly software packages that allow video editing, audio, and picture (Aversano et al., 2021). The adoption of smartphone applications that conduct automated procedures, including lip-syncing, audio instrumentation, and face swaps, has simplified media manipulation (Tolosana et al., 2020). Furthermore, DL-driven technical developments have resulted in a slew of AI-driven technologies that make manipulations very convincing and believable (Castillo Camacho & Wang, 2021). Every one of these techniques seems to be an unquestionably valuable addition to the toolbox of a digital artist (Afchar et al., 2018; Y. Li, Zhang, et al., 2021). Nevertheless, when used maliciously to create fake media, they may have significant negative social or personal implications (Muhammad & Hossain, 2022). Deepfakes are a notable example of artificially manipulated media that has caused serious concern. Furthermore, the deepfake AI-driven technique replaces one person's identity in a video with that of another (Sanghvi et al., 2021). This often propagates false information by impersonating politicians and distributing revenge porn (Matern et al., 2019).

Deepfake, which refers to various face modification techniques, incorporates cutting-edge technologies, including computer vision and DL (Lv, Chen, et al., 2021; Masi et al., 2020). Face manipulation is divided into four categories: expression swap, full-face synthesis, attribute manipulation, and identity swap (Balaji et al., 2021). Identity swap, often known as face swap, is one of the most common types of deepfake video wherein the faces of the source people are replaced with the faces of the targeted people (Jafar et al., 2020). Deepfake news can also be coupled with forged videos and photos, making it difficult for users to distinguish them (Matern et al., 2019). This form of deepfake has the potential to spread over social media and negatively influence people's lives (H. H. Nguyen, Yamagishi, et al., 2019). Although some deepfakes could be made using classic visual effects or computer graphics, DL methods, including autoencoders and generative adversarial networks (GANs), have since been extensively used in the computer vision sector and are the most recent popular underlying mechanism for deepfake generation (Gosse & Burkell, 2020). Such models describe a person's emotions and motions and synthesize facial images of someone with similar expressions and actions (Sanghvi et al., 2021). To train models to generate photo-realistic videos and pictures, deepfake techniques often take up a considerable amount of audio/video data (Jung et al., 2020). Celebrities and politicians are the first subjects of deepfakes since they have a vast quantity of videos and photographs available online (Jiang et al., 2021). Deepfakes are also utilized in pornographic photographs and movies to replace the faces of celebrities and politicians with bodies (Schlett et al., 2021). Whenever deepfake techniques could be used to make movies of international leaders with fake statements for falsification reasons, it represents a risk to global security (Fernandes & Jha, 2020). Deepfakes can thus be used to incite political or religious tensions between countries, deceive the public and influence election results, or create havoc in financial markets by spreading false information (Jeyaraj & Dwivedi, 2020; Nirkin et al., 2021). However, it could be used to make artificial satellite photos of the Earth that include features that do not correspond to reality to deceive military analysts, including a fake bridge across a river when there is none in actuality (Biswas et al., 2021; Heidari, Jafari Navimipour, et al., 2022). It could lead to a troop being misled when crossing a bridge during combat (Baygin et al., 2022; Siegel et al., 2021). Also, transfer learning (TL) is critical since the model must be trained in a reasonable period and with little resources to achieve the necessary accuracy for classification over a wide range of

instances in the dataset (Garg & Mago, 2021; Kim & Cho, 2021; Zong & Wan, 2022; Zong & Wang, 2022). It enables models to be trained fast and accurately by finding relatively useful spatial features from large datasets across various domains at the start of the training process (Heidari, Navimipour, et al., 2022; Luo, Zhang, et al., 2022). As a result, TL might be a means to combine the necessary computing power and encourage more efficient DL techniques, which could help with the resolution of several problems (Dwivedi et al., 2021; P. Yu et al., 2021). Because of the size of datasets, the importance of datasets is significant. The size of the dataset used to generate the findings has a massive effect. Various datasets, including faceforensics, CelebA, TIMIT, and HFM, are used in this context (Hinton et al., 2015; Zi et al., 2020).

Nonetheless, there is no thorough and detailed evaluation of the application of DL approaches in the deepfake detection area that considers all categories such as image, video, and multimedia content. The study focused on emphasizing the achievements and detection of deepfakes utilizing DL approaches to combat fraud, stigmas, scams, spoofing, and so on to offer a complete overview of the modern systems using these technologies. Given recent reviews on the various DL applications developed for deepfake detection techniques, this review substantially contributes by addressing the most intriguing study area. This study employs a systematic literature review (SLR) (Doewes et al., 2023; Zadeh et al., 2023) to find, analyze, and combine findings from related studies. Image deepfake detection, video deepfake detection, sound deepfake detection, and hybrid multimedia deepfake detection are the four primary kinds of DL strategies used in deepfake detection. We evaluated numerous features such as advantages, obstacles, dataset, usages, simulation environments, security, and TL for each category and approach that applied DL methods in this area. This article discusses the use of DL approaches in deepfake detection and addresses a variety of concerns. We have also gone over future work in great detail, highlighting all issues to rectify. In a nutshell, the following are the contributions of this article:

- Providing a wide overview of the topic of DL in the deepfake detection area;
- Providing a full overview of existing methods for DL-deepfake detection;
- Providing an overview of the key methods in DL that incorporate deepfake detection;
- Investigating each strategy for DL-deepfake detection with critical features;
- Highlighting the essential areas where the mentioned techniques can be improved.

The following classification provides the structure of this article. The basic concepts and corresponding terminologies of DL in deepfake detection are discussed in the next section. The relevant research studies are examined in Section 3. The research methodology and techniques for article selection are presented in Section 4. The categories of the selected articles are described in Section 5. The results and comparisons are presented in Section 6. Finally, the open issues are discussed in Section 7, and the conclusion is defined in the last sections. Moreover, the used abbreviations are shown in Table 1.

2 | BASIC CONCEPTS AND CORRESPONDING TERMINOLOGIES

This section covers the fundamentals of deepfakes, ML, and DL in general and how they are used in the deepfakes domain.

2.1 | Fake content creation

There are several techniques for manipulating visual material and tracking different targets (Niknejad et al., 2020; Yan et al., 2022). The following section will provide a quick overview of some of the more common and promising among them. Adding, replicating, and deleting objects are all very frequent actions (Meskys et al., 2020). A new item can be added by copying it from another picture (splicing) or the same image (copy-move). An existing item can be removed by expanding the backdrop to cover it (inpainting), as in the renowned exemplar-based inpainting (S. R. A. Ahmed & Sonuç, 2021). All of these procedures are simple to carry out with common picture editing software. Furthermore, appropriate post-processing, including scaling, rotation, or color correction, could be used to match the item to the scene better, improve the visual appeal, and ensure consistent perspective and scale (Deshmukh & Wankhade, 2021). Meanwhile, sophisticated computer graphics techniques, DL, and multi-layer networks have produced similar

TABLE 1 Table of abbreviations.

Abbreviation	Definition	Abbreviation	Definition
AI	Artificial intelligence	MFNN	Multi-layer fusion neural network
AMTEN	Adaptive manipulation traces extraction network	ML	Machine learning
AUC	Area under the curve	NLP	Natural language processing
AUROC	Area under the receiver operating characteristic	NN	Neural network
BP-DANN	Backpropagation based on domain-adversarial neural network	PoA	Proof of authenticity
CNN	Conventional neural networks	RGB	Red green blue
DL	Deep learning	RCNN	Region conventional neural networks
DNN	Deep neural network	RNN	Recurrent neural networks
ELA	Error level analysis	SA-DTH-Net	Speaker authentication dynamic talking habit network-based
GAN	Generative adversarial networks	SCNN	Set CNN
HF	Hyperledger fabric	SFFN	Shallow-FakeFaceNet
HFF	Hybrid fake face	SLR	Systematic literature review
HFM	Handcrafted face manipulation	SRM	Spatial rich model
HMM	Hidden Markov model	SVM	Support vector machine
HRNet	High-resolution network	TL	Transfer learning
JPEG	Joint photographic experts group	TTS	Text-to-speech
LBP	Local binary patterns	VSA	Visual speaker authentication
MCNet	Manipulation classification network	YOLO	You only look once

outcomes with improved semantic consistency in current history (Hossain et al., 2022; Zheng, Liu, & Yin, 2021). Manipulations that do not necessitate the use of advanced AI technologies are frequently referred to as "shallow fakes" or "cheap fakes" (Dixit & Silakari, 2021). However, its ability to alter perception could be significant (Heidari, Toumaj, et al., 2022). For instance, a video's meaning can be altered by deleting, adding, or cloning entire sets of frames (L. Verdoliva, 2020). Aside from these "conventional" manipulations of particular portions of an image or video, DL and computer graphics are now providing many new ones (Iqbal & Qureshi, 2022). Occasionally media assets are created entirely from scratch. To that aim, auto-encoders and GAN enabled effective solutions, particularly for face synthesis, and with a high degree of photo-realism attained (Habeeba et al., 2021). A segmentation map may also be used to produce a synthetic video or picture. Image synthesis can also be accomplished using simply a drawing or a text description (Soares & Parreiras, 2020). Similarly, a person's face can be modified depending on an auditory input sequence (Nasar et al., 2020). Often, the modification alters existing pictures or videos. Style transfer is very well-conceived since it allows you to alter the style of a painting, convert peaches to oranges, or recreate an image in a new season. Face manipulation has received much attention because of its high semantic value and wide range of uses (Kietzmann et al., 2020).

2.2 | Deepfake detection, fake image detection, and fake video detection

Deepfake detection is typically considered a binary arrangement problem in which classifiers are used to distinguish between reliable and interfering movies (Singh et al., 2021). This technique requires a big library of actual and false videos to train classification models. The number of fraudulent videos is growing, even though it is insufficient to provide a standard for validating various finding methods (Q. Liu & Celebi, 2021). Several prior studies on detecting deepfake images or videos were focused on DL techniques, which could be split into two detection methods: Conventional Neural Networks (CNN)-based methods and Region Conventional Neural Networks (RCNN)-based methods (X. Zhou et al., 2020). CNN-based techniques take facial pictures from video frames and feed them into the CNN for training and prediction to get an image-level result. So, in deepfake videos, such algorithms only employ spatial

information from a single frame. RCNN-based techniques, on the other hand, require a series of video frames for training to produce a video-level result. This approach, known as RCNN, combines CNN and Recurrent Neural Networks (RNN) (Tariq et al., 2021). As a result, RCNN-based techniques could fully use spatial and temporal information in deepfake movies (Chung et al., 2015). Furthermore, several deepfake detection methods are based on the standard ML methods, including utilizing a support vector machine (SVM) as a classifier and extracting handmade characteristics, including biological signals, as a classifier (Yazdinejad et al., 2020). Face swap offers several compelling uses in video compositing, portrait alteration, and, most notably, individuality protection, as it can exchange faces in photos with ones from a group of standard images. Nonetheless, it is one of the methods used by imposters to acquire access to authentication systems. Besides, face photographs have become increasingly hard for forensics as DL, including CNN, GAN, SVM, random forest, and multi-layer perceptron, can preserve the photos' position, facial appearance, and light. Among the DL images, those created by GAN are perhaps the most difficult to spot because they are real and high-quality, thanks to GAN's capacity to learn the supply of input data and produce fresh results with the same input distribution. Though this advancement of GAN continues and various new modifications of GAN are frequently introduced, most works on the recognition of GAN creating samples do not examine the capability of the recognition models. A step of image preparation was also used in various methods (Sanghvi et al., 2021). This increases the statistical comparison between the actual image and the false image at the pixel level, requiring the forensic classifier to examine important associated and meaningful features with a higher simplification competence than previous image forensics methods or image stage assessment systems. Also, because of the robust deprivation of frame data following audio-visual compression, maximum image-based deepfake recognition techniques cannot be applied to videos (Du et al., 2020). Furthermore, videos have a variety of chronological properties that vary between sets of frames, making it challenging for techniques designed to detect individual fraudulent images. This subcategory focuses on deepfake video recognition methods and divides them into two groups: those that use chronological characteristics and those that investigate visual artifacts inside frames (Mehta et al., 2021).

In the meantime, certain deepfake video generation models cannot synthesize all of the textures of the face, resulting in some somewhat rough fake faces. For example, the small wrinkles on the face cannot be created accurately. At the same time, the created face is fused into the backdrop during the final phase of creating the false frame. Smoothing techniques are frequently employed to alleviate the border inconsistencies created by this process, which results in the loss of face texture features. Figure 1a shows a frame from the genuine videos VidTIMIT dataset, whereas Figure 1b shows a frame from the face manipulation videos DeepFake-TIMIT dataset. Human eyes, following the figure, have difficulty distinguishing which frame is real. Nevertheless, the actual frame has a more realistic texture, for example, double eyelids and wrinkle features around the eyes, but the false frame lacks texture details.

2.3 | Attribute manipulation, expression, and identity swap

The face of one individual in a film is replaced with another person's face in identify swap methods. Two approaches are often considered: the first is traditional computer graphics-based technologies, such as FaceSwap, and the second is innovative deepfakes techniques, such as the current ZAO smartphone app (Hongmeng et al., 2020). Also, on Youtube and commercial websites, you can find quite realistic examples of this type of modification. Such alteration could assist various industries, including the entertainment industry (Rana & Sung, 2020). On the other hand, it could be used for

FIGURE 1 Real and fabricated frames are shown. (a) is a genuine frame, whereas (b) is a fake (H. Zhao, Zhou et al., 2021).

nefarious objectives, including producing celebrity pornographic videos, frauds, and financial fraud, to name a few (Pokroy & Egorov, 2021). In addition, attribute manipulation, sometimes referred to as face alteration or face retouching, entails changing aspects of the face, including hair or skin color, gender, age, and the addition of glasses, among other things (Fink et al., 2020). Then, GAN, including the StarGAN technique introduced in, is commonly used to carry out this modification procedure. The popular FaceApp smartphone application is an example of this kind of manipulation (Akhtar et al., 2020). Users might utilize this innovation to test a wide range of products in a virtual environment, including cosmetics and makeup, spectacles, and hairstyles (Y. Zhang, Gao, et al., 2021). Also, face reenactment is a type of manipulation that involves altering a person's facial expression. Although several modification approaches are proposed in the literature, such as at the image level using famous GAN architectures, this group focuses on the most popular techniques, Face2Face, and NeuralTextures, which substitute one person's facial expression in a video with that of another person. This form of deception could have catastrophic ramifications if it is employed (Tu et al., 2021).

2.4 | Creating a fake speech and fake speech detection

Synthetic speech production is a subject that has been explored for many years and has been approached in various ways (Bekci et al., 2020). As a result, the literature has a vast number of approaches that provide amazing results, and there is no single consistent method of producing a synthetic voice track. Text-to-Speech (TTS) generation was traditionally based on concatenative waveform synthesis, which means that provided a sentence as input, the output speech is created by picking the proper diphone units from a huge dataset of diphone waveforms and integrating them to ensure acceptability (Katarya & Lal, 2020). Additional post-processing tools help for smoother transitions between diphones, simulation of human prosody, and a high degree of naturalness (Tjon et al., 2021). The primary disadvantage of concatenative synthesis is the complexity of changing the tonal features of the sound, for example, to alter the speaker or embed emotional resonance in the voice. Also, some approaches, such as the Hidden Markov Models (HMM) based speech synthesis system, are proposed to improve the range of voice characteristics or talking patterns (Zi et al., 2020). These use contextual HMMs trained on huge datasets of auditory characteristics collected from diphones and triphones. Another method, called parametric TTS synthesis techniques, seeks to increase the number of produced voices. In this scenario, a speech signal is produced as an auto-regressive process, given a collection of speech characteristics, including frequency (Hong et al., 2021), spectral envelope, and excitation signal. Nevertheless, parametric TTS synthesis yields fewer natural-sounding outcomes than concatenative TTS synthesis (Chi et al., 2020).

Neural networks (NNs) have created natural and versatile synthetic voices (Z. Wang, Ramamoorthy, et al., 2022; Z. Wang, Ramamoorthy, et al., 2022). Modeling audio sample by sample, for example, has traditionally been thought to be extremely difficult because speech signals often count 100 of samples per second and preserve essential features at many timeframes (Nasar et al., 2020). However, in recent decades, CNN and RNN have permitted the construction of entirely auto-regressive models, allowing the direct generation of raw audio waveforms (Burroughs et al., 2020). It is far from simple to determine if a speech recording belongs to a genuine person or was created artificially. Nevertheless, artificial speeches may be created using several methods, each with its own set of characteristics. It is difficult to establish a comprehensive forensic framework that shows all potential synthetic speech approaches (D. Xie et al., 2020). Furthermore, as ML solutions become more popular, new and improved methods of creating synthetic voice recordings are regularly presented (A. Li, Masouros, et al., 2021). Keeping up with the growth of speech synthesis literature is therefore difficult. Even so, the sector has developed a set of detectors to prevent the proliferation of false audio recordings. One of several aims of this study is to examine deepfake detection in the area of fake speech (Mitra et al., 2020).

2.5 | Domain adaptation and TL

Understanding TL in general and its use in deepfake detection is an important subject that is an intrinsic component of this area. However, TL is a subfield of DL that leverages information from source domains to help the method acquire knowledge from the target domain quicker and more effectively. TL has received much interest in the realm of forensics. Loading ImageNet's pre-trained weight to the model before training is a straightforward TL (Marei et al., 2021). The majority of works utilizing deep domain adaptation rely on discrepancy measurement. In addition, several efforts are based on domain-adversarial learning and discrepancy assessment (I. Ahmed et al., 2021). So, TL is an approach for

solving a new task (apparently related or unrelated) using previously learned model information with little retraining or fine-tuning. In comparison to traditional ML methods, DL requires a considerable amount of training data. The time it takes to train for a domain-specific problem is greatly reduced when using TL (Lewis et al., 2020). The dataset is alleviated using a TL-based DL approach (Hung & Chang, 2021). Because the growth of deepfakes has been so terrifying on the internet, detecting them has become a top priority in the current circumstances (Heidari et al., 2022). Researchers in this sector are working hard to develop some potential TL mechanisms that can assist in alleviating the challenges in such a circumstance.

3 | RELEVANT REVIEWS

Multiple methods for manipulating material such as photos, video, audio, and so on have been successfully created and are now accessible to the public, such as FaceSwap, Face2Face, deepfake, and so forth. It allows anybody to quickly and simply modify faces in video sequences, resulting in amazingly realistic outcomes with no effort. Furthermore, in this era of fake multimedia, easy access to large-scale public databases and rapid advances in DL techniques, notably GANs, have resulted in incredibly realistic fake content with societal ramifications. Similarly, deepfake detection is a significant application of DL and ML that assists detect forgeries in media, including videos and photos. An amount of research has indeed been done on it, including a thorough study and implementation of many common algorithms (Albahar & Almalki, 2019; T. T. Nguyen, Nguyen, et al., 2019; Pashine et al., 2021; Saif & Tehseen, 2022; Shelke & Kasana, 2021; Swathi & Sk, 2021; Weerawardana & Fernando, 2021). So, DL has been recognized as an excellent method for detecting synthetic media in the deepfake realm. For this reason, L. Verdoliva (2020) examined approaches for visual media integrity verification or identifying altered pictures and videos. She focused on the rising topic of deepfakes, or fake media generated by DL algorithms, as well as current data-driven forensic approaches for combating them. The article examined integrity verification methods, beginning with traditional ways and progressing to DL-based approaches before concluding with particular deepfake detection algorithms. The review emphasized the limitations of present forensic techniques, the most pressing concerns, impending difficulties, and future research paths. The work does not define a method category and does not compare parameters between methods.

Also, Tolosana et al. (2020) presented a detailed overview of facial picture manipulation techniques, particularly deepfake strategies, as well as methods to identify such alterations. Four methods of facial manipulation are discussed in detail: (i) complete face synthesis; (ii) identity swap; (iii) attribute manipulation; and (iv) expression swap. They present details about manipulation techniques, existing public databases, and essential standards for performance review of false detection systems, as well as a summary of the results of those evaluations for each manipulation group. They focused on the newest versions of deepfakes, stressing its advances and problems for fake identification, among all the topics mentioned in the study. They also talk about unresolved difficulties and future trends that should be explored as the field progresses.

Mirsky and Lee (2021) specialized in human reenactment and replacement deepfakes. They went into how these techniques function, the variations in their structures, and what is being done to identify them in detail. They look into the production and identification of deepfakes and give a detailed look into how these systems work. Their goal is to give the reader a better grasp of how deepfakes are made and recognized and the latest trends and developments in this field. They also point out flaws in present defense measures as well as topics that need additional study and emphasis.

Besides, Castillo Camacho and Wang (2021) provided a broad grasp of the detecting methods used in the field of image forensics. They gathered and presented a variety of DL-based methods grouped into three main categories, emphasizing the unique characteristics of picture forensics approaches. They discovered that a preprocessing procedure to achieve a specific feature or a customized initialization on the network's first layer was utilized in many pioneer works and is still employed in current ones. They provided a detailed overview of image forensics approaches, with a particular emphasis on profound algorithms. They addressed a wide range of image forensics issues, such as detecting regular picture alterations, detecting deliberate image falsifications, identifying cameras, classifying computer graphics images, and detecting developing deepfake images.

Finally, P. Yu et al. (2021) presented the deepfake video production technologies, analyzed the available detection system, and discussed the path of the research directions. Their paper provided a state of research in deepfake video detection, namely the production process, various detection algorithms, and current standards. The many sorts of detection techniques are then discussed. The review relies on new difficulties with current detection algorithms as well as potential developments. Their study places a strong emphasis on generalization and resilience.

Rana et al. (2022) performed a broad analysis in their paper to offer an overview of the research efforts in Deepfake detection, synthesizing 112 published documents from 2018 to 2020 that provided a range of techniques. They classified them into four categories: DL-based approaches, traditional ML-based techniques, statistical techniques, and blockchain-based methods. They also compared the detection capabilities of the various algorithms across different datasets and concluded that DL-based methods outperform other methods in Deepfake detection.

When these publications were evaluated, it was determined that none of them employed the SLR strategy, and none of the deepfakes analysis articles had all deepfakes classifications for examining DL methods. Also, according to the findings, most mentioned publications either focused on a certain component of deepfake administration or described a specific type of data collection. Furthermore, most of these surveys included no comparison examination and only looked at less than 20 articles. Our paper focuses on peer-reviewed publications that propose DL algorithms for deepfake detection tasks, including image, video, sound, and hybrid detection. As a result, we use a slightly different method than most articles. Table 2 includes a summary of related works. Also, the methodology of the research will be described in the following section.

4 | METHODOLOGY OF RESEARCH

In the previous section, we discussed some related works that looked into deepfake detection techniques. This section uses the SLR approach (Esmailiyan et al., 2021; Vahdat & Shahidi, 2020) to comprehend deepfake detection better. The SLR is a comprehensive examination of all studies on a particular topic. This section provides a detailed look at how DL approaches are employed in deepfake detection. The validity of the study selection procedures is next examined. The following subsections describe the search procedure, including research questions and selection criteria.

4.1 | Question formalization

The study's main goals are to identify, distinguish, assess, and evaluate all relevant publications found in deepfake detection methods from DL. An SLR can be used to explore the aspects and features of the methodologies to attain the aims specified before. Another goal of SLR is to understand better the major concerns and problems that this industry encounters. A few Research Questions (RQs) that have been defined are as follows:

RQ 1. What are the applications of DL in deepfake detection?

This question was addressed in Section 1.

RQ 2. What are DL methods in deepfake, and what usages do they have?

This question was answered in Section 2.

RQ 3. Is there any study that has been published as a review article in this field? What distinguishes this article from past research?

This question was answered in Section 3.

RQ 4. What are this area's leading issues and unanswered problems?

Section 5 will present the answers to this topic, while Section 7 will present the open problems.

RQ 5. How can we find the article and choose the deepfake detection DL methods?

This is dealt with in Section 4.2.

TABLE 2 Compilation of related works.

Researchers	Contributions	Scope	Advantage	Weakness
L. Verdoliva (2020)	Presenting an overview of contemporary manipulation techniques	Fake media	• Deepfake's backstory is presented. • Issues and possible solutions are explored.	• There has not been any in-depth review of the articles.
Tolosana et al. (2020)	Examining face-altering techniques	Image deepfake detection	• Different criteria for evaluating articles are taken into account.	• It is unclear how articles are chosen for review.
Mirsky and Lee (2021)	Providing deepfake creation and detection services	Deepfake in general	• Challenges and potential guidance are discussed.	• It is unclear how articles are chosen for review.
Castillo Camacho and Wang (2021)	Examining DL-based image forensic methods	Image forensic	• Taking into account all aspects of the criteria for image forensics.	• It is unclear how articles are chosen for review.
P. Yu et al. (2021)	Focusing on deepfake video detection, its history, current research, and plans.	Deepfake video	• An in-depth description of future work. • In-depth examination of datasets.	• There is no comparison between the articles.
(Rana et al., 2022)	Demonstrating several cutting-edge deepfake algorithms.	DL-ML and statistical models	• There is a comparison between the articles.	• There is no discussion of all kinds of deepfake applications.
Ours	Providing a comprehensive review of the literature on deepfake detection techniques based on DL-based algorithms	DL-ML methods in the video, image, audio, and hybrid multimedia detection	• An in-depth description of future work. • In-depth examination of datasets. • Challenges and potential guidance are discussed.	• Papers published before 2018 are not allowed.

RQ 6. How can the DL mechanisms in multimedia deepfake detection be classified? What are some of their examples?

The answer to this question can be found in Section 5.

RQ 7. What procedures do the researchers employ to conduct their research?

This question is answered in Sections 5.1–5.4.

4.2 | The process of article selection

This study's article search and selection process is divided into four sections. This is depicted in Figure 2. Table 3 shows the keywords and terms used to search the publications in the first step. The articles in this collection are the result of an electronic database search. Some of the used electronic databases are Google Scholar, Scopus, IEEE, ACM, Springer, Elsevier, Emerald Insight, Taylor & Francis, Wiley, Peerj, and MDPI. The other items uncovered are journals, conference articles, books, chapters, notes, technical studies, and special issues. The first stage delivered 651 articles. In addition, the distribution of articles by the publisher is depicted in Figure 3.

The ultimate number of articles to study is determined in Stage 2 through two processes. The articles are initially evaluated (stage 2.1) using the criteria outlined in Figure 4. There are currently 254 articles left. In Figure 5, the publisher's distribution of articles is illustrated. Besides, stage 2.2 excludes review articles; at the previous stage, 8.87% of the remaining 254 articles were review articles. IEEE publishes the majority of research publications (21%). Springer

and Elsevier publish the majority of review articles (5.54%). There are currently 210 articles available. In Stage 3, the titles and abstracts of the articles were reviewed. The articles' methodology, assessment, discussion, and conclusion have all been double-checked to ensure that they are relevant to the study. At this time, 88 articles have been selected for further review. Finally, 34 articles were picked to review and examine the other publications since they matched the tight criteria. The distribution of the chosen articles by their publishers is depicted in Figure 6. IEEE publishes the majority of the selected articles. Wiley, IGI Global Publishing, and Tech Science Press have the lowest. The number of publications published in 2021 and 2020 is equal (50%, 16 articles). The journals that publish the publications are shown in Figure 7. The IEEE Access journal publishes the most articles (9.4%, three articles). Table 4 also lists the

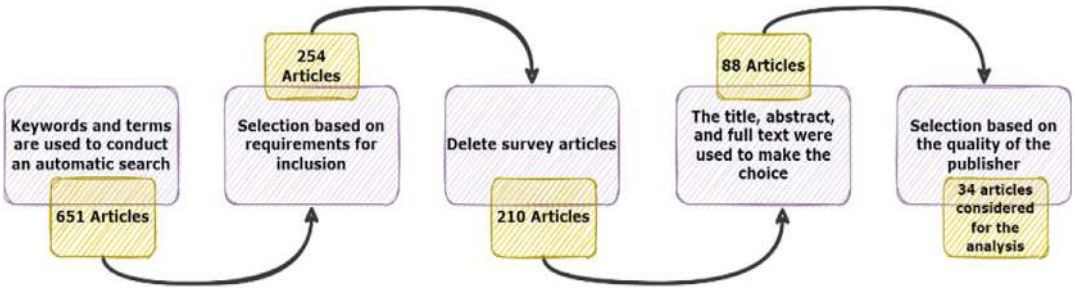

FIGURE 2 The steps involved in the article search and selection procedure.

TABLE 3 Keywords and search phrases.

S#	Search terms and keywords
S1	"Deep learning" and "Deepfake detection"
S2	"Machine learning" and "Deepfake detection"
S3	"Deep learning" and "Image forensic"
S4	"Neural network" and "Deepfake detection"
S5	"AI methods deepfake detection" or "Artificial intelligence deepfake"
S6	"Deep transfer learning" and "Deepfake"
S7	"Transfer learning" and "Deepfake detection"

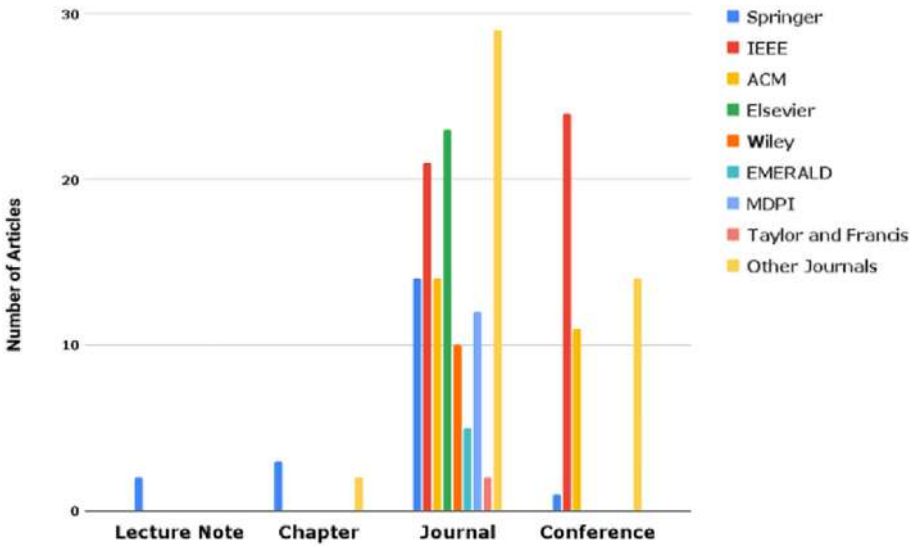

FIGURE 3 The first stage is the distribution of the articles by the publisher.

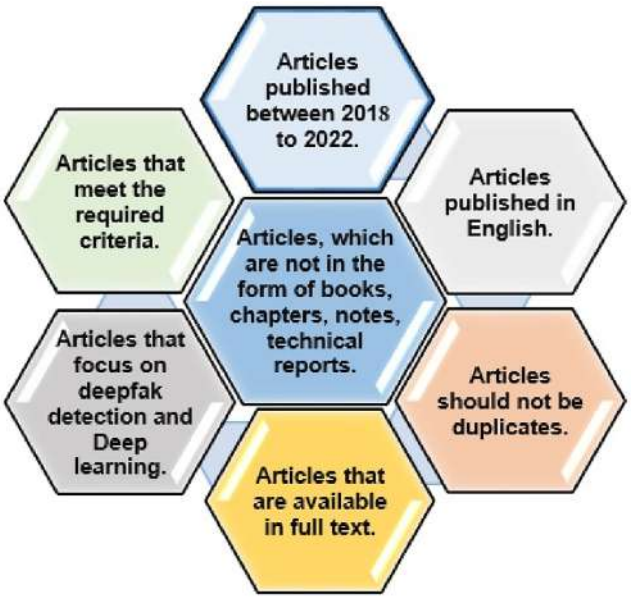

FIGURE 4 Inclusion criteria for the paper selection process.

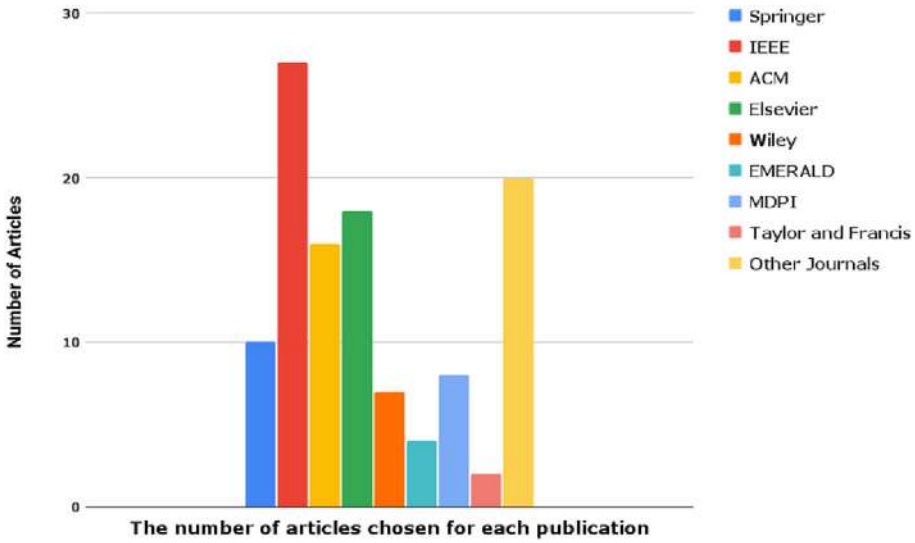

FIGURE 5 The articles are provided by the publishers at Stage 2.1.

chosen articles' specifications (updated on July 19, 2021). In addition, in the following part, the deepfake detection mechanism is explored in-depth, and its properties are discussed.

5 | DEEP FAKE DETECTION MECHANISMS

In Section 4, we explored how we choose articles and the factors that are significant to us, and then we listed the articles chosen and appraised based on their merits. In this section, the DL approaches for identifying deepfake, and associated circumstances are discussed in this section. In this part, 32 articles will be discussed, all of which fulfill our selection criteria. First, we divide the techniques into four categories based on their intended use: Figure 8 displays the suggested taxonomy of DL-deepfake detection techniques, including image, video, sound, and hybrid multimedia content.

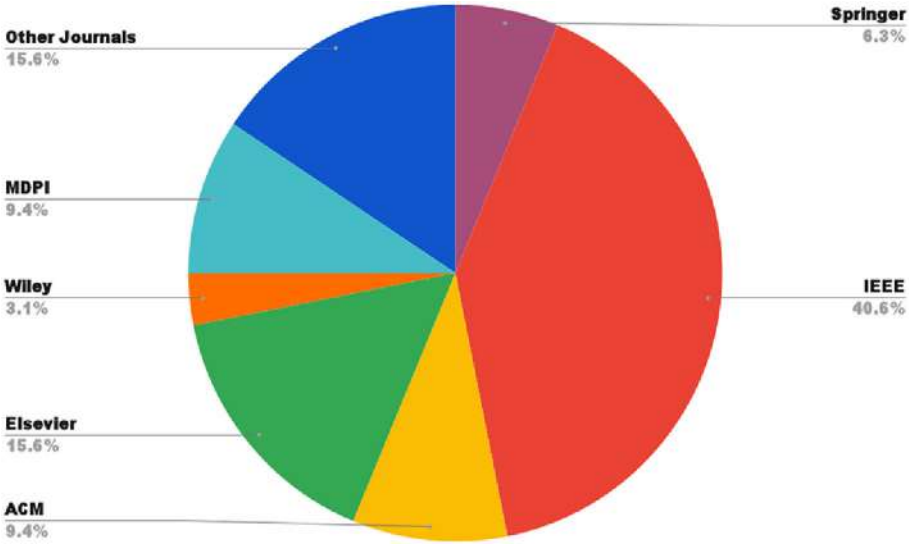

FIGURE 6 The distribution of the selected articles by the publishers.

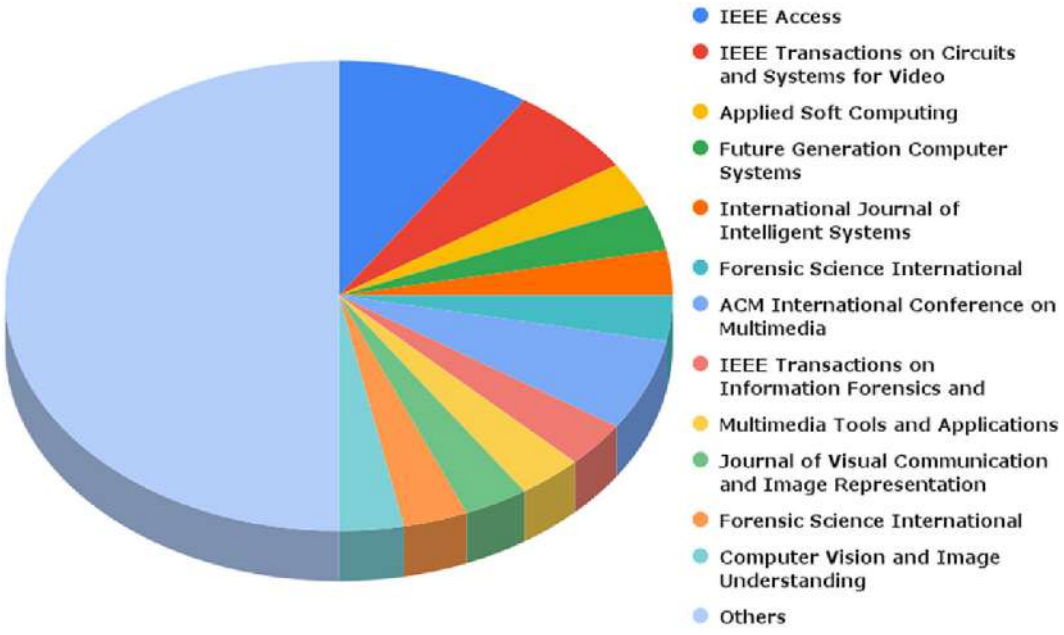

FIGURE 7 Distribution of the selected articles based on journals.

5.1 | Fake image detection

As described in Section 2, Fake face image detection is the most difficult challenge in the field of image forgery detection. Phony photos can construct fake identities on social media platforms, allowing for the illicit theft of personal information. The fake picture generator, for example, can be used to create photos of celebrities with inappropriate content, which might be dangerous. This section will look at detecting fake images using DL approaches. Deepfake's development significantly lowers the bar for face fabrication techniques. Deepfake, in particular, uses GANs to substitute an original image's face with another person's face. Because the GAN models were trained on 10 of 100 of photos, they are more likely to produce realistic faces that can be accurately spliced into the primary image. The image may be made more real by using appropriate post-processing. Furthermore, with the growth of DL, face-swap innovation has been employed in various settings, encompassing privacy protection, multimedia synthesis, and other novel applications (M. Zhang, Chen, et al., 2021; M. Zhang, Chen, et al., 2020). So, W. Zhang, Zhao, et al. (2020) investigated a comprehensive DL-based counterfeit feature

TABLE 4 The chosen articles' specifications (updated on 15 January, 2022).

Publisher	Author	Year	Citation	JCR based 2021	Scopus based 2021	Journal name	Country of journal	H-index based 2021
MDPI	W. Zhang, Zhao, and Li (2020)	2020	7	Q2	Q2	Entropy	Switzerland	74
Elsevier	Lee et al. (2021)	2021	10	Q1	Q1	Applied Soft Computing	Netherlands	143
Elsevier	Guo et al. (2021)	2021	8	Q2	Q1	Computer Vision and Image Understanding	USA	138
IEEE	Guarnera et al. (2020)	2020	14	Q2	Q1	IEEE Access	USA	127
IEEE	I.-J. Yu et al. (2020)	2021	4	Q2	Q1	IEEE Access	USA	127
Elsevier	J. Yang et al. (2021)	2021	2	Q1	Q1	Future Generation Computer Systems	Netherlands	119
MDPI	Hsu et al. (2020)	2020	53	Q2	Q2	Applied Sciences	Switzerland	52
IEEE	(Güera & Delp, 2018)	2018	608	-	-	IEEE International Conference on Advanced Video and Signal Based Surveillance	USA	-
Elsevier	X. H. Nguyen et al. (2021)	2021	5	Q2	Q1	Forensic Science International	Ireland	120
IEEE	Jung et al. (2020)	2020	36	Q2	Q1	IEEE Access	USA	127
-	Karandikar et al. (2020)	2020	2	-	-	International Journal of Advanced Trends in Computer Science and Engineering	India	18
IGI Global Publishing	Zhao, Wang, and Lu (2021)	2021	0	-	Q4	International Journal of Digital Crime and Forensics	USA	15
Elsevier	Z. Xu et al. (2021)	2021	1	Q2	Q1	Journal of Visual Communication and Image Representation	USA	81
Springer	Kohli and Gupta (2021)	2021	0	Q2	Q1	Multimedia Tools and Applications	Netherlands	70
Hindawi	Chen and Tan (2021)	2021	0	Q4	Q2	Security and Communication Networks	Egypt	43
IEEE	A. Yan, Fan, et al. (2021)	2021	4	Q1	Q1	IEEE Transactions on Circuits and Systems for Video Technology	USA	168
Elsevier	Caldelli et al. (2021)	2021	3	Q2	Q1	Pattern Recognition Letters	Netherlands	57
John Wiley	L. Yan, Yin-He, et al. (2021)	2021	3	Q1	Q1	International Journal of Intelligent Systems	United Kingdom	87
IEEE	Mitra et al. (2020)	2020	3	-	-	International Symposium on Smart Electronic Systems (iSES)	USA	-
IEEE	Suratkar et al. (2020)	2020	1	-	-	International Conference on Computing, Communication and Networking Technologies (ICCCNT)	USA	-

(Continues)

TABLE 4 (Continued)

Publisher	Author	Year	Citation	JCR based 2021	Scopus based 2021	Journal name	Country of journal	H-index based 2021
IEEE	Bonettini et al. (2021)	2021	82	-	-	International Conference on Pattern Recognition	Italy	-
IEEE	Cozzolino et al. (2019)	2019	42	-	-	IEEE/CVF Conference	USA	-
Springer	Borrelli et al. (2021)	2021	2	Q3	Q2	Eurasip Journal on Advances in Signal Processing	USA	88
IEEE	C.-Z. Yang et al. (2020)	2020	8	Q1	Q1	IEEE Transactions on Information Forensics and Security	USA	133
ACM	Mittal et al. (2020)	2020	40	-	-	Proceedings of the 28th ACM international conference on multimedia	USA	-
ACM	R. Wang et al. (2020)	2020	21	-	-	in Proceedings of the 28th ACM International Conference on Multimedia	USA	-
IEEE	Wijethunga et al. (2020)	2020	2	-	-	International Conference on Advancements in Computing (ICAC)	USA	-
MDPI	Khalil et al. (2021)	2021	3	-	Q2	Future Internet	Switzerland	28
IEEE	Chintha et al. (2020)	2020	28	Q1	Q1	IEEE Journal of Selected Topics in Signal Processing	USA	120
IEEE	Kong et al. (2021)	2021	3	Q1	Q1	IEEE Transactions on Circuits and Systems for Video Technology	USA	168
IEEE	Sun et al. (2020)	2020	1	-	-	IEEE International Workshop on Information Forensics and Security	USA	-
IEEE	Nasar et al. (2020)	2020	0	-	-	Recent Advances in Intelligent Computational Systems (RAICS)	USA	-
ACM	Chugh et al. (2020)	2020	16	-	-	Proceedings of the 28th ACM International Conference on Multimedia	USA	-
IEEE	Chan et al. (2020)	2020	3	-	-	IEEE/ITU International Conference on Artificial Intelligence for Good (AI4G)	USA	-

extraction approach that could successfully differentiate between real and fake pictures generated using DL. Their forgery feature extraction approach may disclose face-swap pictures based on DL and error level analysis (ELA), which can differentiate DL-generated facial images successfully. Instead of the original photos, they utilized these ELA images for training the CNN model. Converting the original picture to an ELA image is one way to increase the CNN model's training efficiency. Because the ELA picture has less information than the original image, this could enhance efficiency. The ELA

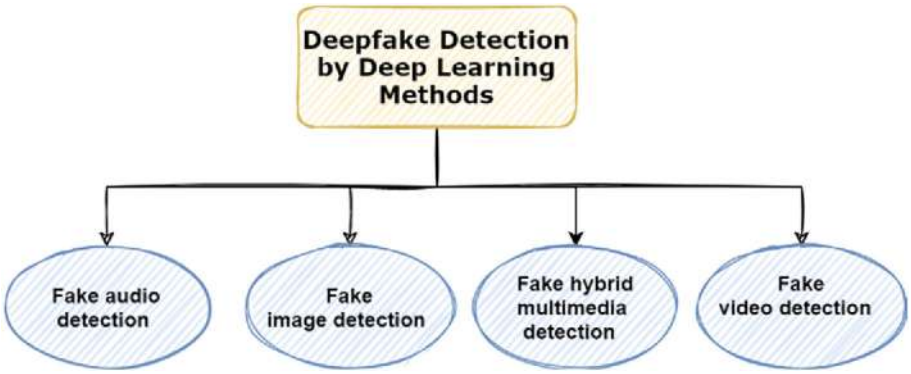

FIGURE 8 The suggested DL-deepfake detection taxonomy separated four distinct methods.

image feature concentrates on the part of the original picture in which the error level exceeds the threshold value. Furthermore, the pixels in the ELA picture are frequently significantly distinct from the neighboring pixels, and the contrast is quite visible; therefore, the image processed by ELA develops the effectiveness of the training CNN model. They trained a CNN model to extract the counterfeit aspects of the ELA photos and then determined whether or not the input image was fake. They examined the approach and demonstrated its usefulness in practice. Because the ELA images created in the previous phases may emphasize the features of the original image when the error level is greater than the threshold value, just two convolution layers are necessary. As a result, it is simpler to isolate counterfeit characteristics and evaluate if the image is authentic. It demonstrates that the characteristics in the picture analyzed by ELA could be effectively employed to identify image authenticity. Additionally, their tests showed that the ELA technique greatly improves the training efficiency of the CNN model. They assessed the strategy and demonstrated its efficacy in practice. The findings showed that the characteristics in the picture processed by ELA might effectively be utilized to determine image authenticity. Additionally, their research shows that the ELA approach may considerably improve the training efficiency of the CNN model. Especially, the amount of processing may be greatly decreased without any loss of accuracy if the needed floating-point computing power is lowered by more than 90%.

Also, handcrafted face manipulation (HFM) images dataset and soft computing neural network models (Shallow-FakeFaceNets) with an effective facial manipulating detection process were presented by Lee et al. (2021). Shallow-FakeFaceNet (SFFN), the neural network classifier model, demonstrates the capacity to recognize false pictures by focusing on altered facial landmarks. The detection process simply uses Red Green Blue (RGB) information to detect fraudulent facial pictures and does not use any metadata, which could be readily altered. The technique performs better in Area Under the Receiver Operating Characteristic (AUROC), gaining 3.99% F1-score and 2.91% AUROC on identifying handcrafted fake facial images, and 93.99% on detecting small GAN-generated fake images, gaining 1.98% F1-score and 10.44% AUROC. The approach is intended to develop an automated defensive system to battle false pictures used during various online services and apps, utilizing their cutting-edge HFM and SFFN. Furthermore, the study offers a variety of test findings that could be used to successfully drive future applied soft computing studies to battle and successfully identify human and GAN-generated false face pictures.

Besides, as preprocessing, Guo et al. (2021) developed a simple yet successful Adaptive Manipulation Traces Extraction Network (AMTEN) component that uses the convolution layer as a predictor to retrieve photo manipulation traces. During the backpropagation pass, the weights are dynamically adjusted. Traces were repeated in successive layers to optimize manipulation traces. They additionally, created a false face detector, AMTENnet, by combining AMTEN with CNN. AMTEN's manipulation traces are put through CNN to develop different discriminative characteristics. A set of tests were carried out in which many common post-processing processes were chosen to replicate actual forensics in complicated circumstances. Their findings demonstrate that AMTENnet obtained higher detection accuracy as well as desired generalization capabilities. Also, AMTENnet outperformed other approaches in terms of average detection accuracy on the Hybrid Fake Face (HFF) dataset by 7.61%. Indeed, AMTEN is the primary advantage here, as it produces superior residual extraction than Constrainedconv and Spatial Rich Model (SRM). It is indeed worth noting that AMTEN can be used as a simple residual predictor for additional face forensic procedures.

Also, Guarnera et al. (2020) presented the completion of a previous study on deepfake image analysis. An expectation–maximization technique is used to extract the Convolutional Traces (CTs): a unique fingerprint that can be

used to determine if a photo is a deepfake and the GAN architecture that produced it. The retrieved CT is a fingerprint with strong discriminative power and resistance to attacks and is independent of high-level picture ideas. The results also revealed that a simple and fast-to-compute method could outperform the state-of-the-art. Furthermore, CT is connected to the image production process, and better performance could well be achieved by rotating input pictures to locate the most significant direction. With an overall classification accuracy of more than 98% and deepfakes from 10 distinct GAN architectures included in pictures of faces, the CT proves to be trustworthy and independent of image semantics. Lastly, testing on deepfakes created by the FACE application, which achieved 93% accuracy in the deepfake detection task, proved the efficiency of the method in a real-world setting.

Besides, I.-J. Yu et al. (2020) introduced the Manipulation Classification Network (MCNet) to categorize the various manipulation techniques used on Joint Photographic Experts Group (JPEG) compressed pictures by utilizing several domain characteristics. To optimize the efficiency of manipulation classification for JPEG compressed images, the MCNet uses spatial, frequency, and compression domain data. They created domain-specific preprocessing and network topologies. The results indicate that the MCNet outperforms existing manipulation detection networks as well as numerous low-level vision networks. The suggested network is excellent for fine-tuning current forensics tasks such as deepfake detection and JPEG picture integrity verification since it considers many alternative alteration methods. The results specify that the MCNet outperforms existing manipulation detection networks as well as numerous low-level vision networks. The suggested network is excellent for fine-tuning current forensics tasks such as deepfake detection and JPEG picture integrity verification since it considers many alternative alteration methods.

J. Yang et al. (2021) recognized the tiny textural variations between the actual and fake facial images and used the picture saliency approach to demonstrate them. Depending on this finding, they use the enhanced guided filter to do picture preprocessing on all false and real photos, intending to improve the texture artifacts included in the face modification image, which we term guided characteristics. However, this distortion cannot be caught visually; these magnified texture variations will be learned using a Resnet18 network that can constantly downsample and resample to identify actual and false facial pictures accurately. Studies showed that their technique has the best detection accuracy in both whole picture and face image training, which is at the cutting edge of current research. Simultaneously, due to the extremely expanded texture characteristics, the network could record differences rapidly and correctly, achieving convergence quicker, and reducing training time. Nevertheless, their approach has certain flaws, which is a frequent issue in investigating real and fake face recognition. Network model training still needs enormous amounts of data to obtain reasonable accuracy. Rather than building a universal detection network, training a new authentication network is still required for the unknown face tampering approach.

Finally, Hsu et al. (2020) proposed a fake feature network-based paired learning method to recognize fake faces and generic pictures created by state-of-the-art GANs. They can learn middle- and high-level and discriminative false features by combining cross-layer feature representations. The reduced DenseNet is then transformed into a two-streamed network structure that can accept paired data as input. The suggested network is then trained to employ paired learning to differentiate between the characteristics of the fake and real photos. Lastly, a classification layer is added to the suggested common fake feature network to determine if the input picture is real or false. Their paired learning system permits fake feature learning, allowing the trained fake image detector to recognize the false picture created by a new GAN even though it was not included in the training phase. The findings showed that their technique beat other state-of-the-art methods in terms of precision and recall rate. Table 5 discusses the deepfake image applications used in deepfake and their properties.

5.2 | Fake video detection

As described in Section 2, one of the most difficult problems is detecting video deepfakes. To create a very advanced face-swap movie using deepfake, only one GPU and a vast amount of training data are required. Several technology enthusiasts have posted a large number of live-performance films on the short video platform that replace regular people's faces with synthetic stars' faces, resulting in a topic to be debated. Concerns have been raised about this technique. Before this technique, it was widely assumed that videos were dependable and that they could even be used as video evidence in multimedia forensics. In the digital age, video deepfake technology has posed a threat to public confidence. Some even believe that this innovation would stifle societal growth. Thanks to the rapid rise in the availability of open-source datasets, significant advancements in the study of topics like GANs, and significant technological developments in the field of high-speed computing, creating and manipulating fake videos has become a relatively simple task in

TABLE 5 Image deepfake techniques' approaches, attributes, and features.

Authors	Main idea	Advantages	Research challenges	Security used?	Dataset	Simulation environment	Using TL?	Method	Usage?
Zhang, Zhao, and Li (2020)	Using a binary classifier trained by a CNN.	• The achieved accuracy was 97% • The AUC stood at 97.6%	• Poor robustness	No	Milborrow University of Cape Town dataset	Python	No	CNN	Image
Lee et al. (2021)	Suggesting an approach with an effective end-to-end false face detection pipeline that can identify fake face pictures.	• Obtaining 72.52% AUROC • Obtaining 93.99% accuracy on difficult low-resolution pictures	• Robustness is not taken into account	No	HFM dataset	Keras Python DL library	No	GAN	Image
Guo et al. (2021)	Providing a preprocessing module named AMTEN for face image forensics.	• AMTENnet achieves an average accuracy of up to 98.52% • Achieves desirable preprocessing	• Detector's robustness is low	No	HFF dataset	Caffe framework	No	CNN	Image
Guarnera et al. (2020)	Presenting an expectation–maximization method trained to identify and extract a fingerprint.	• Obtaining a 93% accuracy rate • In a real-world scenario, efficacy is demonstrated • High robustness	• High delay • High energy consumption	No	The FACE APP, a dataset, and CELEBA images.	Keras	No	Expectation–maximization + CNN	Image
I.-J. Yu et al. (2020)	Presenting the MCNet to utilize multi-domain spatial, frequency, and compression domain characteristics.	• High robustness • High accuracy	• High complexity • High energy consumption	No	ALASKA dataset	PyTorch	Yes	CNN	Image
J. Yang et al. (2021)	Using the image saliency to determine the texture depth and pixel difference between actual and fake facial images.	• Detection accuracy is 0.9990	• Poor robustness	No	Faceforensics++ dataset	Pytorch	No	CNN + simple linear iterative clustering	Image
Hsu et al. (2020)	Presenting an image detector comprised of an enhanced DenseNet backbone network and Siamese network architecture.	• Achieving a modest level of precision and recall	• High energy consumption • Low robustness	No	CelebA	Not mentioned	No	Pairwise learning	Image

terms of the cost of creating a manipulated sequence. By superimposing a politician's face over the face of a target actor, a deepfake film might be used to slander and promote false political propaganda. Also, it could then be used to embarrass, shame, or cause a schism among major political institutions, jeopardizing national peace by disrupting the delicate balance of peace between people, governments, and nations worldwide. So, detecting such fraudulent videos has become incredibly important. More academics have begun to look at detecting techniques and the distinctions between authentic and phony films. Traditional techniques and DL approaches have been the most commonly employed. Additionally, living body recognition in face recognition and multimedia forensics can provide options. It is thought that DL can currently build realistic phony faces, detect and investigate invisible forgery traces, and recognize forgery films. DL approaches, as opposed to standard picture forensic techniques, incorporate feature extraction and feature classification into a network structure and achieve an end-to-end effective automatic feature learning classification methodology (Zheng, Liu, Ni, et al., 2021). On the other hand, typical image forensic approaches are typically unsuitable for video since compression significantly reduces data quality. As DL improves in digital forensics, it upends standard image forensic methodologies. As a result, automated approaches for detecting deepfake movies have become extremely essential, given the global influence and the extent to which these videos could harm peace and security. In this section, we will look at how to detect fake videos using DL approaches. So, Güera and Delp (2018) presented a temporal-aware method for detecting deepfake films automatically. They demonstrated a temporal-aware system for detecting deepfake videos automatically. They demonstrated a robust trainable recurrent deepfake video identification system. For processing frame sequences, the suggested system employs a convolutional LSTM structure. A convolutional LSTM has two important components: (1)_ CNN for extracting frame features. (2) Temporal sequence analysis using LSTM. Given an unknown test sequence, they retrieved a collection of features produced by the CNN for each frame. The characteristics of numerous consecutive frames were then concatenated and sent to the LSTM for analysis. Ultimately, they calculated the odds of the sequence being a deepfake or a non-manipulated video. These attributes are then used to train an RNN to determine whether or not a video has been manipulated. They tested their strategy against a vast collection of deepfake films gathered from various video websites. They demonstrate how, despite a minimal architecture, the system can obtain competitive outcomes in this task. Their experiments with a large collection of digitally altered videos revealed that a simple convolutional LSTM structure could indeed effectively forecast whether a video has been manipulated or not with as little as 2 s of video data.

Also, X. H. Nguyen et al. (2021) proposed creating three-dimensional (3D) pictures as well as a 3D CNN model that can learn characteristics of spatial and temporal dimensions from the 3D images. They used 3D convolution kernels to create a deep 3D CNN to learn spatio-temporal characteristics in short consecutive frame sequences to recognize deepfake videos. In their method, the feature maps in the convolutional layers are linked to the successive frames of the previous layer. As a result, it can collect information about faces in spatial and temporal dimensions in a series of frames. Studies showed that their suggested model performs very well on deepfake FaceForensic and VidTIMIT datasets. The suggested model excelled on both datasets by more than 99% due to the elevated videos.

Besides, Jung et al. (2020) devised a way to detect deepfakes created by the GANs model by analyzing substantial variations in eye blinking, a spontaneous and unconscious human behavior. Blinking patterns differ depending on a person's gender, age, cognitive activities, and time of day. Consequently, their system detected these changes utilizing ML, various methods, and a heuristic way to verify the integrity of deepfakes. Their method, which was developed utilizing prior experiments' findings, constantly showed a substantial possibility of validating the integrity of deepfakes and regular videos, correctly recognizing deepfakes in seven of eight videos (87.5%). Nevertheless, one study drawback is that blinking is also associated with mental illness and dopamine activity. The integrity verification may not be relevant to persons suffering from mental disorders or who have issues with nerve conduction pathways. Conversely, because cyber-security attacks and defenses are always evolving, this may be enhanced by a variety of techniques. The suggested technique points towards a new way to overcome the limitation of integrity verification algorithms.

However, Karandikar et al. (2020) suggested technique employs TL on a CNN model to train the dataset and focuses on face modification to identify counterfeit. Their classifier is built on a VGG-16 model, which is then supplemented by batch normalization, dropout, and a proprietary two-node dense layer. The two nodes in the suggested architecture's last dense layer are being used for two final classes (real and fake). The batch normalization layer is used to normalize and scale the inputs from the preceding layer. Dropout is also included to prevent overfitting and improve weight optimization. At each epoch, the dropout layer will arbitrarily transmit some nodes as off from the preceding layer. It will result in better training because this layer introduces some unpredictability during weight updates. Their method performed well and can effectively gather features necessary for additional processing to test for deepfakes. The suggested model's accuracy falls with low-quality pictures, and with medium-quality videos, the accuracy must be enhanced

further by utilizing mixed models for training on temporal parameters. As a result, better training will result from a better dataset of higher quality.

In addition, Z. Zhao, Zhou et al. (2021) presented a method for detecting face fraud in deepfake videos that is more accurate. This has offered a Multi-layer Fusion Neural Network (MFNN) to catch diverse artifacts caused by deepfake forgeries in three distinct layers. The technique offered a deepfake detection method by combining feature maps generated from multiple levels in the detection network. They constructed shortcut connections for feature maps from various layers, reduced their size, and sent them directly to the final layer to categorize the extracted features. The features maps collected by the shallow, medium, and deep layers are output and fused simultaneously during categorization. The FaceForensics++ dataset is also used to train and test the network. The findings demonstrate that the novel detection approach beats the previous methods in detecting deepfake videos. MFNN has significantly increased accuracy in identifying low-quality deepfake movies, which is generally more challenging for previous techniques.

Z. Xu et al. (2021) proposed a perspective on the faces of numerous video frames as a set to examine facial altered video recognition and a unique framework Set CNN (SCNN). Their system provides three examples: t-MesoNet, t-XceptionNet, and t-XceptionNet. Also, comprehensive studies show demonstrated the approach outperforms prior techniques and may be used as a foundation for combining with other backbone networks. Nevertheless, they discovered that a stronger backbone network could produce superior outcomes. As a result, seeking a better backbone network could be the next step in the right way, and as well as the set reduction technique is critical for SCNN. The three distinct set reduction techniques are straightforward and intuitive in this case.

Also, Kohli and Gupta (2021) presented a strategy for exploiting the frequency domain characteristics of face forgery. A Frequency CNN (FCNN) is used in their technique to evaluate and categorize clean and counterfeit faces. To assess the efficacy of the FCNN, the FaceForensics++ dataset is employed. Studies showed that FCNN detects forgeries efficiently in actual circumstances, including such high and low video quality. Furthermore, the display of activation maps demonstrates that FCNN learns different frequency characteristics for deepfake, Face2Face, and FaceSwap manipulation methods. Among all other facial modification techniques, the FCNN detects deepfakes with the greatest recall of 0.9256, 0.8639, and 0.8399 for raw, c23, and c40, respectively. This technique is also tested on a Celeb-DF (v2) dataset as well as an automated FaceForensic benchmark. The findings demonstrate the usefulness of the suggested approach for detecting face manipulation.

Likewise, Chen and Tan (2021) introduced feature transfer, a two-stage deepfake detection approach that relies on unsupervised domain adaption. The feature vectors derived from CNN are utilized in Backpropagation Based on Domain-Adversarial Neural Network (BP-DANN) for adversarial TL, leading to higher efficiency than end-to-end adversarial learning. The face detection network is initially utilized in the preprocessing stage to extract the face region of the video frame, which is then enlarged by 1.2 times to crop and save the face picture. The CNN is trained on the large-scale deepfake dataset, which a third party provides. Lastly, the facial images are input into the CNN to obtain the 2048-dimensional feature vectors. The collected feature vectors are stored to easily load them into the BP-DANN for the training of unsupervised domain adaptive (B. Xie et al., 2023). Furthermore, the features extraction CNN pretrained on a big deepfake dataset might well be utilized to extract additional transferrable feature vectors, reducing the gap between the source and target throughout unsupervised domain adaptive training.

Also, A. Yan, Yin-He, et al. (2021) worked on compressed deepfake movies with a low-quality factor to cater to scenarios commonly seen on social media. In reality, compressed videos are widespread on social media platforms like Instagram, Wechat, and Tiktok. As a result, determining ways to detect compressed deepfake films becomes a critical challenge. In addition, they use the frame-level stream with a low complexity network and prune the model to avoid fitting the noise because of the noise introduced by compression. The temporality-level stream is used to extract the inconsistency between frames to discover the temporal characteristics of compressed movies. The two streams extract compressed video's frame-level and temporal-level information. They tested their suggested two-stream technique on deepfakes, FaceSwap, Face2Face, NeuralTextures, and Celeb-DF datasets, and the results outperformed previous work. The accuracy of cross-compression detection findings demonstrates that their approach is robust to compression factors.

Furthermore, Caldelli et al. (2021) suggested optical flow field dissimilarities differentiate between deepfakes and real videos using CNN. Their research is based on the usage of CNNs that have been trained to detect potential motion dissimilarities in the temporal structure of a video clip using optical flow fields. The test results produced on the FaceForensics++ dataset are intriguing and demonstrate that this feature is well suited to extract distinctive characteristics between the fake and real instances, particularly when dealing with the tough cross-forgery scenario. Moreover, it is demonstrated how this technique leverages discrepancies on the temporal axis, which improves their efficiency when integrated with well-known state-of-the-art frame-based approaches.

L. Yan, Yin-He, et al. (2021) proposed a lightweight 3D CNN. The CT module is intended to extract features with fewer parameters at a higher level. 3D CNNs are used as a spatial-temporal module to merge spatial information in the time dimension. However, SRM characteristics are collected from the input frames to suppress frame content and emphasize frame texture, allowing the spatial-temporal module to function better. Also, the CT module seeks to extract deep-level features from the outcome of the spatial-temporal module using as few parameters as possible. As previously stated, 3D CNNs have larger parameters than traditional two-dimensional (2D) CNNs, which harms convergence and generalization capabilities. The CT module is intended to replace the standard 3D convolution layers to minimize the number of parameters. The results demonstrate that their network has fewer parameters than other networks and outperforms existing state-of-the-art deepfake detection techniques on major deepfake data sets. Because their methodology overcomes the problem of high deployment consumption while retaining good detection performance, it is clear that deepfake detection on edge devices will be used soon.

Also, Mitra et al. (2020) demonstrated a DL-based method for detecting deepfake videos on social media with excellent accuracy. They categorize and modify footage using a neural network-based technique. A model is developed that consists of a CNN and a classifier network. The CNN modules were chosen from three different structures: XceptionNet, InceptionV3, and Resnet50, and a comparison study was conducted. They analyzed three existing CNN modules and decided on Xception net as the feature extractor for the most accurate model, along with the proposed classifier. They used intermediate compression to train the network and achieved good accuracy even in a high-loss environment. Even without training with extremely compressed videos, their approach is the key to achieving great accuracy. The amount of frames retrieved from the movie determines the algorithm's complexity.

Suratkar et al. (2020) developed a system for detecting false videos that use a CNN architecture and the TL methodology. Their approach uses a CNN to collect features from each movie frame to build a binary classifier that can efficiently distinguish between real and altered videos. The process is tested on many deepfake films culled from diverse datasets. So, the findings demonstrate that using TL to develop a relatively robust model for deep fake detection is doable. With the notion of TL at its helm, the technique employed allows any model to be trained significantly faster, resulting in a significant reduction in training time. TL also makes it easier to transfer valuable knowledge from one work to another, which was demonstrated in this situation. To obtain generality in mind, their paper aimed for maximal generality by combining datasets collected using various methodologies from various sources.

Bonettini et al. (2021) handled the difficulty of detecting face alteration in video sequences by means of recent facial manipulation techniques. They investigated the assembly of several trained CNN models. In the suggested method, several models are created by beginning with a base network (EfficientNetB4) and employing two distinct concepts: (i) attention layers and (ii) siamese training. On two publicly available datasets with over 119,000 videos, they demonstrated that merging these networks yields promising face modification detection results.

Finally, Cozzolino et al. (2019) developed a neural network that enhanced the model-related traces concealed in a video, extracting a form of camera fingerprint known as video noiseprint, inspired by recent work on photos. The net is trained using a Siamese method on pristine films, limiting distances between same-model patches and maximizing distances between unrelated patches. Experiments demonstrated that systems based on video noiseprints work well in important forensic tasks, including camera model identification and video forgery localization, with no requirement for previous information or fine-tuning.

However, Table 6 lists the deepfake video applications that use DL techniques and their features.

5.3 | Fake sound detection

Section 2 describes voice deepfake detection as a mostly unexplored field. Machine-generated voices mostly inhabit the daily lives. People are more likely to control daily tasks using speech as technology becomes more automated. Virtual assistants such as Google Assistant, Alexa, Siri, Bixby, and others increasingly use machine-generated or synthetic voices. Despite having access to such technologies, celebrities, politicians, and other well-known individuals are victims. The main source of concern is a cutting-edge technique known as deepfake, which primarily uses GAN to generate synthetic audio to impersonate actual people. Because of the numerous ways to generate synthetic speech, the challenge of synthetic speech detection is extremely difficult. Nevertheless, synthetic speech may be created using basic cut-and-paste methods that accomplish waveform concatenation, which is often accessible as open-source toolkits. It may also be produced using the voice stream's source or bandpass filter model (K.-D. Xu et al., 2022). Lately, numerous CNNs-based approaches for generating synthetic audio have been presented. These provide incredibly realistic results

TABLE 6 Video deepfake techniques' methods, attributes, and features.

Authors	Main idea	Advantages	Research challenges	Security used?	Dataset	Simulation environment	Using TL?	Method	Usage?
Güera and Delp (2018)	Presenting a temporal-aware method for detecting automatically.	• High accuracy • High robustness	• High complexity	No	Deepfake-TIMIT and FaceForensics++ dataset	Python	No	CNN + RNN	Videos
X. H. Nguyen et al. (2021)	Using deep 3D CNN that collects spatio-temporal characteristics from a sequence of brief consecutive frames.	• The findings perform with binary accuracy rates of 99.4 and 94.5 on the high-quality and low-quality data sets, respectively	• High complexity • High energy consumption	No	FaceForensicsþþ and VidTIMIT datasets	Python	No	CNN	Videos
Jung et al. (2020)	Providing a method for detecting deepfakes created by the GANs model.	• Accuracy rate of 87.5%	• The number of eye blinks was linked to a mental disorder that was linked to dopamine activity	No	Their deepfake dataset	Python	No	GAN	Eye blinking pattern
Karandikar et al. (2020)	Using a pre-trained CNN to collect face characteristics to extract hidden features.	• High accuracy • Low complexity	• High complexity	No	Celeb-DF Dataset	Keras for python	Yes	CNN	Videos
Z. Zhao, Zhou et al. (2021)	Proposing the MFNN technique for capturing artifacts at various levels.	• Greater benefit in identifying low-quality deepfake videos • High accuracy	• High energy consumption • High complexity	No	FaceForensics++	TensorFlow + Keras DL library	No	CNN	Videos
Z. Xu et al. (2021)	Providing an SCNN approach for detecting face modification techniques deepfakes.	• High accuracy • Low delay	• High complexity • High energy consumption • Low robustness	No	Deepfake-TIMIT dataset, FaceForensics++ dataset, and DFDC Preview dataset	Python	No	CNN	Videos
Kohli and Gupta (2021)	Developing a frequency-based CNN method.	• Strong robustness	• High energy consumption • High delay	No	FaceForensics++ and Celeb-DF (v2) dataset	MATLAB R2018b	No	Frequency CNN	Videos
Chen and Tan (2021)	Proposing feature transfer, a technique based on unsupervised domain adaptation.	• Solves the overfitting problem	• High energy consumption • High complexity	No	Deepfake-TIMIT + FaceForensics++	Python	Yes	CNN + BP-DANN	Videos

(Continues)

TABLE 6 (Continued)

Authors	Main idea	Advantages	Research challenges	Security used?	Dataset	Simulation environment	Using TL?	Method	Usage?
A. Yan, Yin-He, et al. (2021)	Evaluating the frame-level and temporality-level of compressed deepfake films.	• Robust to the compression factor	• A generic model cannot recognize several forms of manipulated facial videos	No	Celeb-DF and FaceForensics++	PyTorch	No	CNN	Videos
Caldelli et al. (2021)	Proposing an approach that makes use of optical flow fields.	• Improved robustness in a more realistic cross-forgery operational situation	• High energy consumption	No	FaceForensics++	Python	No	CNN	Videos
L. Yan, Yin-He et al. (2021)	Presenting a lightweight 3D CNN for deepfake detection.	• Lower network parameters • Achieved accuracy with SRM is 98.07	• Low robustness • Low reliability	No	FaceForensics, deepfake-TIMIT, DFDC-pre and celeb-DF	Python	No	CNN	Videos
Mitra et al. (2020)	Presenting a CNN-based model and a classifier.	• High accuracy	• High complexity • Low robustness	No	FaceForensics++	Keras with the TensorFlow	Yes	CNN	Videos
Suratkar et al. (2020)	Using a CNN architecture and the TL methodology for exposing such fake videos.	• Strong robustness • Archived accuracy is 98%	• Security risks were not taken into account.	No	FaceForensics++, Google Deep Fake Dataset, and Deep Fake Dataset from Facebook	DLIB for python	Yes	CNN	Videos
Bonettini et al. (2021)	Detecting facial alteration in video sequences, targeting both traditional computer graphics and DL-produced fake movies.	• High accuracy	• Moderate delay	No	FF++ and DFDC datasets	Pytorch	No	CNN	Detecting face alteration
Cozzolino et al. (2019)	Constructing a method that improved the model-related traces hidden in a video.	• High accuracy • High robustness	-	No	Vision dataset	-	No	CNN	Video forensics

that are difficult to distinguish from genuine speech as well as human listeners. The audio anti-spoofing research group has addressed the more broad topic of synthetic speech synthesis detection. Various methods based on either hand-crafted or data-driven feature analysis have been presented in this context. Furthermore, with the recent development of CNN-based algorithms for synthetic audio synthesis, many of the earlier detectors are sure to fail. In this section, we will look at how to detect fake voices using DL approaches. So, the approach developed by Borrelli et al. (2021) is centered on short-term and long-term prediction traces. Considering a spoken audio recording, the purpose is to determine if the speech is synthetic—for example, created by some methods, or genuine, that is, belongs to a real human speaker. They focused on both closed-set and open-set settings. The suggested approach determines whether the voice is genuine or synthetic in the closed-set situation. It also recognizes which algorithm was utilized to produce the voice in the case of synthetic speech. In the open-set scenario, the suggested technique can also detect whether a phony voice was made using a previously unseen methodology. They integrated a set of characteristics inspired by the speech processing literature to collect traces from various types of synthetically created voice recordings. They presented, in particular, a set of characteristics based on the notion of modeling speech as an auto-regressive process. They examined numerous auto-regressive orders at the same time to construct this feature set. A set of features is derived from the audio under investigation; a supervised classifier is trained to handle the extracted features' classification issue. They offer a set of characteristics based on the notion of modeling speech as an auto-regressive mechanism. This characteristic framework is defined by taking into account many distinct auto-regressive orders at the same time. Furthermore, they explored the impact of integrating the suggested characteristics with the bicoherence-based features to see whether they enhance each other. The suggested features are driven by the widespread use of the source-filter paradigm to analyze and synthesize speech signals. In essence, they suggest extracting various statistics derived by filtering the signal under investigation using short-term and long-term predictors while taking into account different prediction phases. The findings indicate that their results are more accurate than other solutions. In certain situations, combining all of the qualities is advantageous.

Also, C.-Z. Yu et al. (2020) proposed a lip-based visual speaker authentication technique to protect against human imposters as well as deepfake threats. This technique could distinguish deepfake attacks despite previous knowledge of the video-generation process on the basis that the attackers have insufficient information about the customer and cannot perfectly recreate the client's talking habit while speaking arbitrary immediate texts. Speaker Authentication Dynamic Talking Habit Network-based (SA-DTH-Net) is a newer DNN that extracts information about the customer's specific talking pattern to distinguish the customer's lip image sequence from actual imposters and deepfake forgeries. By combining all word-level authentication results produced from SA-DTH-Net, the ultimate authentication result for a speaker reciting a random prompt text may be achieved. Also, the analysis indicates that their strategy can effectively reject the majority of deepfake attempts created by various manipulation methods. As a result, their technique might be a viable option for universal deepfake detection in Visual Speaker Authentication (VSA) systems.

In addition, Mittal et al. (2020) demonstrated a learning-based technique for distinguishing between genuine and phony deepfake multimedia content. They collect and evaluate the similarity between the two auditory and visual modalities from within the same video to optimize information for training. They also compare and extract effective clues relating to observed emotion from the two modalities inside a video to determine if it is "genuine" or "fake." Motivated by the Siamese network architecture and the triplet loss, they suggested a DL network. To validate their model, they give the Area Under the Curve (AUC) metric on two large-scale deepfake detection datasets, the deepfake-TIMIT Dataset and DFDC. They evaluate their technique to multiple SOTA deepfake detection approaches, reporting per-video AUCs of 84.4% on the DFDC dataset and 96.6% on the DF-TIMIT dataset. Their method is the first to use both audio and video modalities and perceived emotions from two modalities to detect deepfakes.

Besides, R. Wang et al. (2020) introduced DeepSonar, a method that monitors the learned neuron behaviors from a voice synthesis system to detect AI-generated phony sounds. Their research offers a novel approach to detecting AI-assisted multimedia forgeries by observing neuron activities to develop a reliable and effective detector. Analyses of the three datasets show that it is effective and robust and that it has real-world promise in a loud background. They stated that while creating a detector, robustness should be prioritized because numerous manipulations of voices can be readily disguised as routine activities. In contrast, manipulation of images is restricted and detectable. They use the power of layer-wise neuron activation patterns to provide a clearer signal to classifiers than raw inputs in the hopes of capturing the tiny variations between real and DL-generated fake voices. Studies on three datasets encompassing both Chinese and English languages were done to confirm DeepSonar's high detection rates (98.1% average accuracy) and low false alarm rates (approximately 2% mistake rate) in detecting phony voices. Furthermore, extensive testing results demonstrate its resistance to manipulation attacks.

Finally, Wijethunga et al. (2020) proposed a method for training and validating models to get some of the greatest accuracy outcomes. Audio signal processing techniques, speaker diarization procedures, and synthetic speech detection processes are all discussed in their study. Although these signal denoising tasks were completed successfully, the approaches could be enhanced by using better filtering techniques and enhancing the datasets. The speech-denoising element helps clean and preprocesses audio using the Multilayer-Perceptron and CNN architectures, respectively, with 93% and 94% accuracy. Natural Language Processing (NLP) for text conversion with 93% accuracy and the RNN model for speaker tagging with 80% accuracy and 0.52 Diarization-Error Rate was used to accomplish speaker diarization. The final component uses a CNN architecture to discern actual and false sounds accurately. NLP for text conversion with 93% accuracy, RNN model for speaker tagging with 80% accuracy, and 0.52 Diarization-Error-Rate was used to accomplish speaker diarization. The final module takes a CNN architecture to discriminate between real and fraudulent audio with a 94% accuracy. With these findings, this study will significantly contribute to the field of speech analysis. Also, the speech-denoising component cleans and preprocesses audio using the Multilayer-Perceptron and CNN architectures, respectively, with 93% and 94% accuracy. Table 7 lists the deepfake sound applications that use DL techniques and their features.

5.4 | Fake hybrid multimedia detection

Deepfakes (algorithmically modified footage, photos, audio, and videos) could generate huge social upheaval when combined with the virality of social media. To verify the validity of digital content, anti-disinformation technologies such as deepfake detection algorithms or immutable metadata are needed. It is critical to have correctly developed tools to detect, analyze, and defeat deepfake digital content when interacting with multimedia deepfake technology. Multimedia material could take many forms, including films, artwork, photos, and sound recordings, to mention a few. Even if there is a trustworthy, secure, and trusted mechanism to trace the history of digital content, achieving this goal can be difficult. This part will look at how to use DL techniques to recognize hybrid multimedia content. So, Khalil et al. (2021) presented a deepfake detection approach that combines texture feature extraction with a High-Resolution Network (HRNet)-based strategy for training a capsule network to enhance classification accuracy. The core idea behind the HRNet would be to prevent the loss of any spatial information, this matched the capsule network's fundamental idea rather well—except that the original HRNet also had a pooling layer at the very end. The pooling layer from the HRNet is taken out, and the raw feature vector produced by the HRNet has been used to best suit the two ideas. When a capsule network (CapsNet) is combined with an improved routing approach (Xiao et al., 2020; Xiao et al., 2021) for classification, it offers greater feature abstraction and representation capabilities without requiring a lot of data or parameters. The use of You Only Look Once (YOLO-v3) as a face detector resulted in a significant decrease in false-positive predictions, eliminating the requirement for additional preprocessing steps. The model was enhanced and improved as a result of the merging of two distinct in-nature feature extraction techniques. The texture-based method employing Local Binary Patterns (LBPs) gave an artifacts-specific flavor to their model by counting on the alteration between the textures of the false face and the surrounding backdrop. CNN-based HRNet supplied informative representations to their capsule network while maintaining the capsule idea of not sacrificing any spatial information to get the most out of little details. The HRNet's findings were improved by performing an augmentation–normalization step on the entered pictures. With its superior feature abstraction and classification performance, CapsNet determined if the presented picture was real or false.

Also, Chintha et al. (2020) developed the XcepTemporal convolutional RNN system for deepfake detection. They utilized an XceptionNet CNN to identify prominent and efficient face features. This representation is fed through bidirectional recurrence layers, which help discover temporal discrepancies. They train their model with both standard cross-entropy and KL divergence loss functions. They offer a companion architecture that obtains audio feature representations by stacking several convolution modules on the audio side. This audio embedding is also routed through a bidirectional recurrent layer. They show the robustness of their techniques for visual deepfake identification to both compression and out-of-sample inference on the prominent FaceForensics++ and Celeb-DF datasets, as well as the newly released deepfake detection challenge. The ASVSpoof 2019 challenge dataset illustrates the resilience of their techniques for audio spoof detection against compression and out-of-sample inference. Their approaches achieve new benchmark standards for deepfake visual and spoof audio detection and show that they generalize well to unexpected attacks.

Plus, Kong et al. (2021) suggested a cross-modality approach that uses visual and auditory information to discover the hidden face behind the deepfake material. Every recovered audio segment is matched with all false faces retrieved from the related deepfake video, and each fake face is coupled with one matching ground truth face. The pipeline

TABLE 7 Sound deepfake techniques' methods, properties, and characteristics.

Authors	Main idea	Advantages	Research challenges	Security used?	Dataset	Simulation environment	Using TL?	Method	Usage?
Borrelli et al. (2021)	Providing a collection of characteristics built on the notion of describing the speech as an auto-regressive mechanism.	• High accuracy • Low complexity	• High response time • It is difficult to recognize some families of synthetic speech recordings in the open-set situation	No	The ASVSpoof 2019 logical access audio dataset	Scikit-learn—python library	No	LSTM	AI-synthesized fake voices
C.-Z. Yang et al. (2020)	Proposing a deep CNN-based visual speaker authentication method.	• High accuracy • Low energy consumption	• Not taking into account various scenarios • High delay	No	GRID dataset and MOBIO dataset	Python	No	Deep CNN	Speaker authentication
Mittal et al. (2020)	Using the similarities between audio-visual modalities and the similarity between affective cues.	• AUC of 84.4% on the DFDC and 96.6% on the DF-TIMIT datasets	• High complexity • High delay	No	TIMIT and DFDC dataset	PyAudioAnalysis	No	RNN + LSTM+	Audio-visual deepfake detection
R. Wang et al. (2020)	Developing a method based on observing neuron activities.	• Strong robustness • Obtained accuracy is 98.1	• High complexity • High energy consumption	No	FakeOrReal dataset, lab, MC-TTS, and Sprocket-VC	Python	No	CNN-DNN	AI-synthesized fake voices
Wijethunga et al. (2020)	Suggesting a system that can differentiate between actual and synthetic speech in a group conversation.	• NLP for text conversion with 93% accuracy • RNN model for speaker labeling with 80% accuracy • −0.52 for Diarization-Error-Rate	• High delay • High energy consumption • Low robustness	No	FakeOrReal dataset + the AMI Corpus dataset	TensorFlow	Yes	CNN + RNN	AI-synthesized fake voices

comprises two parts: (a) an invertible face autoencoder and (b) an audio-visual conditional face reconstruction. They train a generic face autoencoder to produce representations of ground truth faces with a specific resolution. The approach is the first of its kind in predicting the genuine face from all available data. The modality transfer is conducted at the feature level to accurately predict the face appearance, resulting in a promising performance in quantitative and qualitative assessments. It can have a significant influence on a wide range of applications in multimedia security and forensics.

Besides, Sun et al. (2020) presented a Landmark Breaker technique to prevent deepfake formation by breaking the step-facial landmark extraction requirement. Landmark Breaker is considered a new loss function that encourages the difference between anticipated and original heatmaps and optimizes it with the momentum iterative fast gradient sign approach. They primarily target CNN-based facial landmark extractors due to their superior performance. To accomplish it, they introduce adversarial perturbations into the facial landmark extraction process, preventing the input faces to the deepfake model from being adequately matched. Also, Landmark Breaker has been validated using the Celeb-DF dataset, demonstrating its efficiency in extracting unpleasant face landmarks. They further look into how Landmark Breaker performs with different parameter values. Unlike detecting approaches that only operate after a deepfake has been generated, Landmark Breaker goes one step further and prevents deepfake development.

In addition, Nasar et al. (2020) created a toolkit that uses cutting-edge technology to detect deep fakes in films, pictures, and audio. As a result, their scheme employs a combinational strategy that combines image processing with the CNN network to identify forgeries in deepfakes. Their system is divided into four components. The first is the data preparation module, which converts the input samples to picture samples. The second module is the data enhancement module, which removes the image's noise component. Following that is the CNN network, where the model is trained and tested, and finally, the detection module, which does real-time detection against various media files from multiple platforms. The model's accuracy was tested against different datasets, including the deepfake-TIMIT dataset, Face Forensic++, and others, and the accuracy was nearly 0.9 for all three model files.

Chugh et al. (2020) introduced a bimodal deepfake detecting model based on the Modality Dissonance Score (MDS), it measures the similarity between audio and visual streams for real and fake videos, allowing them to be distinguished. The MDS is represented using the contrastive loss estimated over segment-level audio-visual data and constraints authentic audio-visual streams to be closer than their fake counterparts. In addition, the unimodal streams are subjected to cross-entropy loss to ensure that they acquire discriminative features independently. Simulations indicate that (a) on the DFDC dataset, the MDS-based fake detection framework can attain state-of-the-art performance, and (b) on top of the contrastive loss, the unimodal cross-entropy losses provide additional benefits to improve fake detection performance. The suggested approach's explainability and interpretability are proven using audio-visual distance distributions for actual and fraudulent films, Grad-CAM results representing MDS network attention regions, and forgery localization results.

Finally, Chan et al. (2020) proposed a technique that develops on the experience acquired through the use of various LSTM-CNN to analyze visual and audio parts of digital media, as well as developing a descriptive caption to go along with the extracted feature information to form a representative hash that is distinctive to the content. If this material were not incorporated into a Hyperledger Fabric (HF) framework, deepfakes created from it would lack historical provenance. Permissioned blockchain, particularly HF combined with LSTMs for audio/video/descriptive captions, is a step in making deepfake media a viable tool. The original artist's verification of untampered data would be required for original content. The smart contract integrates several LSTM networks into a method that enables the tracing and tracking of digital content's historical origin. The outcome is a conceptual model for digital media Proof of Authenticity (PoA) employing a decentralized blockchain with several LSTMs as deep encoders to create unique discriminative characteristics, that are then compressed and hashed into a transaction. Their research implies that they trust the video at the point of reception. However, Table 8 lists the deepfake hybrid multimedia applications that use DL techniques and their features. Also, we will thoroughly examine and analyze the results with all of their features in the next part.

6 | RESULTS AND COMPARISONS

In the previous section, we looked at four different forms of deepfake detection methods that used different sorts of DL models. However, public opinion in media information has been eroded by deepfakes, as seeing them no longer equates to trust in them. They have the power to intensify political tensions, cause public indignation, inspire violence, or even start a war. They also can boost hate speech and disinformation. This is particularly important today because deepfake

TABLE 8 Hybrid multimedia deepfake systems' methodologies, traits, and features.

Authors	Main idea	Advantages	Research challenges	Security used?	Dataset	Simulation environment	Using TL?	Method	Usage?
Kohli and Gupta (2021)	Using LBP to highlight the differences between textures of the actual and forged parts of the picture.	• Improvement of 20.25% in the area under the curve • Very few false positives	• High complexity • High energy consumption	No	Deepfake detection challenge preview (DFDC-P) dataset	Not mentioned	No	LBP and CNN	Image and video detection
Chintha et al. (2020)	Presenting a convolutional bidirectional recurrent architecture.	• Strong robustness	• High complexity • High delay	No	FaceForensics++, Celeb-DF, and the ASVSpoof 2019 logical access audio datasets	Keras for python	Yes	CNN	Deepfake video and audio detection
Kong et al. (2021)	Proposing a method to use learning to produce features of the real face.	• High accuracy • Strong robustness	• High complexity • High delay	No	VoxCeleb2, CelebDF2, and authentic videos audios from Youtube	Python	Yes	Auto-encoders	Deepfake video and audio detection
Sun et al. (2020)	Proposing Landmark Breaker to block the development of deepfake by disrupting face landmark extraction.	• Strong robustness • High reliability	• High energy consumption	No	Celeb-DF dataset	PyTorch	No	CNN	Image and video detection
Nasar et al. (2020)	Combining image processing with the CNN network to provide a combinational method.	• Obtained accuracy as 0.9	• Low robustness • Low reliability	No	Deepfake-TIMIT dataset, face Forensic++	Keras	Yes	CNN	Images, videos, and audio
Chugh et al. (2020)	Using modality dissonance measured across tiny temporal segments.	• Increasing the AUC score by up to 7%	• High delay • Low reliability	No	DFDC and deepfake-TIMIT	Python	No	CNN	Videos, and audio
Chan et al. (2020)	Providing a decentralized blockchain and discriminative digital media.	• High security • Moderate accuracy	• High complexity	Yes	VidTIMIT database	Python	No	LSTM+ CNN+ blockchain	Videos, and audio

techniques are becoming more accessible, and social media platforms can swiftly propagate malicious information. Also, deepfakes do not always need to be broadcast to a large audience to have a negative impact. People who build deepfakes for malevolent purposes need to disseminate them to target audiences as part of a sabotaging campaign. They do not need to use social media to do it. For instance, security agencies could use this strategy to sway influential people's judgments, including politicians, posing national and international security risks (Wu et al., 2021; Z. Zhang, Luo, et al., 2020; Wei Zheng, Xun, Wu, et al., 2021). The scientific community has intended to build deepfake detection algorithms to address the worrying problem of deepfake, and various results have been published. So, this paper gives an overview of common methodologies as well as an assessment of cutting-edge techniques. A struggle is brewing between people who utilize powerful ML to build deepfakes and others who attempt to detect them. While the accuracy of deepfakes improves, so does the effectiveness of detection systems. The idea is that what DL has broken can also be mended by DL. Detection systems are now in their infancy, and various approaches have indeed been presented and tested, although on scattered datasets. Creating a constantly updated benchmark dataset of deepfakes to test the ongoing development of detection systems is one way to improve detection method effectiveness. This will make it easier to train detection algorithms, particularly those based on DL, involving a significant training set. Also, the detection mistake was seen in the results for all detection systems. As we spoke about in Section 5.4, the research revealed that a lip-syncing-based method could not identify face-swapping because GANs could indeed generate high-quality facial expressions that match audio speech, implying that only image-based algorithms could detect deepfake videos with high accuracy and that potential more modern technologies for face exchanging will be harder to identify. Several of the prevalent preventative measures indicated across articles included a dependable screening as well as filtering method between all systems, including such social networks, YouTube, and others, to automate the detection and remediation of fake news, improve legislator and legal requirements, and the use of watermarking tools into devices as then digital content creates immutable metadata that can be tracked. Although the general subgroup accuracy for detecting deepfake may vary, it is surprising to see that there is little to no notable difference between deepfake and the actual image or video.

During the last 2 years, deepfake films have gained widespread notice, representing a great danger to social security. Towards this purpose, the authors have conducted a large number of investigations and achieved great progress, as we discussed in Section 5.2. In the previous deepfake datasets, recent detection methods obtain about 100% detection accuracy. Nevertheless, with newly constructed datasets, the accuracy of current detection techniques is not perfect. Also, the average accuracy of detection methods offered in the latest methods competition was just 65%, demonstrating that existing detection approaches are still far from addressing the needs of practical settings. At the same time, existing research seeks to identify abstract information using a complicated network structure. Although better detection performance is achieved, the improvement in network complexity increases computation costs. The detection impact vs. network complexity is a tradeoff. A reasonable solution, we suggest, should provide improved precision detection while reducing network complexity. It is necessary to develop appropriate detection techniques to summarize past algorithms and explore new research approaches in such conditions. Our research study provides a detailed panoramic view of the domain, which includes up-to-date specifics: types of facial manipulations, public databases for research, facial manipulation techniques, and benchmarks for detecting each facial manipulation group, which would include crucial outcomes obtained by much more representative mobsters. In general, most contemporary face manipulations appear to be straightforward to identify in controlled circumstances, namely when false detectors are tested in the same conditions they were taught. It has been established in the majority of the criteria in this study, with very low mistake rates in manipulation identification. This situation, nevertheless, would not be very practical because fraudulent pictures and films are frequently shared on social media, with significant differences in compression level, scaling, noise, and other factors. Facial manipulation techniques are also getting better all the time. Those certain factors motivate more studies into the potential of fake detectors to generalize to new situations. Various works have looked at this topic in the past. A deeper analysis could follow in the footsteps of the most recent articles, as they do not necessitate the use of false movies for training, allowing for improved generalization of hidden threats.

A solution that aims to make accurate choices on the integrity of videos and photos must have numerous tools to address most of the relevant operating conditions. In truth, each approach only works when appropriate hypotheses are met, and when these are not met, the technique becomes entirely useless. In the case of splicing, a tool for copy-move detection, for example, and will be useless. However, combining various tools is necessary to broaden the range of detectable forgeries and increase the detection capability for each one. In essence, the traces to be identified are typically quite weak and could be rapidly camouflaged by deliberate attacks, such as those mentioned in the previous section, as well as by ordinary processing processes. As a result, combining numerous techniques designed to identify

comparable attacks using various attitudes is likely to increase performance, particularly robustness, in the face of both benign and malicious disturbances (D. Yang, Zhu, et al., 2022). On the other hand, maximizing the number of clues is a common investigative technique. A forensic tool's basic methodology is identifying appropriate low-level features from the original data, analyzing them to get a scalar score or a probability vector, and then evaluating the latter to reach the final judgment. As indicated in Section 5.4, a combination can occur at any three levels (feature, measurement, or abstract), each with its own advantages and disadvantages. When many tools are used, dealing at the feature level has considerable issues, as shown in various studies, especially when there are many features to deal with, and the complexity of establishing representative datasets for all probable important cases. Also, abstract-level integration could have already destroyed valuable information, diminishing the capacity to exploit cross-tool relationships. Studying at the level of measurement could be a good middle ground between these two extremes. While integration is unquestionably a valuable tool for improving performance, it is unclear how to merge disparate data bits effectively. Interestingly, the fraudulent user went one step further and replicated some content inside the poster, which was identified by a copy-move-based detector. Nevertheless, several alterations were required to construct a plausible forgery, and so multiple traces were left. It is not easy to combine these results.

Modern audio generation methods exhibit fingerprint artifacts and repeated inconsistencies across temporal and spectral domains. As we saw in previous sections, these artifacts could be well captured by frequency domain analysis over the spectrogram. Mel Frequency Cepstral Coefficients (MFCCs) were used in various speech processing techniques. It has been discovered that extracting features from an audio signal and feeding them into the base model produces significantly better results than directly feeding the raw audio signal into the model. The result of applying MFCC is a matrix containing feature vectors extracted from all of the frames. The rows in this output matrix represent the corresponding frame numbers, and the columns represent the corresponding feature vector coefficients. Finally, this output matrix is used in the classification process. A spectrogram is a visualization of a signal's frequency spectrum, where the frequency spectrum of a signal is the frequency range that the signal contains. According to research, humans do not perceive frequencies on a linear scale, so the Mel scale mimics how the human ear works. A mel-spectrogram also renders frequencies above a certain threshold logarithmically (the corner frequency). Mel-spectrograms are generated by applying a Fourier transform to a signal's frequency content and converting it to the mel-scale, whereas MFCCs are generated by a Discrete Cosine Transform (DCT) into a mel-frequency spectrogram. The linear audio spectrogram is best suited for applications in which all frequencies are equally important, whereas mel-spectrograms are better suited for applications in which human hearing perception must be modeled. Data from mel-spectrograms can also be used in audio classification applications.

6.1 | Analysis of results

However, Section 5 looked into four categories and 32 publications about applying DL approaches in deepfake detection applications. The most severe flaw in these methods was a lack of comparison across the offered approaches, which is predictable given the domain's novelty. As a result, identifying flaws in existing approaches might be difficult. Another point of contention has been the proposed methods' algorithm, which was not explicitly stated; the very least that could be expected of the proposed methods is a thorough explanation of the algorithms used, which, unfortunately, hardly any of the plans, algorithms, or pseudo-code did not provide. As a result, we provide a complete overview of the approaches covered in a variety of domains, including the simulation environment, applications, parameters targeted by the articles, protection mechanisms, TL methods, datasets, and so on. In addition, Python is the most popular programming language for this type of task in the case of a simulation, theoretical, or implementation setting about the proposed approaches, which is a very enticing component for researchers to employ in future work and has a wide variety of applications. As shown in Figure 9, 90% of strategies used Python for their simulations, with Keras being the most popular Python library with 24% usage. MATLAB is a rare environment, accounting for only 3% of all environments. With 15% usage, Pytorch is in the second position. TensorFlow is in third place with 9%, followed by Scikit-learn, Dlib, PyAuidoAnalys, and Caffe, each with 3%. Additionally, 30%, with the remainder of the works being done in Python, but without specifying which library was used. Also, two publications failed to mention their environments. Furthermore, as demonstrated in Figure 10. China (14 articles), the United States (7 articles), India (4 articles), South Korea (3 articles), and Italy (3 articles) have the most published articles in this area, according to a map of the author's nations based on their affiliation, with a focus on deepfake detection applications. So, the authors from these five countries have contributed the most to the project.

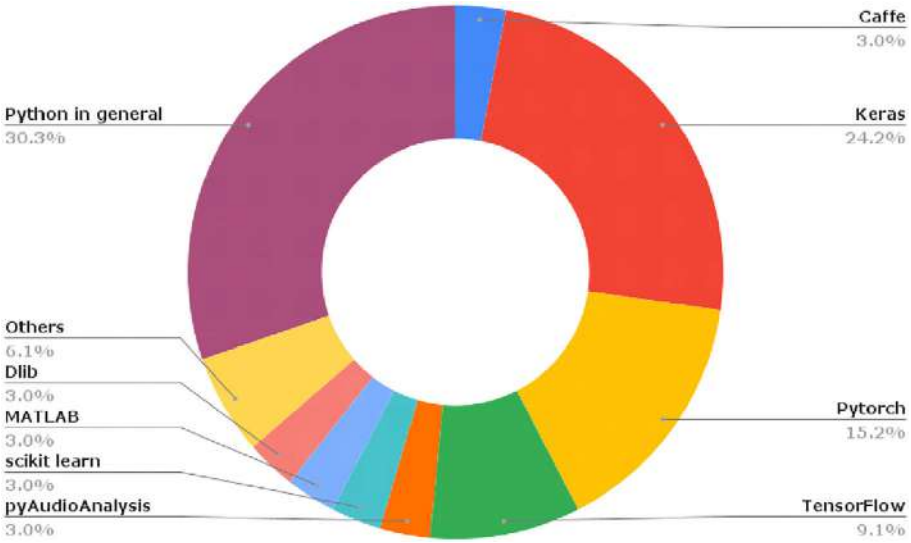

FIGURE 9 The distribution of simulation environments used in the DL-deepfake detection techniques.

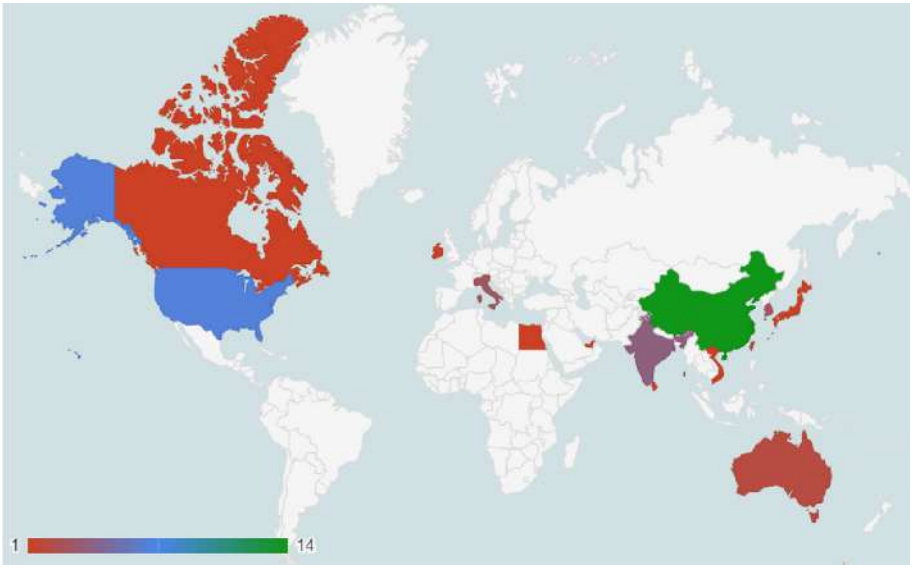

FIGURE 10 The most contributed countries by the studied publications are depicted in a geo-chart.

Another investigation of DL applications of deepfake detection in this domain found that 13 articles (40.62%) employed apps to detect fake videos. Given the global reach and the extent to which these videos could affect peace and security, one of the advantages of this strategy is that algorithms can find fake videos that humans cannot identify from authentic ones. The second place jointly and simultaneously goes to image and multimedia fake detection, with seven articles (21.87%). With five articles (15%), the fake audio attack becomes a big danger to the speaker verification system and is the third most intriguing issue. This is one of the most important uses of DLs in this sector, although it receives the least attention. The various uses of DL in the context of deepfake detection are depicted in Figure 11. Section 2 discusses TL in general. This section explains how TL could aid in mitigating errors in detecting fake content such as video, images, and sounds. As previously stated, developing enough datasets of fake images or any other deepfakes in a short period is challenging. As a result, utilizing the benefits of DL to deal with deepfakes is a bit difficult. So, TL could be useful in this situation. A DL model could be trained using a large-scale benchmark dataset, just like TL, and learned features can detect deepfakes. Only four studies in our evaluation seek to exploit this TL for diverse purposes in the realm of deepfake detection. The number of studies is small and the majority of the existing research and experiments have been utilized to detect fraudulent videos. Many of these activities could be made easier if DL interacts with IoT or

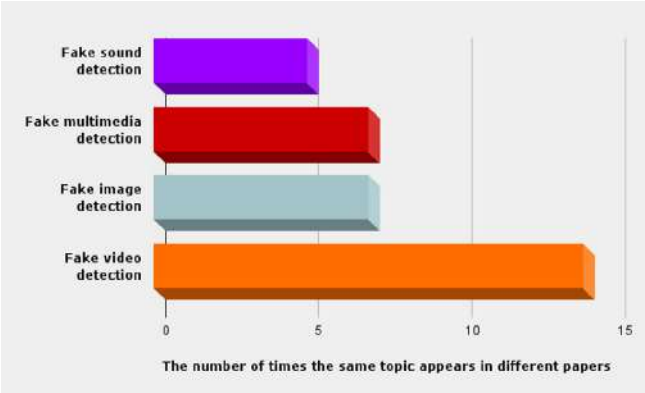

FIGURE 11 In the context of deepfake detection, DL application usages.

edge computing (B. Cao, Fan et al., 2021). In some situations, TL could be able to assist. A better system can be given when the most appropriate algorithm is used for edge computing.

Another significant feature is security, as the information about the person is kept confidential for ethical and legal reasons (Wei Zheng, Xun, Wu, et al., 2021). As a result, data security must be prioritized while establishing security structures for these processes. Only one of the procedures analyzed is regarded as a security mechanism for maintaining security in this area and the confidentiality of deepfake detection mechanisms, which is one of the study's most shocking conclusions. It is the principal disadvantage of the methods under consideration. Ensuring the confidentiality and security of users' information and processed data could be one of the essential areas for future research. As indicated in Table 9, the lowest metric evaluated in the articles is security; nevertheless, only one of the offered articles has considered this crucial element. The accuracy metric is the one that gets the most attention in articles, while the robustness metric comes in second. The complexity is ranked fourth, and the AUC is placed third. Furthermore, the problem with articles is that most have one target metric and ignore the others almost entirely. Since deepfake usage is one of the most dangerous human behaviors ever, with numerous individual and social repercussions, it is desired that additional research be conducted to address the phenomenon, its ramifications, and complexities, taking into account all potentially dangerous factors' features.

Furthermore, our findings showed that CNN is the most frequently employed method in articles (61%) and is used in almost every category, particularly in image and video deepfake detection. Furthermore, auto-encoders are the least used technique in these articles, with a 2.8% usage rate. In addition, Figure 12 depicts the frequency of DL approaches in deepfake detection applications. As a result, our final analysis focuses on the frequency of datasets utilized in the publications, which is an important aspect of our assessment which is depicted in Figure 13. Faceforensics is the most often used dataset (24.4%), followed by CelebA and TIMIT. HFM is ranked last with only one use. In this critical subject, we hope that our research will aid researchers in creating techniques to avoid and detect deepfake content.

6.2 | Criteria of DL/ML methods

The quality of functions is defined by mathematical metrics that show profitable feedback and analysis of a deepfake detection method's performance. We have to name a few critical parameters: accuracy, MCC, Confusion Matrix, recall, precision, and $F1$ score. As a result, as previously stated, accuracy is the most significant indicator for demonstrating the fraction of accurately recognized observed to satisfy the predicted observation demand. In the time of combining total values in a confusion matrix, the True Negative to True Positive rate is exploited. The total quantity of patterns successfully detected is demonstrated by n, and the entire number of patterns is given by t in this equation.

$$A = \frac{n}{t} \times 100 \tag{1}$$

The given number of exact predictions is indicated by P and the rate of True Positive forecasted compared with the total positively forecasted. Moreover, S_{TP} is the representation of the sum of total true positives, when A_{FP} is the representation of total false positives.

TABLE 9 Considered performance evaluation metrics in the examined articles.

Type	Authors	Method	Security	TL	Energy consumption	Response time	Accuracy	Delay	Complexity	AUC	Robustness
Fake image detection	Zhang, Zhao, and Li (2020)	CNN	•	•	✓	•	✓	•	•	✓	•
	Lee et al. (2021)	GAN	•	•	•	•	✓	•	•	✓	•
	Guo et al. (2021)	CNN	•	•	•	✓	✓	•	•	•	•
	Guarnera et al. (2020)	Expectation–Maximization+ CNN	•	•	•	•	✓	•	•	•	✓
	I.-J. Yu et al. (2020)	CNN	•	✓	•	•	✓	•	•	•	✓
	Yang et al. (2021)	CNN+ SLIC	•	•	•	•	✓	•	•	•	•
	Hsu et al. (2020)	Pairwise learning	•	•	•	•	✓	•	•	•	•
Fake video detection	(Güera & Delp, 2018)	CNN + RNN	•	•	•	•	✓	•	•	•	✓
	Nguyen et al. (2021)	CNN	•	•	•	•	✓	•	•	•	•
	Jung et al. (2020)	GAN	•	•	•	•	✓	•	•	•	•
	Karandikar et al. (2020)	CNN	•	✓	•	•	✓	•	•	•	•
	Z. Zhao, Zhou, et al. (2021)	CNN	•	•	•	•	✓	•	•	•	•
	Z. Xu et al. (2021)	CNN	•	•	•	•	✓	✓	•	•	•
	Kohli and Gupta (2021)	Frequency CNN	•	•	•	•	•	•	•	•	✓
	Chen and Tan (2021)	CNN + BP-DNN	•	✓	•	•	•	✓	•	•	•
	A. Yan, Yin-He, et al. (2021)	CNN	•	•	•	•	•	•	•	•	✓
	Caldelli et al. (2021)	CNN	•	•	•	•	•	•	•	•	✓
	L. Yan, Yin-He, et al. (2021)	CNN	•	•	•	•	•	•	✓	•	•
	Mitra et al. (2020)	CNN	•	✓	•	•	✓	•	•	•	•
	Suratkar et al. (2020)	CNN	•	✓	•	•	✓	•	•	•	✓
	Bonettini et al. (2021)	CNN	•	•	•	•	✓	•	•	•	•
	Cozzolino et al. (2019)	CNN	•	•	•	•	✓	•	•	•	•
Fake sound detection	Borrelli et al. (2021)	LSTM	•	•	•	•	✓	•	✓	•	•
	Yang et al. (2020)	Deep CNN	•	•	✓	•	✓	•	•	•	•
	Mittal et al. (2020)	RNN + LSTM	•	•	•	•	•	•	•	✓	•
	R. Wang et al. (2020)	CNN-DNN	•	•	•	•	✓	•	•	•	✓
	Wijethunga et al. (2020)	CNN + RNN	•	✓	•	•	✓	•	•	•	•

TABLE 9 (Continued)

Type	Authors	Method	Security	TL	Energy consumption	Response time	Accuracy	Delay	Complexity	AUC	Robustness
Fake multimedia	Khalil et al. (2021)	LBP and CNN	•	•	•	•	•	•	•	✓	•
	Chintha et al. (2020)	CNN	•	✓	•	•	•	•	•	•	✓
	Kong et al. (2021)	Auto-encoders	•	✓	•	•	✓	•	•	•	✓
	Sun et al. (2020)	CNN	•	•	•	•	•	•	•	•	✓
	Nasar et al. (2020)	CNN	•	✓	•	•	✓	•	•	•	•
	Chugh et al. (2020)	CNN	•	•	•	•	•	•	•	✓	•
	Chan et al. (2020)	LSTM+CNN+ Blockchain	✓	•	•	•	✓	•	•	•	•

Note: The bullets indicate that metrics are not being examined.

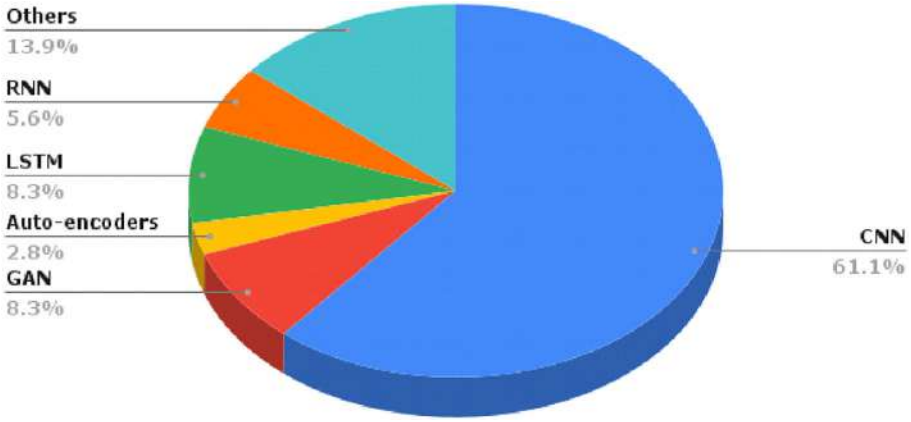

FIGURE 12 The utilization of DL approaches and their frequency in selected articles in deepfake detection applications.

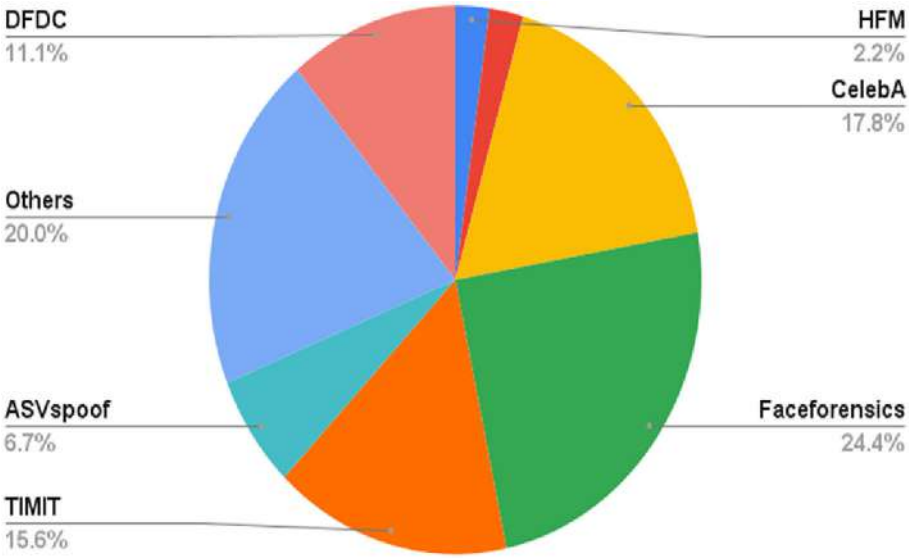

FIGURE 13 The frequency of datasets used in the articles.

$$P = \frac{S_{\mathrm{TP}}}{S_{\mathrm{TP}} + A_{\mathrm{FP}}} \times 100 \tag{2}$$

A criterion of the amount of real positive observations is demonstrated by $\mathrm{Re}_{\mathrm{Call}}$ named recall which can precisely forecast. Besides, A_{FN} specifies total false negatives in Equation (3).

$$\mathrm{Re}_{\mathrm{Call}} = \frac{S_{\mathrm{TP}}}{S_{\mathrm{TP}} + A_{\mathrm{FN}}} \times 100 \tag{3}$$

In addition, the $F1$ score is a total functionality criterion determination of Recall and Precision and representation of Harmonic achieved by Precision and Recall.

$$F1_{\mathrm{score}} = \frac{2 \times \mathrm{Re}_{\mathrm{Call}} \times P}{\mathrm{Re}_{\mathrm{Call}} + P} \times 100 \tag{4}$$

Additionally, functionality matrix measurement which weighs forecasted and real observations is a confusion matrix that utilizes True Negatives, True positive, False Positives labels, and False Negatives. All true predictions are

the total number of Positives and Negatives, so all wrong predictions are the aggregated False Negatives and False Positives.

$$\begin{vmatrix} S_{\mathrm{TP}} & A_{\mathrm{FP}} \\ A_{\mathrm{FN}} & \mathrm{TN} \end{vmatrix} \tag{5}$$

Further, a class that is both true and positive is predicted by True Positives. As well, True Negative are wrong predictions but are Negative to a clasNegatives are Negative predictions but are incorrect. Also, an individual value functionality demonstrator which encapsulates the total Confusion Matrix is the MCC. It presents a more instructive and correct result than the F1 score and accuracy in evaluating assortment challenges. If the prediction findings are advantageous in four Confusion Matrix zones, provide a high score.

7 | OPEN ISSUES

In the preceding section, we comprehensively examined the mechanisms under consideration. So, considering the context of modern research on deepfake detection, we have compiled a list of upcoming projects requiring considerable study addresses.

7.1 | Identity switching

Even though various methods have already been reported in the literature, determining which is the best is undoubtedly difficult. This is the result of a variety of circumstances. Initially, general methods are trained for a certain database and compression level, yielding generally successful performance. Nonetheless, they all demonstrate low generalization to unobserved circumstances. Furthermore, various measures (e.g., accuracy, AUC, etc.) and empirical validation do not allow for fair evaluations between studies. All of these factors should be taken into account further to develop in the area. In addition, we want to emphasize the identification conclusions reached in the most recent deepfake datasets of the second generation, including such deepfake detection challenges and Celeb-DF. Because when fake detectors already obtain AUC outcomes near to 100% in datasets of the first generation, such as in the UADFV dataset and FaceForensics++, they all struggle from such a significant performance degradation on the newer releases, with AUC outcomes underneath 60% in most case scenarios for the Celeb-DF dataset.

7.2 | Synthetic speech detection

Several circumstances necessitate more research into synthetic speech detection. Because of the wide range of synthetic speech generation technologies, it is still difficult to recognize some families of synthetic voice tracks in an open-set situation. Furthermore, the majority of the studies only addressed the logical access synthetic speech recognition problem, but other articles examined a clean recording of each speech. Future research should investigate what occurs if voice recordings are damaged by noise, coding, or transmission problems. This situation is especially relevant when authors consider synthetic voice recordings posted on social media sites or utilized live during phone calls.

7.3 | Interpretability

The interpretability of NN-based methods has become an issue. The NN, being a black box model, cannot offer human-understandable reasons for its result (Fan et al., 2022; L. Zhou et al., 2023). Nevertheless, the detection method must be interpretable in real-world forensic settings, or convincing findings will be achieved. Several areas have done significantly related studies on interpretability, but research on interpretability in deepfake detecting fields has not developed. The interpretability of deepfake detection techniques remains a significant issue that must be addressed in the upcoming.

7.4 | Generalization

Generalization is an essential indication for measuring algorithm performance and is frequently used to assess the system's performance on unknown datasets. The suggested detection methods primarily focus on supervised learning, which would be prone to overfitting its datasets. Relevant studies conducted in a few articles have demonstrated that the generalization performance of the existing detection techniques is still inadequate for cross dataset detection tasks. In practice, there seem to be significant variations in the choosing of source videos and the post-processing of produced films, leading to various data sets implying unique distributions. To the authors' knowledge, considerable effort has been made to improve method generalization. Nevertheless, due to their unique design, these algorithms have intrinsic faults. Some rely significantly on blending stages, making it impossible to identify artifacts in fully synthetic pictures. As a result, generality remains a pressing issue that must be addressed.

7.5 | Face synthesis

The majority of contemporary modifications focus on GAN architectures, including Style GAN, which produces extremely realistic pictures. Nonetheless, most detectors can differentiate between actual and fake pictures with near-perfect accuracy. It is caused by the fact that fake photos are distinguished using unique GAN fingerprints. But even if authors could erase the GAN fingerprints or bring multiple noise patterns, still maintain highly realistic synthetic images? Modern methods have concentrated on this study path, which is a problem even for the finest modification detection methods.

7.6 | Manipulation of attributes

Because most manipulations are based on GAN structures, the same issue mentioned for face synthesis (GAN fingerprint removal) generally applies here. Furthermore, the paucity of database searches for study and the lack of standard experimental methods to make meaningful comparisons among experiments is worth noting. A further intriguing future study subject is the creation of various counter-forensic procedures that we feel have a fundamental right to exist. Furthermore, technologies to fool forensic detectors introduce a new and fascinating participant into the game, challenging the detectors of false multimedia material. It is indeed worth noting that specific initializations on the first layer of a network design have already been utilized in the anti-forensics domain. Consideration of anti-forensic technique assaults might increase the robustness of forensic detectors. As we observed in previous parts, nearly all existing DL-based anti-forensic algorithms employ a GAN model; it has been shown to provide good results. Nonetheless, several techniques for eliminating forensic cues might be investigated, ranging from the creation of appropriate network topologies to the detailed analysis and elimination of forensic traces using tailored layers and loss functions. Therefore, it is worth noting that specific initializations on the first layer of a network design have been utilized in the anti-forensics sector. Consideration of anti-forensic technique assaults might increase the robustness of forensic detectors. We consider that rivalry between forensics and anti-forensics would indeed be good for the progress of both fields and is an intriguing issue to watch. Also, some methods, such as the fast optimization method (Xi et al., 2021), recursive neural net (J. Li et al., 2017), and Region-Based Convolutional Neural Network (R-CNN) (Wenfeng Zheng, Yin, Chen, et al., 2021), can increase the efficiency of the related mechanisms.

7.7 | Changing expressions

To the best of our knowledge, the only public database in expression swap is FaceForensics++, as opposed to identity swap, which has rapidly expanded with the introduction of better deepfake databases. This dataset is distinguished by easily detectable visual artifacts, yielding close to 100% AUC resulting in many false detection techniques. We motivate investigators to create and make public more comprehensive datasets depending on cutting-edge methodologies. Also, authors can intend to enhance the amount and quality of the HFM dataset in the future by incorporating different face situations to decrease bias against a particular group and by creating various handcrafted fake facial pictures using other software programs, including Gimp, Pixlr, and Photoshop. Additional direction for research consideration is by

using GANs to supplement manufactured false facial pictures for training or to use TL, a domain adaptation approach, to increase detection accuracy with a short dataset. Authors could adapt this information to identifying handmade facial manipulations by saving the knowledge obtained when solving a comparable facial fake detection issue by recognizing GAN-produced fake faces, deepfakes, and FaceSwap to multiple models. Future research can also provide more effective 3D CNN structures to better fuse the spatial properties, and a lighter CT module will be developed to extract features with fewer parameters.

7.8 | Triplet training

Due to the heterogeneous distribution of datasets, the most difficult challenge for deepfake detection tasks is that generalization performance is insufficient to satisfy the demands of actual scenarios. It is challenging for detection algorithms to understand the inherent difference between real and fraudulent movies in such conditions. To overcome this issue, a triplet training approach might be used. In the feature space, triplet training reduces the distance between samples of the same classification and maximizes the distance between other categories' samples. The triplet training approach, in particular, guarantees that the distance between samples belonging to distinct groups is greater than the distance between samples belonging to the same classification. As a result, the optimization objective of triplet training would be to leverage the inherent difference between actual and false videos, therefore aiding in later classification tasks. In the domain of face liveness detection, triplet training is being used for domain adaptation tasks, proving the ability of the triplet training method to discover intrinsic differences between genuine and false films, even though the datasets have different characteristics.

7.9 | Anti-forensics

Anti-forensic technology has been introduced in response to flaws in present forensic techniques. NNs are frequently employed in the field of deepfake detection to differentiate counterfeit videos. However, because of intrinsic flaws, NNs cannot withstand adversarial sample assaults. To that aim, investigators must develop more robust methods to survive hypothetical laboratory attacks to avoid such attacks in complex real-world settings. The advancement of anti-forensics technology allows for the prediction of potential assaults and the discovery of flaws in current algorithms, allowing for the improvement of existing methodologies.

7.10 | Multitask learning

Compared to single-task learning, multitask learning involves doing multiple tasks simultaneously and has been shown to enhance prediction performance. It has been discovered that doing both forgery location and deepfake recognition simultaneously time improves accuracy in deepfake identification activities. Multitask learning allows designers to execute two tasks simultaneously, accounting for losses produced by both tasks and enhancing the model's performance. Moreover, we demonstrate that forging location is critical in the deepfake detection job. As a result, multitask learning has a significant potential for further improving deepfake detection.

7.11 | Robustness and resilience

Robustness is a term that is frequently used to assess the effectiveness of detection algorithms when they are subjected to various degradations. Compressed videos are harder to identify when compared to original videos because they disregard a lot of visual information to achieve a greater compression rate. When confronted with low-quality movies, detection techniques frequently show a reduction in performance when compared to high-quality videos. Video files may be subjected to procedures including picture reshaping, rotation, and compression. In those kinds of cases, robustness would become a crucial feature to consider when building detection systems. Adding a noise layer to the detection network to account for different data degradation conditions is an excellent method to increase resilience. Increasing the resilience of present detection systems will be critical in the future.

7.12 | Response time

Processing time has become a critical factor when used in a real-world scenario. Deepfake detection techniques will also be tightly integrated on streaming media platforms shortly to mitigate the harmful impact of deepfake videos on social security. Nevertheless, due to their high time consumption, existing detection algorithms are far from being widely implemented in actual settings. Given that the movies in practice are far longer than 300 frames, that time consumption falls short of satisfying the demands of large video detection. In the existing studies on deepfake detection, detection accuracy is considered the only benchmark, with just a few studies focusing on deepfake detection time consumption. Additional emphasis should be paid in the future to researching how to build an efficient and high-accuracy detecting technique.

7.13 | Human performance

Even though the dangers of internet deepfake movies are well-known, there seems to be currently no systematic or quantitative research on the perceptual and psychological components that contribute to their deception. Intriguing questions remain unanswered; for instance, it is feasible that deepfake films have an uncanny valley; what is the sole noticeable difference between high-quality deepfake movies and actual videos to human eyes; and what types/aspects of deepfake videos are considerably more successful in deceiving consumers. It will take proper cooperation between digital media forensics and perceptual and social psychology academics to answer these problems. Such investigations are undeniably helpful for both research into detection systems and a greater understanding of the societal impact that deepfake videos might have.

7.14 | Measures to protect

Nevertheless, considering the speed and accessibility of online media, even with the most advanced forensic tools would primarily be postmortem in nature, relevant just after DL-generated phony face photographs or videos emerge. In addition to forensic tools, investigators can develop preventative measures to protect persons from being victims of this kind of attack. The addition of specially designed patterns known as adversarial perturbations, which are undetectable to human eyes but can result in detection failures, is one such strategy that academics have lately examined. The reasoning behind this is as follows: high-quality DL face synthesis models require a huge number of training face photos, often in the 100, if not millions, gathered via automatic face identification techniques, referred to as face sets. Adversarial perturbations "pollute" a face set, resulting in a small number of actual faces and a large number of non-faces that are of little or no use as training material for DL face synthesis models. The suggested adversarial perturbation generation approach can either be implemented as a feature of photo/video sharing platforms before a user uploads their own photos or movies or as a separate tool that the user can use to edit photos and videos before they are shared online.

7.15 | Blockchain

Identity sovereignty is possible with a permissioned blockchain like HF, which means the original artist retains complete control over their work. To understand the complexities inherent in this proposed system, further in-depth discussion, and implementation outcomes are required. If detection techniques reduce inaccuracy, the system may also attach detection models as a preprocessing step prior to HF. As DL technology advances, such as the development/improvement in the standards of DL model creation and compression for archival purposes, we hope to implement more of the interaction of cutting-edge content-unique hashing methods, integrity methods, security measures, and widely used blockchains in an effort to ensure that.

7.16 | Deepfakes in other forms

Though face-swapping has been the most well-known deepfake video technique, it is far from the most effective. Face-swapping deepfake videos, for example, have significant limits when it comes to impersonating someone. According to

psychological studies, human face recognition depends heavily on facial shape and hairstyle information. As a result, the person whose face is to be replaced (the target) must have a comparable facial shape and haircut to the person whose face is swapped to generate a realistic imitating impression. Furthermore, because the synthetic faces must be spliced into the original video frame, discrepancies between the generated region and the rest of the original structure could be significant and challenging to hide. Within those ways, the other two types of deepfake videos, notably head puppetry and lip-syncing, seem to be more efficient and must therefore be the subject of future deepfake detection research. During recent decades, techniques for researching complete face synthesis or reenactment have advanced rapidly. Even though there have not been as numerous convenient and free public software tools for creating such kinds of deepfake videos as there are for face-swapping videos, the growing complexity of the generation algorithms could soon change that. Since this generated region differs from that of face-swapping deepfake films, identification systems that focus on face-swapping artifacts are unlikely to be useful for such videos. As a result, researchers should work on developing detecting systems for these types of deepfake movies. Performance analysis, experiments with big datasets, scalability, consistency, and reliability (K. Cao, Wang, et al., 2021) are all planned for the future.

7.17 | GAN-based fake datasets

The GAN approach is the most extensively utilized for image classification and detection. It has recently gained popularity in the medical and healthcare sectors (Lv et al., 2022), and it is one of the most enticing methodologies for investigators. Furthermore, GANs are an intriguing ML approach. Because GANs are generative models, they generate new data instances comparable to the training data. GANs, for example, may produce images that resemble photos of real faces, even though the faces do not belong to any actual person. The lack of large datasets and databases of high-quality pictures for training is one of the most significant obstacles preventing using GAN approaches to generate artificially produced datasets. According to the paper, GAN techniques can create fake datasets without enormous picture datasets.

8 | CONCLUSION AND LIMITATION

Deepfakes have eroded people's faith in digital content when seeing them no longer equates to believing in them. This becomes especially important today, even as capabilities for making deepfakes are becoming more accessible, and online platforms can quickly distribute fake content. People's beliefs and truths can be jeopardized in the absence of deepfake detection methods. In this regard, such methods can help prevent the spread of fake multimedia content worldwide while also making it easier for the media to detect them. This work offered a systematic review of the DL mechanisms for deepfake detection. Before presenting the goal of this research, we addressed the advantages and disadvantages of some systematic and peer-reviewed studies about DL-deepfake detection algorithms. Also, the advantages and disadvantages of each mechanism were explored in four categories based on their applicability. The detecting tools and platforms for DL-deepfake were also examined. According to articles based on qualitative characteristics, most publications are assessed based on accuracy, AUC, latency, robustness, and complexity. Meanwhile, some functions, such as security and delay, go unused. Besides, various programming language libraries are utilized to analyze and implement the discussed methods, with Keras accounting for 24% of the effort. According to the report, recognizing video deepfakes is classified by 40% of apps. In the fight against deepfakes, we also hope this research will serve as a valuable reference for future investigation on DL and deepfake applications. In terms of the prospective findings, applying DL to process deepfake detection takes a lot of time and effort and tight collaboration between government, business, and academia. On the other hand, DL has been acclaimed as a wonderful methodology for developing intelligent solutions to these challenges. The findings of this work could assist in developing deepfake detection-based DL algorithms in real-world scenarios.

Also, we have encountered several obstacles, including the inaccessibility of non-English articles, which has prevented us from participating in several development programs. Another limitation of our research is that several articles we looked at had severe flaws in clear descriptions of the algorithms they utilized. Because this is a brand-new issue with little maturity in articles and methodologies, constraints such as not comparing the suggested strategy to other existing methods made it difficult to judge the success of the approaches. Also excluded from the review were studies that did not directly address deepfake detection. Besides, other articles that were nearly identical to deepfake

detection were also removed from consideration. Another issue we ran into was the inaccessibility of numerous articles published by specific publications.

AUTHOR CONTRIBUTIONS

Arash Heidari: Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); project administration (equal); resources (equal); supervision (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Nima Jafari Navimipour:** Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); project administration (equal); resources (equal); supervision (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Hasan Dag:** Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); project administration (equal); resources (equal); supervision (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Mehmet Unal:** Investigation (equal); methodology (equal); resources (equal); supervision (equal); validation (equal); writing – review and editing (equal).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The article contains all of the data.

ORCID

Arash Heidari https://orcid.org/0000-0003-4279-8551

RELATED WIREs ARTICLES

Deepfake attribution: On the source identification of artificially generated images

REFERENCES

Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). Mesonet: A compact facial video forgery detection network. *Paper presented at the 2018 IEEE International Workshop on Information Forensics and Security (WIFS)*.

Ahmed, I., Ahmad, M., Rodrigues, J. J., & Jeon, G. (2021). Edge computing-based person detection system for top view surveillance: Using CenterNet with transfer learning. *Applied Soft Computing*, *107*, 107489.

Ahmed, S. R. A., & Sonuç, E. (2021). Deepfake detection using rationale-augmented convolutional neural network. *Applied Nanoscience*, *13*, 1485–1493.

Akhtar, Z., Mouree, M. R., & Dasgupta, D. (2020). Utility of deep learning features for facial attributes manipulation detection. *Paper presented at the 2020 IEEE International Conference on Humanized Computing and Communication with Artificial Intelligence (HCCAI)*.

Albahar, M., & Almalki, J. (2019). Deepfakes: Threats and countermeasures systematic review. *Journal of Theoretical and Applied Information Technology*, *97*(22), 3242–3250.

Aversano, L., Bernardi, M. L., Cimitile, M., & Pecori, R. (2021). A systematic review on deep learning approaches for IoT security. *Computer Science Review*, *40*, 100389.

Balaji, T., Annavarapu, C. S. R., & Bablani, A. (2021). Machine learning algorithms for social media analysis: A survey. *Computer Science Review*, *40*, 100395.

Baygin, M., Yaman, O., Baygin, N., & Karakose, M. (2022). A blockchain-based approach to smart cargo transportation using UHF RFID. *Expert Systems with Applications*, *188*, 116030.

Bekci, B., Akhtar, Z., & Ekenel, H. K. (2020). Cross-dataset face manipulation detection. *Paper presented at the 2020 28th Signal Processing and Communications Applications Conference (SIU)*.

Biswas, A., Bhattacharya, D., & Kakelli, A. K. (2021). DeepFake detection using 3D-Xception net with discrete Fourier transformation. *Journal of Information Systems and Telecommunication*, *3*(35), 161–168.

Bonettini, N., Cannas, E. D., Mandelli, S., Bondi, L., Bestagini, P., & Tubaro, S. (2021). Video face manipulation detection through ensemble of CNNS. *Paper presented at the 2020 25th International Conference on Pattern Recognition (ICPR)*.

Borrelli, C., Bestagini, P., Antonacci, F., Sarti, A., & Tubaro, S. (2021). Synthetic speech detection through short-term and long-term prediction traces. *EURASIP Journal on Information Security*, *2021*(1), 1–14.

Burroughs, S. J., Gokaraju, B., Roy, K., & Khoa, L. (2020). DeepFakes detection in videos using feature engineering techniques in deep learning convolution neural network frameworks. *Paper presented at the 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*.

Caldelli, R., Galteri, L., Amerini, I., & Del Bimbo, A. (2021). Optical flow based CNN for detection of unlearnt deepfake manipulations. *Pattern Recognition Letters*, *146*, 31–37.

Cao, B., Fan, S., Zhao, J., Tian, S., Zheng, Z., Yan, Y., & Yang, P. (2021). Large-scale many-objective deployment optimization of edge servers. *IEEE Transactions on Intelligent Transportation Systems*, *22*(6), 3841–3849.

Cao, K., Wang, B., Ding, H., Lv, L., Tian, J., Hu, H., & Gong, F. (2021). Achieving reliable and secure communications in wireless-powered NOMA systems. *IEEE Transactions on Vehicular Technology*, *70*(2), 1978–1983.

Castillo Camacho, I., & Wang, K. (2021). A comprehensive review of deep-learning-based methods for image forensics. *Journal of Imaging*, *7*(4), 69.

Chan, C. C. K., Kumar, V., Delaney, S., & Gochoo, M. (2020). Combating deepfakes: Multi-LSTM and blockchain as proof of authenticity for digital media. *Paper presented at the 2020 IEEE/ITU International Conference on Artificial Intelligence for Good (AI4G)*.

Chen, B., & Tan, S. (2021). FeatureTransfer: Unsupervised domain adaptation for cross-domain deepfake detection. *Security and Communication Networks*, *2021*, 1–8.

Chi, H., Maduakor, U., Alo, R., & Williams, E. (2020). Integrating deepfake detection into cybersecurity curriculum. *Paper presented at the Proceedings of the Future Technologies Conference*.

Chintha, A., Thai, B., Sohrawardi, S. J., Bhatt, K., Hickerson, A., Wright, M., & Ptucha, R. (2020). Recurrent convolutional structures for audio spoof and video deepfake detection. *IEEE Journal of Selected Topics in Signal Processing*, *14*(5), 1024–1037.

Chugh, K., Gupta, P., Dhall, A., & Subramanian, R. (2020). Not made for each other-audio-visual dissonance-based deepfake detection and localization. *Paper presented at the Proceedings of the 28th ACM International Conference on Multimedia*.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2015). Gated feedback recurrent neural networks. *Paper presented at the International conference on machine learning*.

Cozzolino, D., Poggi, G., & Verdoliva, L. (2019). Extracting camera-based fingerprints for video forensics. *Paper Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Deshmukh, A., & Wankhade, S. B. (2021). Deepfake detection approaches using deep learning: A systematic review. In V. E. Balas, V. B. Semwal, A. Khandare, & Patil, M. (Eds.), *Intelligent Computing and Networking*. Lecture Notes in Networks and Systems, Vol 146. Springer. https://doi.org/10.1007/978-981-15-7421-4_27

Dixit, P., & Silakari, S. (2021). Deep learning algorithms for cybersecurity applications: A technological and status review. *Computer Science Review*, *39*, 100317.

Doewes, R. I., Gharibian, G., Zadeh, F. A., Zaman, B. A., Vahdat, S., & Akhavan-Sigari, R. (2023). An updated systematic review on the effects of aerobic exercise on human blood lipid profile. *Current Problems in Cardiology*, *48*(5), 101108. https://doi.org/10.1016/j.cpcardiol.2022.101108

Du, M., Pentyala, S., Li, Y., & Hu, X. (2020). Towards generalizable deepfake detection with locality-aware autoencoder. *Paper presented at the Proceedings of the 29th ACM International Conference on Information & Knowledge Management*.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., & Eirug, A. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994.

Esmailiyan, M., Amerizadeh, A., Vahdat, S., Ghodsi, M., Doewes, R. I., & Sundram, Y. (2021). Effect of different types of aerobic exercise on individuals with and without hypertension: An updated systematic review. *Current Problems in Cardiology*, 101034. https://doi.org/10.1016/j.cpcardiol.2021.101034

Fan, Q., Zhang, Z., & Huang, X. (2022). Parameter conjugate gradient with secant equation based Elman neural network and its convergence analysis. *Advanced Theory and Simulations*, *5*(9), 2200047.

Feng, Y., Zhang, B., Liu, Y., Niu, Z., Dai, B., Fan, Y., & Chen, X. (2021). A 200–225-GHz manifold-coupled multiplexer utilizing metal waveguides. *IEEE Transactions on Microwave Theory and Techniques*, *69*(12), 5327–5333.

Fernandes, S. L., & Jha, S. K. (2020). Adversarial attack on deepfake detection using RL based texture patches. *Paper presented at the European Conference on Computer Vision*.

Fink, O., Wang, Q., Svensen, M., Dersin, P., Lee, W.-J., & Ducoffe, M. (2020). Potential, challenges and future directions for deep learning in prognostics and health management applications. *Engineering Applications of Artificial Intelligence*, *92*, 103678.

Garg, A., & Mago, V. (2021). Role of machine learning in medical research: A survey. *Computer Science Review*, *40*, 100370.

Gosse, C., & Burkell, J. (2020). Politics and porn: How news media characterizes problems presented by deepfakes. *Critical Studies in Media Communication*, *37*(5), 497–511.

Guarnera, L., Giudice, O., & Battiato, S. (2020). Fighting deepfake by exposing the convolutional traces on images. *IEEE Access*, *8*, 165085–165098.

Güera, D., & Delp, E. J. (2018). Deepfake video detection using recurrent neural networks. *Paper presented at the 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*.

Guo, Z., Yang, G., Chen, J., & Sun, X. (2021). Fake face detection via adaptive manipulation traces extraction network. *Computer Vision and Image Understanding*, *204*, 103170.

Habeeba, M. S., Lijiya, A., & Chacko, A. M. (2021). Detection of Deepfakes using visual artifacts and neural network classifier. In *Innovations in electrical and electronic engineering* (pp. 411–422). Springer.

Heidari, A., Jafari Navimipour, N., Unal, M., & Toumaj, S. (2022). Machine learning applications for COVID-19 outbreak management. *Neural Computing and Applications*, *34*, 15313–15348. https://doi.org/10.1007/s00521-022-07424-w

Heidari, A., Navimipour, N. J., & Unal, M. (2022). Applications of ML/DL in the management of smart cities and societies based on new trends in information technologies: A systematic literature review. *Sustainable Cities and Society*, 104089. https://doi.org/10.1016/j.scs.2022.104089

Heidari, A., Navimipour, N. J., Unal, M., & Toumaj, S. (2022). The COVID-19 epidemic analysis and diagnosis using deep learning: A systematic literature review and future directions. *Computers in Biology and Medicine*, *141*, 105141.

Heidari, A., Toumaj, S., Navimipour, N. J., & Unal, M. (2022). A privacy-aware method for COVID-19 detection in chest CT images using lightweight deep conventional neural network and blockchain. *Computers in Biology and Medicine*, *145*, 105461.

Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv Preprint arXiv:1503.02531*, *2*(7).

Hong, T., Guo, S., Jiang, W., & Gong, S. (2021). Highly selective frequency selective surface with ultrawideband rejection. *IEEE Transactions on Antennas and Propagation*, *70*(5), 3459–3468.

Hongmeng, Z., Zhiqiang, Z., Lei, S., Xiuqing, M., & Yuehan, W. (2020). A detection method for DeepFake hard compressed videos based on super-resolution reconstruction using CNN. *Paper presented at the Proceedings of the 2020 4th High Performance Computing and Cluster Technologies Conference & 2020 3rd International Conference on Big Data and Artificial Intelligence*.

Hossain, M. S., Cucchiara, R., Muhammad, G., Tobón, D. P., & Saddik, A. E. (2022). Special section on AI-empowered multimedia data analytics for smart healthcare. *ACM Transactions on Multimedia Computing, Communications and Applications*, *18*, 1–2.

Hsu, C.-C., Zhuang, Y.-X., & Lee, C.-Y. (2020). Deep fake image detection based on pairwise learning. *Applied Sciences*, *10*(1), 370.

Hung, J. C., & Chang, J.-W. (2021). Multi-level transfer learning for improving the performance of deep neural networks: Theory and practice from the tasks of facial emotion recognition and named entity recognition. *Applied Soft Computing*, *109*, 107491.

Iqbal, T., & Qureshi, S. (2022). The survey: Text generation models in deep learning. *Journal of King Saud University-Computer and Information Sciences*, *34*(6), 2515–2528.

Jafar, M. T., Ababneh, M., Al-Zoube, M., & Elhassan, A. (2020). Forensics and analysis of deepfake videos. *Paper presented at the 2020 11th International Conference on Information and Communication Systems (ICICS)*.

Jeyaraj, A., & Dwivedi, Y. K. (2020). Meta-analysis in information systems research: Review and recommendations. *International Journal of Information Management*, *55*, 102226.

Jia, T., Cai, C., Li, X., Luo, X., Zhang, Y., & Yu, X. (2022). Dynamical community detection and spatiotemporal analysis in multilayer spatial interaction networks using trajectory data. *International Journal of Geographical Information Science*, *36*, 1–22.

Jiang, J., Li, B., Wei, B., Li, G., Liu, C., Huang, W., & Yu, M. (2021). FakeFilter: A cross-distribution Deepfake detection system with domain adaptation. *Journal of Computer Security*, *29*(4), 1–19.

Jung, T., Kim, S., & Kim, K. (2020). DeepVision: Deepfakes detection using human eye blinking pattern. *IEEE Access*, *8*, 83144–83154.

Karandikar, A., Deshpande, V., Singh, S., Nagbhidkar, S., & Agrawal, S. (2020). Deepfake video detection using convolutional neural network. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(2), 1311–1315.

Katarya, R., & Lal, A. (2020). A study on combating emerging threat of deepfake weaponization. *Paper presented at the 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*.

Khalil, S. S., Youssef, S. M., & Saleh, S. N. (2021). iCaps-Dfake: An integrated capsule-based model for Deepfake image and video detection. *Future Internet*, *13*(4), 93.

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, *63*(2), 135–146.

Kim, E., & Cho, S. (2021). Exposing fake faces through deep neural networks combining content and trace feature extractors. *IEEE Access*, *9*, 123493–123503.

Kohli, A., & Gupta, A. (2021). Detecting DeepFake, FaceSwap and Face2Face facial forgeries using frequency CNN. *Multimedia Tools and Applications*, *80*(12), 18461–18478.

Kong, C., Chen, B., Yang, W., Li, H., Chen, P., & Wang, S. (2021). Appearance matters, so does audio: Revealing the hidden face via cross-modality transfer. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(1), 423–436.

Lee, S., Tariq, S., Shin, Y., & Woo, S. S. (2021). Detecting handcrafted facial image manipulations and GAN-generated facial images using shallow-FakeFaceNet. *Applied Soft Computing*, *105*, 107256.

Lewis, J. K., Toubal, I. E., Chen, H., Sandesera, V., Lomnitz, M., Hampel-Arias, Z., & Palaniappan, K. (2020). Deepfake video detection based on spatial, spectral, and temporal inconsistencies using multimodal deep learning. *Paper presented at the 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*.

Li, A., Masouros, C., Swindlehurst, A. L., & Yu, W. (2021). 1-bit massive MIMO transmission: Embracing interference with symbol-level precoding. *IEEE Communications Magazine*, *59*(5), 121–127.

Li, A., Spano, D., Krivochiza, J., Domouchtsidis, S., Tsinos, C. G., Masouros, C., & Ottersten, B. (2020). A tutorial on interference exploitation via symbol-level precoding: Overview, state-of-the-art and future directions. *IEEE Communications Surveys & Tutorials*, *22*(2), 796–839.

Li, D., Ge, S. S., & Lee, T. H. (2020). Fixed-time-synchronized consensus control of multiagent systems. *IEEE Transactions on Control of Network Systems*, *8*(1), 89–98.

Li, J., Xu, K., Chaudhuri, S., Yumer, E., Zhang, H., & Guibas, L. (2017). Grass: Generative recursive autoencoders for shape structures. *ACM Transactions on Graphics*, *36*(4), 1–14.

Li, S., Liu, C. H., Lin, Q., Wen, Q., Su, L., Huang, G., & Ding, Z. (2020). Deep residual correction network for partial domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(7), 2329–2344.

Li, Y., Zhang, C., Sun, P., Ke, L., Ju, Y., Qi, H., & Lyu, S. (2021). DeepFake-o-meter: An open platform for DeepFake detection. *Paper presented at the 2021 IEEE Security and Privacy Workshops (SPW)*.

Liu, Q., & Celebi, N. (2021). Large feature mining and deep learning in multimedia forensics. *Paper presented at the Proceedings of the 2021 ACM Workshop on Security and Privacy Analytics*.

Liu, R., Wang, X., Lu, H., Wu, Z., Fan, Q., Li, S., & Jin, X. (2021). SCCGAN: Style and characters inpainting based on CGAN. *Mobile Networks and Applications*, *26*(1), 3–12.

Luo, G., Yuan, Q., Li, J., Wang, S., & Yang, F. (2022). Artificial intelligence powered mobile networks: From cognition to decision. *IEEE Network*, *36*(3), 136–144.

Luo, G., Zhang, H., Yuan, Q., Li, J., & Wang, F.-Y. (2022). ESTNet: Embedded spatial-temporal network for modeling traffic flow dynamics. *IEEE Transactions on Intelligent Transportation Systems*, *23*, 19201–19212.

Lv, Z., Chen, D., Feng, H., Zhu, H., & Lv, H. (2021). Digital twins in unmanned aerial vehicles for rapid medical resource delivery in epidemics. *IEEE Transactions on Intelligent Transportation Systems*, *23*(12), 25106–25114.

Lv, Z., Li, Y., Feng, H., & Lv, H. (2021). Deep learning for security in digital twins of cooperative intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, *22*, 4281–4290.

Lv, Z., Qiao, L., & You, I. (2020). 6G-enabled network in box for internet of connected vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *22*(8), 5275–5282.

Lv, Z., Yu, Z., Xie, S., & Alamri, A. (2022). Deep learning-based smart predictive evaluation for interactive multimedia-enabled smart healthcare. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *18*(1s), 1–20.

Marei, M., El Zaatari, S., & Li, W. (2021). Transfer learning enabled convolutional neural networks for estimating health state of cutting tools. *Robotics and Computer-Integrated Manufacturing*, *71*, 102145.

Masi, I., Killekar, A., Mascarenhas, R. M., Gurudatt, S. P., & AbdAlmageed, W. (2020). Two-branch recurrent network for isolating deepfakes in videos. *Paper presented at the European conference on Computer Vision*.

Matern, F., Riess, C., & Stamminger, M. (2019). Exploiting visual artifacts to expose deepfakes and face manipulations. *Paper presented at the 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*.

Mehta, V., Gupta, P., Subramanian, R., & Dhall, A. (2021). FakeBuster: A DeepFakes detection tool for video conferencing scenarios. *Paper presented at the 26th International Conference on Intelligent User Interfaces*.

Meskys, E., Liaudanskas, A., Kalpokiene, J., & Jurcys, P. (2020). Regulating deep fakes: Legal and ethical considerations. *Journal of Intellectual Property Law & Practice*, *15*(1), 24–31.

Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Computing Surveys (CSUR)*, *54*(1), 1–41.

Mitra, A., Mohanty, S. P., Corcoran, P., & Kougianos, E. (2020). A novel machine learning based method for deepfake video detection in social media. *Paper presented at the 2020 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS)*.

Mittal, T., Bhattacharya, U., Chandra, R., Bera, A., & Manocha, D. (2020). Emotions don't lie: An audio-visual deepfake detection method using affective cues. *Paper presented at the Proceedings of the 28th ACM International Conference on Multimedia*.

Muhammad, G., & Hossain, M. S. (2022). Light deep models for cognitive computing in intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, *24*(1), 1144–1152.

Nasar, B. F., Sajini, T., & Lason, E. R. (2020). Deepfake detection in media files-audios, images and videos. *Paper presented at the 2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*.

Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Capsule-forensics: Using capsule networks to detect forged images and videos. *Paper presented at the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Nguyen, T. T., Nguyen, C. M., Nguyen, D. T., Nguyen, D. T., & Nahavandi, S. (2019). Deep learning for deepfakes creation and detection: A survey. *arXiv Preprint arXiv:1909.11573*.

Nguyen, X. H., Tran, T. S., Nguyen, K. D., & Truong, D.-T. (2021). Learning spatio-temporal features to detect manipulated facial videos created by the Deepfake techniques. *Forensic Science International: Digital Investigation*, *36*, 301108.

Niknejad, N., Ismail, W. B., Mardani, A., Liao, H., & Ghani, I. (2020). A comprehensive overview of smart wearables: The state of the art literature, recent advances, and future challenges. *Engineering Applications of Artificial Intelligence*, *90*, 103529.

Nirkin, Y., Wolf, L., Keller, Y., & Hassner, T. (2021). DeepFake detection based on discrepancies between faces and their context. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(10), 6111–6121.

Niu, Z., Zhang, B., Dai, B., Zhang, J., Shen, F., Hu, Y., & Zhang, Y. (2022). 220 GHz multi circuit integrated front end based on solid-state circuits for high speed communication system. *Chinese Journal of Electronics*, *31*(3), 569–580.

Pashine, S., Mandiya, S., Gupta, P., & Sheikh, R. (2021). Deep fake detection: Survey of facial manipulation detection solutions. *arXiv Preprint arXiv:2106.12605*.

Pokroy, A. A., & Egorov, A. D. (2021). EfficientNets for DeepFake detection: Comparison of pretrained models. *Paper presented at the 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)*.

Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake detection: A systematic literature review. *IEEE Access*, *10*, 25494–25513.

Rana, M. S., & Sung, A. H. (2020). Deepfakestack: A deep ensemble-based learning technique for deepfake detection. *Paper presented at the 2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)*.

Saif, S., & Tehseen, S. (2022). Deepfake videos: Synthesis and detection techniques–A survey. *Journal of Intelligent & Fuzzy Systems*, *42*(4), 2989–3009.

Sanghvi, B., Shelar, H., Pandey, M., & Sisodia, J. (2021). Detection of machine generated multimedia elements using deep learning. *Paper presented at the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*.

Schlett, T., Rathgeb, C., & Busch, C. (2021). Deep learning-based single image face depth data enhancement. *Computer Vision and Image Understanding*, *210*, 103247.

Shelke, N. A., & Kasana, S. S. (2021). A comprehensive survey on passive techniques for digital video forgery detection. *Multimedia Tools and Applications*, *80*, 6247–6310.

Siegel, D., Kraetzer, C., Seidlitz, S., & Dittmann, J. (2021). Media forensics considerations on DeepFake detection with hand-crafted features. *Journal of Imaging*, *7*(7), 108.

Singh, P., Masud, M., Hossain, M. S., Kaur, A., Muhammad, G., & Ghoneim, A. (2021). Privacy-preserving serverless computing using federated learning for smart grids. *IEEE Transactions on Industrial Informatics*, *18*(11), 7843–7852.

Soares, M. A. C., & Parreiras, F. S. (2020). A literature review on question answering techniques, paradigms and systems. *Journal of King Saud University-Computer and Information Sciences*, *32*(6), 635–646.

Sun, P., Li, Y., Qi, H., & Lyu, S. (2020). Landmark breaker: Obstructing DeepFake by disturbing landmark extraction. *Paper presented at the 2020 IEEE International Workshop on Information Forensics and Security (WIFS)*.

Suratkar, S., Johnson, E., Variyambat, K., Panchal, M., & Kazi, F. (2020). Employing transfer-learning based CNN architectures to enhance the generalizability of deepfake detection. *Paper presented at the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*.

Swathi, P., & Sk, S. (2021). DeepFake creation and detection: A survey. *Paper presented at the 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*.

Tariq, S., Lee, S., & Woo, S. (2021). One detector to rule them all: Towards a general deepfake attack detection framework. *Paper presented at the Proceedings of the Web Conference 2021*.

Tjon, E., Moh, M., & Moh, T.-S. (2021). Eff-YNet: A Dual Task Network for DeepFake Detection and Segmentation. *Paper presented at the 2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM)*.

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.

Tu, Y., Liu, Y., & Li, X. (2021). Deepfake video detection by using convolutional gated recurrent unit. *Paper presented at the 2021 13th International Conference on Machine Learning and Computing*.

Vahdat, S., & Shahidi, S. (2020). D-dimer levels in chronic kidney illness: a comprehensive and systematic literature review. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences, 1-18*.

Verdoliva, L. (2020). Media forensics and deepfakes: An overview. *IEEE Journal of Selected Topics in Signal Processing*, *14*(5), 910–932.

Wang, R., Juefei-Xu, F., Huang, Y., Guo, Q., Xie, X., Ma, L., & Liu, Y. (2020). Deepsonar: Towards effective and robust detection of AI-synthesized fake voices. *Paper presented at the Proceedings of the 28th ACM International Conference on Multimedia*.

Wang, Y., Wang, H., Zhou, B., & Fu, H. (2021). Multi-dimensional prediction method based on Bi-LSTMC for ship roll. *Ocean Engineering*, *242*, 110106.

Wang, Z., Ramamoorthy, R., Xi, X., & Namazi, H. (2022). Synchronization of the neurons coupled with sequential developing electrical and chemical synapses. *Mathematical Biosciences and Engineering*, *19*(2), 1877–1890.

Wang, Z., Ramamoorthy, R., Xi, X., Rajagopal, K., Zhang, P., & Jafari, S. (2022). The effects of extreme multistability on the collective dynamics of coupled memristive neurons. *The European Physical Journal Special Topics*, *231*, 1–8.

Weerawardana, M., & Fernando, T. (2021). Deepfakes detection methods: A literature survey. *Paper presented at the 2021 10th International Conference on Information and Automation for Sustainability (ICIAfS)*.

Wijethunga, R., Matheesha, D., Al Noman, A., De Silva, K., Tissera, M., & Rupasinghe, L. (2020). Deepfake audio detection: A deep learning based solution for group conversations. *Paper presented at the 2020 2nd International Conference on Advancements in Computing (ICAC)*.

Wu, X., Zheng, W., Xia, X., & Lo, D. (2021). Data quality matters: A case study on data label correctness for security bug report prediction. *IEEE Transactions on Software Engineering*, *48*(7), 2541–2556.

Xi, Y., Jiang, W., Wei, K., Hong, T., Cheng, T., & Gong, S. (2021). Wideband RCS reduction of microstrip antenna array using coding metasurface with low Q resonators and fast optimization method. *IEEE Antennas and Wireless Propagation Letters*, *21*(4), 656–660.

Xiao, Y., Zhang, Y., Kaku, I., Kang, R., & Pan, X. (2021). Electric vehicle routing problem: A systematic review and a new comprehensive model with nonlinear energy recharging and consumption. *Renewable and Sustainable Energy Reviews*, *151*, 111567.

Xiao, Y., Zuo, X., Huang, J., Konak, A., & Xu, Y. (2020). The continuous pollution routing problem. *Applied Mathematics and Computation*, *387*, 125072.

Xie, B., Li, S., Lv, F., Liu, C. H., Wang, G., & Wu, D. (2023). A collaborative alignment framework of transferable knowledge extraction for unsupervised domain adaptation. *IEEE Transactions on Knowledge and Data Engineering*, *35*(7), 6518–6533.

Xie, D., Chatterjee, P., Liu, Z., Roy, K., & Kossi, E. (2020). DeepFake detection on publicly available datasets using modified AlexNet. *Paper presented at the 2020 IEEE Symposium Series on Computational Intelligence (SSCI)*.

Xu, K.-D., Weng, X., Li, J., Guo, Y.-J., Wu, R., Cui, J., & Chen, Q. (2022). 60-GHz third-order on-chip bandpass filter using GaAs pHEMT technology. *Semiconductor Science and Technology*, *37*(5), 055004.

Xu, Z., Liu, J., Lu, W., Xu, B., Zhao, X., Li, B., & Huang, J. (2021). Detecting facial manipulated videos based on set convolutional neural networks. *Journal of Visual Communication and Image Representation*, *77*, 103119.

Yan, A., Fan, Z., Ding, L., Cui, J., Huang, Z., Wang, Q., & Wen, X. (2021). Cost-effective and highly reliable circuit-components design for safety-critical applications. *IEEE Transactions on Aerospace and Electronic Systems*, *58*(1), 517–529.

Yan, J., Jiao, H., Pu, W., Shi, C., Dai, J., & Liu, H. (2022). Radar sensor network resource allocation for fused target tracking: A brief review. *Information Fusion*, *86*, 104–115.

Yan, L., Yin-He, S., Qian, Y., Zhi-Yu, S., Chun-Zi, W., & Zi-Yun, L. (2021). Method of reaching consensus on probability of food safety based on the integration of finite credible data on block chain. *IEEE Access*, *9*, 123764–123776.

Yang, C.-Z., Ma, J., Wang, S.-L., & Liew, A. W.-C. (2020). Preventing deepFake attacks on speaker authentication by dynamic lip movement analysis. *IEEE Transactions on Information Forensics and Security*, *16*, 1841–1854.

Yang, D., Zhu, T., Wang, S., Wang, S., & Xiong, Z. (2022). LFRSNet: A robust light field semantic segmentation network combining contextual and geometric features. *Frontiers in Environmental Science*, *10*, 996513.

Yang, J., Xiao, S., Li, A., Lan, G., & Wang, H. (2021). Detecting fake images by identifying potential texture difference. *Future Generation Computer Systems*, *125*, 127–135.

Yazdinejad, A., Parizi, R. M., Srivastava, G., & Dehghantanha, A. (2020). Making sense of blockchain for AI deepfakes technology. *Paper presented at the 2020 IEEE Globecom Workshops (GC Wkshps)*.

Yu, I.-J., Nam, S.-H., Ahn, W., Kwon, M.-J., & Lee, H.-K. (2020). Manipulation classification for jpeg images using multi-domain features. *IEEE Access*, *8*, 210837–210854.

Yu, P., Xia, Z., Fei, J., & Lu, Y. (2021). A survey on Deepfake video detection. *IET Biometrics*, *10*(6) 607–624.

Zadeh, F. A., Bokov, D. O., Yasin, G., Vahdat, S., & Abbasalizad-Farhangi, M. (2023). Central obesity accelerates leukocyte telomere length (LTL) shortening in apparently healthy adults: A systematic review and meta-analysis. *Critical Reviews in Food Science and Nutrition*, *63* (14), 2119–2128.

Zhang, M., Chen, Y., & Lin, J. (2021). A privacy-preserving optimization of neighborhood-based recommendation for medical-aided diagnosis and treatment. *IEEE Internet of Things Journal*, *8*(13), 10830–10842.

Zhang, M., Chen, Y., & Susilo, W. (2020). PPO-CPQ: A privacy-preserving optimization of clinical pathway query for e-healthcare systems. *IEEE Internet of Things Journal*, *7*(10), 10660–10672.

Zhang, W., Zhao, C., & Li, Y. (2020). A novel counterfeit feature extraction technique for exposing face-swap images based on deep learning and error level analysis. *Entropy*, *22*(2), 249.

Zhang, Y., Gao, F., Zhou, Z., & Guo, H. (2021). A survey on face forgery detection of Deepfake. *Paper presented at the Thirteenth International Conference on Digital Image Processing (ICDIP 2021)*.

Zhang, Z., Luo, C., & Zhao, Z. (2020). Application of probabilistic method in maximum tsunami height prediction considering stochastic seabed topography. *Natural Hazards*, *104*(3), 2511–2530.

Zhao, H., Zhou, W., Chen, D., Wei, T., Zhang, W., & Yu, N. (2021). Multi-attentional deepfake detection. *Paper presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Zhao, Z., Wang, P., & Lu, W. (2021). Multi-layer fusion neural network for Deepfake detection. *International Journal of Digital Crime and Forensics (IJDCF)*, *13*(4), 26–39.

Zheng, W., Liu, X., Ni, X., Yin, L., & Yang, B. (2021). Improving visual reasoning through semantic representation. *IEEE Access*, *9*, 91476–91486.

Zheng, W., Liu, X., & Yin, L. (2021). Sentence representation method based on multi-layer semantic network. *Applied Sciences*, *11*(3), 1316.

Zheng, W., Xun, Y., Wu, X., Deng, Z., Chen, X., & Sui, Y. (2021). A comparative study of class rebalancing methods for security bug report classification. *IEEE Transactions on Reliability*, *70*(4), 1658–1670.

Zheng, W., Yin, L., Chen, X., Ma, Z., Liu, S., & Yang, B. (2021). Knowledge base graph embedding module design for visual question answering model. *Pattern Recognition*, *120*, 108153.

Zhong, L., Fang, Z., Liu, F., Yuan, B., Zhang, G., & Lu, J. (2021). Bridging the theoretical bound and deep algorithms for open set domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems*.

Zhou, L., Fan, Q., Huang, X., & Liu, Y. (2023). Weak and strong convergence analysis of Elman neural networks via weight decay regularization. *Optimization*, *72*(9), 2287–2309. https://doi.org/10.1080/02331934.2022.2057852

Zhou, X., Wang, Y., & Wu, P. (2020). Detecting deepfake videos via frame serialization learning. *Paper presented at the 2020 IEEE 3rd International Conference of Safe Production and Informatization (IICSPI)*.

Zi, B., Chang, M., Chen, J., Ma, X., & Jiang, Y.-G. (2020). Wilddeepfake: A challenging real-world dataset for deepfake detection. *Paper presented at the Proceedings of the 28th ACM International Conference on Multimedia*.

Zong, C., & Wan, Z. (2022). Container ship cell guide accuracy check technology based on improved 3D point cloud instance segmentation. *Brodogradnja: Teorija i Praksa Brodogradnje i Pomorske Tehnike*, *73*(1), 23–35.

Zong, C., & Wang, H. (2022). An improved 3D point cloud instance segmentation method for overhead catenary height detection. *Computers & Electrical Engineering*, *98*, 107685.

How to cite this article: Heidari, A., Jafari Navimipour, N., Dag, H., & Unal, M. (2024). Deepfake detection using deep learning methods: A systematic and comprehensive review. *WIREs Data Mining and Knowledge Discovery*, *14*(2), e1520. https://doi.org/10.1002/widm.1520