S. No.	Authors & Year	Dataset	Features Used	Techniques Implemented	Outcome	Parameter	Advantages	Disadvantages	Results
[1]	Siddharth Solaiyappan et al. (2022)	LIDC-IDRI CT-GAN	local and global characteristics	Support Vector Machine (SVM), Random Forest, Decision Tree, DenseNet121, DenseNet201, ResNet50, ResNet101, and VGG19,	The main purpose of this work is to detect deepfake tampering in medical photos by detecting injected or excised tumours. The outcome assesses how well machine learning systems classify manipulated images.	Accuracies: SVM-0.591, Random Forest-0.665, Decision Tree-0.641, DenseNet121-0.804, DenseNet201-0.777, ResNet50-0.641, ResNet101-0.706, VGG19-0.657	Robust Detection, Diverse Approach, Higher Accuracy	Generalization and limited details	high potential for real-world application in identifying deepfake attacks in medical imaging scenarios.
[2]	Fabi Prezja et al. (2022)	320K DeepFake KOA Images, 5,556 real X-ray images and 3,20,000 synthetic X-ray (DeepFake) images	structure, density, and alignment features of the knee joint bones	Generator-discriminator using GAN for classification	This work shows the potential of medically accurate synthetic data to address privacy issues and data limitations and the efficacy of these synthetic images in	Loss of 3.79% from baseline accuracy	Privacy Preservation, Data Augmentation and Reduced Regulatory Hurdles	Realism vs. Real Data, Ethical Considerations	The study shows that generative adversarial neural networks may produce medically correct knee joint X-rays. As indicated in the survey, more DeepFakes were mistaken for real photos by medical

					improving osteoarthritis severity assessment tasks.				specialists. Even when replaced by complete training data, synthetic images enhance osteoarthritis severity classification accuracy. This shows how synthetic data can improve medical image analysis while addressing data restrictions and privacy issues.
[3]	Yalamanchili Salini et al. (2022)	DRIVE retinal fundus image dataset	Retinal features like retina, blood vessels etc.	Cycle-GAN framework which includes generation and segmentation.	This study generates synthetic retinal fundus images using the Cycle-GAN architecture, evaluates privacy and hallucination problems, and examines GAN potential and limitations in medical imaging.	Accuracy of 98.19% for the proposed Cycle-GAN model.	Data Accessibility, Privacy Preservation and Innovative approach	Realism and Generalization, Ethical and Clinical Considerations	The Cycle-GAN framework generates and segments retinal fundus pictures in the study. It shows how GANs may solve the medical data shortage and create authentic synthetic images. The suggested model's 98.19% accuracy

									implies it can classify retinal fundus images. For a complete evaluation, further information regarding the classification task, segmentation, and limitations is needed.
[4]	Rajat Budhiraja et al. (2022)	LIDC-IDRI CT scans, X-rays and ultrasound images of a miniscule subset with <100 images.	visual characteristics and patterns like textures, shapes, edges, and structures	Convolutional Reservoir Networks (CoRN)	This study aims to detect deepfake medical images that introduce dangerous tumours into healthy people faster. The work seeks to identify such tampering attacks in a fast and lightweight manner to protect patients'	DenseNet, ResNet, VGG and RC combined architecture increased accuracy to more than 90%	Detection Enhancement, Lightweight Solution, and Generalization.	Dataset Size and Limited Details.	The study uses Convolutional Reservoir Networks (CoRN) to detect deepfake medical imaging inserted with malignant tumours into healthy patients' modalities. It is practical and lightweight. The method reportedly improves categorization metrics. The

					privacy and medical data.				study shows its efficacy with a limited dataset, but further information regarding the technique's performance measures and real-world applications would help explain the results.
[5]	Zeba Ghaffar et al. (2022)	COVID-19 positive and normal cases as well as potential cases of other respiratory diseases.	Texture and density features by grey level and local level	CNN along with MobileNet, EfficientNet, and InceptionV3	The work aims to completely evaluate various CNN architectures' ability to automatically detect COVID-19 infection from chest X-rays. The study seeks to accelerate pandemic testing with novel and efficient technologies. The writers seek infectious respiratory	Accuracies: MobileNet-95% EfficientNet-95% and InceptionV3-94%	Rapid Detection, Non-invasive and scalability.	Limited to Imaging, Variability and Ethical Considerations	The trained CNN models MobileNet, EfficientNet, and InceptionV3 classified COVID-19-infected chest X-ray pictures accurately. These results suggest that these models could help physicians and radiologists speed up COVID-19 testing and diagnosis, improving patient outcomes and

					disease control and prevention methods.				pandemic control.
[6]	Lingzhi Kong et al. (2022)	Pneumonia Healthy COVID-19 Influenza (Lung)	High-level representations capturing complex patterns, textures, and structures within the chest X-ray images.	Feature Fusion, Attention Mechanism and Residual Network	The study proposes an enhanced chest X-ray image classification algorithm for COVID-19 identification. The researchers want to show that their model can accurately categorise X-ray pictures and improve medical diagnosis.	For binary classification (likely COVID-19 positive vs. non-infected), the average accuracy reaches 98.0%. For three-category classification (which might involve categories like COVID-19 positive, other infections, and non-infected), the average accuracy reaches 97.3%.	High Accuracy, Attention Mechanism and Clinical Support.	Dependency on Data Quality, Limited to X-ray Data, and Generalization.	The experimental results show that the proposed model, which combines DenseNet and VGG16 features with attention mechanisms and ResNet segmentation, classifies COVID-19 chest X-ray pictures accurately. This suggests that deep learning, feature fusion, and attention mechanisms can help doctors and radiologists diagnose chest disorders like COVID-19 quickly and accurately.
[7]	Yi-Yang Liu et al. (2022)	KUB images from 104 patients from Kaohsiung Chang	high-level visual representations learned from	Computer-Aided Diagnosis (CAD) system	The work aims to create and propose a	Validation Set: Accuracy 0.977, Sensitivity (True Positive Rate)	Accuracy, Reduced Radiation	External Validation, clinical interpretation	Experimental results show that the deep learning model

		Gung Memorial Hospital	pre-processed KUB images.	using Deep learning, model training, parameter tuning and testing.	deep learning-based CAD system for accurately diagnosing urolithiasis from KUB images. The project attempts to help non-expert clinicians make accurate diagnosis without radiological reports.	0.953, Specificity (True Negative Rate) 1, F1-Measure 0.976. Testing Set: Accuracy 0.982, Sensitivity 0.964, Specificity 1, F1-Measure 0.982.	Exposure and cost effective.	and Dependency on Data Quality.	accurately detects urolithiasis in KUB images. The model's excellent accuracy and other performance measures imply it could help emergency room clinicians make accurate diagnoses quickly, reducing radiation exposure and medical expenditures from unneeded CT scans. Comparing the model against CNN-based approaches shows its efficacy.
[8]	Mohammad Monirujjaman Khan et al. (2022)	Flickr-Faces-HQ dataset, 70,000 real images by Nvidia and 70,000 fake images by styleGAN model. Fake images of 256 pixels.	gray level	CNN with nceptionResNetV2, DenseNet201, InceptionV3, and ResNet152V2. Local Interpretable	The CNN-based deepfake detection method is the study's main result. The study uses LIME to	Accuracy: InceptionV3- 99.68% ResNet152V2- 99.19% DenseNet201- 99.81% InceptionResNetV2- 99.87%	High Accuracy and Explainability.	Adversarial Attacks and Generalization	The most accurate model was InceptionResNetV2 at 99.87%. This technique had the highest accuracy and explainability

				Model-Agnostic Explanations (LIME) algorithm is also used.	accurately distinguish real and deepfake photos and explain the model's decisions.				with the LIME algorithm for Explainable AI (XAI). The study shows that the suggested method detects deepfake images reliably. The study uses CNN models to detect deepfake photos with excellent accuracy. The LIME algorithm for XAI improves transparency and interpretability, making the detection system more reliable. The results show that the proposed technique, especially with InceptionResNetV2, might solve the deepfake content problem.

[9]	R. Saravana Ram et al. (2022)	Group-wise Deep Whitening, Coloring Method (GDWCT), StarGAN, StyleGAN, StyleGAN2	Computer vision features related to digital content integrity.	Fuzzy Clustering Feature Extraction Method, Deep Belief Network (DBN) with Loss Handling and Pairwise Learning Approach.	The paper's main result is a better deepfake image and video detection algorithm. The suggested method uses feature extraction, deep learning, and pairwise learning to improve content manipulation detection.	Detection rate accuracy reported increase of 98%	High Detection Rate and Integration of Techniques	Data Variability and Generalization	The paper shows that fuzzy clustering, deep belief networks with loss handling, and paired learning improve deepfake detection accuracy by 98% across datasets. In conclusion, the paper improves deepfake detection by merging computer vision and deep learning methods to recognise faked content. The results' significant increase in detection accuracy supports the methodology.
[10]	Thanh Thi Nguyen et al. (2022)	UADFV FaceForensics++ Celeb-DF VidTIMIT	Visual cues, artifacts, inconsistencies, and statistical	Algorithms and architectures used in	The paper's main result is a survey and study of	Research study	Comprehensive Overview and Research Direction.	Lack of Original Research and	The paper covers deepfake production and

		CelebA-HQ GDWCT	irregularities in images.	generating realistic fake images and videos using deep learning. Techniques for transferring facial expressions, gestures, and speech from one person to another.	deepfake development and detection techniques. It seeks a complete grasp of the field's current situation.			Rapidly Evolving Field.	detection algorithms and methods. It examines problems, research trends, and the changing landscape of deepfake technology, improving knowledge and guiding future research.
[11]	Andreas Rossler et al. (2019)	Forensics++ Forensic Analysis	The paper employs Convolutional Neural Networks (CNNs) to learn features from manipulated facial images. They use a combination of raw pixel values and handcrafted features, like color histograms and noise patterns.	The authors utilize various deep learning architectures, including Inception ResNet, to train their model on a large dataset of real and manipulated facial images. They also implement data augmentation techniques to improve model generalization.	The primary outcome of the paper is the development of a deep learning-based approach capable of detecting manipulated facial images. The proposed model outperforms traditional forensics methods and other deep learning models in terms of	The specific details of the model's parameters are not provided in the summary. However, the paper likely includes information about the architecture's layers, filter sizes, activation functions, learning rates, etc.	High Accuracy, High Accuracy, Automated Detection	Dependency on Data, Adversarial Attacks	The paper demonstrates that the proposed model outperforms existing methods for detecting manipulated facial images. It achieves high accuracy in distinguishing real images from manipulated ones, even when faced with various manipulation techniques. The experimental

					accuracy and generalization.				results showcase the effectiveness of their approach on a benchmark dataset.
[12]	Brian Dolhansky et al. (2020)	DFDC dataset with 100,000 video clips sourced from 3,426 paid actors.	Frame-level pixel data, facial landmarks, grayscale values, local texture patterns, and structural information from images.	GAN	The large and diversified DFDC collection of face-swapped videos is the result of this study. The dataset supports deepfake detection model training and evaluation. The paper shows that models trained purely on the DFDC dataset can generalise to "in the wild." deepfake films, notwithstanding their difficulties to detect.	MM/NN, DFAE, FSGAN, NTH, StyleGAN with each column shows increasing quality from left to right.	Large and Diverse Dataset and ethical considerations	Lack of Detailed Techniques and privacy concerns.	The DFDC dataset shows that models can generalise to real-world deepfake videos. This shows that the dataset and approach may be promising for deepfake detection.

[13]	Vincent Nozic et al. (2018)	DeepFake Face2Face	The paper employs deep learning features extracted from facial images using a small CNN architecture.	The authors propose MesoNet, a shallow CNN architecture, for detecting facial video forgeries. They train the network on a dataset of manipulated and authentic facial images.	The primary outcome of the paper is the development of MesoNet, which is a compact neural network capable of detecting facial video forgeries with a focus on model efficiency.	Details about the specific parameters of the MesoNet architecture (e.g., number of layers, filter sizes, activation functions) can be found in the paper.	Compact Architecture, Efficiency, Real-Time Detection	Limited Complexity, Dependence on Dataset	The paper demonstrates that MesoNet is effective in detecting facial video forgeries, achieving competitive results compared to more complex architectures. The network's compact design allows for efficient deployment while maintaining good detection performance.
[14]	Ricard Durall et al. (2020)	CelebA Forensics++ CelebA-HQ Flickr-Faces-HQ 5,639 high-quality DeepFake videos.	pixel-level information, frame-level patterns, facial landmarks with local texture details, and structural characteristics	GAN	This effort produced the Celeb-DF dataset, which provides high-quality, realistic DeepFake movies. This dataset is essential for DeepFake detection algorithm creation,	Robust	Realistic Quality and Comprehensive Evaluation.	Potential Bias and Privacy Concerns.	The Celeb-DF dataset raises the bar for DeepFake detection problems, demonstrating its potential to advance DeepFake detection systems.

					training, and evaluation.				
[15]	Lingzhi Li et al. (2020)	FaceForensics++ FaceForensics	Grayscale Face X-ray images	The Face X-ray computation is the key method in this study. The paper does not describe Face X-ray's technique, although it may use image processing and computer vision to identify and visualise face image blending boundaries. The paper also advises training the system without using state-of-the-art face alteration tools' phoney photos.	Facial X-ray was proposed as an effective tool for identifying facial image counterfeiting. It shows how blending boundaries, used in many face manipulation techniques, can be used to detect face manipulations.	Detection accuracy of 97.73% for F2F and 85.69% for FS. AUC: FF++ - 98.52 DFD- 93.47	General Applicability, Effectiveness, Greyscale Image and Transparency.	Lack of Specifics, Limited Dataset Information and Limited Evaluation Information.	Face X-ray outperforms most face forgery detection or deepfake detection algorithms in detecting forgery generated by various face manipulation techniques, including those not seen during training. The research emphasises Face X-ray's efficacy.
[16]	Davide Coccomini et al. (2022)	FaceForensics++ DFDC	CNN, EfficientNet B0 as a feature extractor.	Variational Autoencoders (VAEs), GAN, EfficientNet B7, Vision Transformers	The research focuses on face expression deepfakes in video material. The	AUC: 0.951 F1 Score: 88.0%	High Accuracy, focus on facial expressions and efficient approach.	Lack of Dataset Information, Lack of Technical Details and	The best model detected deepfakes with an AUC of 0.951 and an F1 score of 88.0% on the DFDC

				and voting system.	top model had an AUC of 0.951 and an F1 score of 88.0% on the DFDC dataset. These results are close to the state-of-the-art, proving the suggested technique detects deepfakes.			Limited Discussion on Challenges.	dataset. The paper does not compare the model to existing approaches or assess its robustness to different deepfake modifications.
[17]	Ruben Tolosana et al. (2020)	ImageNet CelebA FFHQ CelebA-HQ FaceForensics++ Celeb-DF DFDC FaceForensics	GAN	Fake detection and deepfake detection methods.	The survey covers facial picture manipulation and detection methods, including DeepFakes. Facial manipulation includes whole face synthesis, identity switch (DeepFakes), attribute manipulation, and expression exchange. The survey	Summarizes the results from multiple studies.	Comprehensive Coverage, Focus on DeepFakes, Addresses Societal Implications & Highlights Public Databases	Limited Discussion of Open Issues.	The survey may reveal the present state of the art and trends in the sector, which can inform future research and development to address bogus material in society.

					provides a comprehensive grasp of technology evaluation methods, databases, and benchmarks.				
[18]	Nicolò Bonettin et al. (2020)	FaceForensics++ DFDC ImageNet, more than 119,000 videos to tackle the problem.	Deep learning-based features extracted from video frames	Ensembling of CNN Models, Base Network (EfficientNetB4), Attention Layers & Siamese Training.	The work addresses the essential topic of identifying face modification in video sequences, particularly contemporary facial manipulation techniques. Combining CNN models yielded good results for recognising altered faces in videos.	On datasets- DFDC: Model - EfficientNetB4 + EfficientNetB4ST + B4Att, Metric name – LogLoss, Metric value – 0.4640. FaceForensics++: Model - EfficientNetB4 + EfficientNetB4ST + B4Att + B4AttST, Metric name – AUC, Metric value - 0.9444. FaceForensics++: Model - EfficientNetB4 + EfficientNetB4ST + B4AttST, Metric name – LogLoss, Metric value - 0.3269.	Addresses a Societal Issue, Large Dataset and Ensembling Approach.	Limited Discussion on challenges.	It implies that ensembling CNN models with attention layers and siamese training can detect face alteration in video sequences.
[19]	Hasam Khalid et al. (2021)	FakeAVCeleb VoxCeleb2 Celeb-DF DFDC	multimodal	Deepfake Generation Methods and Multimodal	Its main result is the FakeAVCeleb dataset,	Accuracy (best): Model - Xception-comp UADFV - 91.2	Addressing Emerging Threats, Racial Bias Mitigation	Limited Technical Information	It introduces FakeAVCeleb, a new Audio-Video

		DeeperForensics KoDF, deepfake videos and synthesized lip-synced fake audios created using popular deepfake generation methods and real YouTube videos of celebrities with four different ethnic backgrounds.		Deepfake Detection.	which provides high-quality data to detect audio and video deepfakes concurrently. The dataset seeks realism, ethnic diversity, and racial bias-free data.	DF-TIMIT LQ - 95.9 DF-TIMIT HQ - 94.4 FF-DF - 99.7 DFD - 85.9 Celeb-DF - 65.3 FakeAvCeleb - 72.5 Model - Meso4 DFDC - 72.2	and Realistic Multimodal Data.		multimodal dataset that can detect both deepfake videos and audios. FakeAVCeleb includes cloned audios and deepfake videos. Authors created FakeAVCeleb to be gender and racially impartial, featuring footage of men and women from four major races across various age categories. Also used various common deepfake video and audio creation technologies to create almost lip-synced videos and audios. Evaluated FakeAVCeleb dataset against seven other

									deepfake detection datasets. Conducted tests employing cutting-edge approaches in unimodal, ensemble-based, and multimodal environments (see to Appendix C for findings). FakeAVCeleb aims to strengthen deepfake detectors and equip researchers with a solid base for multimodal detectors.
[20]	Joel Frank et al. (2021)	WaveFake LFSpeech with 13,100 short audio clips JSUT with basic5000 corpus	Spectral Features.	Deepfake Detection Dataset Creation, Frequency Statistics Analysis and Implementation of Classifiers.	This study created a dataset for audio deepfake identification, filling a gap in research that has mostly concentrated on image-based	Neural network-based approaches performed better on average, but more traditional models proved to be more robust.	Filling a Research Gap, Novel Dataset and Baseline Classifiers.	Absence of Specificity	The study underpins audio signal research. The current TTS landscape, signal processing methods, and feature formats were covered first. Next, they

					detection. The research also sheds light on frequency statistics across network designs and gives baseline classifiers for detection technique development.				presented their main contribution: a new data set with samples from six cutting-edge architectures in two languages. They identified modest differences between models by visualising the frequency spectrum, especially at higher frequencies. They next calculated each data set's average energy per frequency using prosody analysis. The study indicates that all models approximate training data, yet differences remain. To prepare future practitioners, they trained numerous baseline

									models. Their performance was evaluated in numerous data sets and circumstances. Our GMM and neural network solutions were trained. Although neural networks performed better, GMM classifiers were more robust, which may be a benefit in real life. The final classifier analysis used attribution. It was found that high-frequency and low-frequency information are important.
[21]	Sanjay Saha et al. (2023)	FaceForensics++ Celeb-DF FSh NT DF F2F	Embeddings Vision Transformer-based spatial features and a Timeseries Transformer for learning temporal	Deepfake Detection Method, Benchmark Dataset Creation, Vision Transformer (ViT) and	The main result of this research is a deepfake detection algorithm that can detect small changes in	IoU and AuC of more than 90% for all the datasets	Novel Benchmark Dataset, Spatial and Temporal Features and Potential for Targeted Moderation.	Limited Technical specificity	The research describes a deepfake detection algorithm that can detect small changes in generative films. It also

			features from the videos	Timeseries Transformer.	generated movies. The approach identifies suspicious deepfakes at the frame and video levels to moderate them.				mentions a new benchmark dataset for evaluation.
[22]	Peter Lorenz et al. (2023)	CIFAR-10 ImageNet CelebA-HQ LAION-5B Diffusion DB ArtiFact CIFAKE use of known and newly created datasets for extensive experiments on diffusion detection and model identification.	multi–Local Intrinsic Dimensionality (multiLID)	Diffusion Model Detection, Identification of Generator Networks and Benchmark for Diffusion-Generated Images.	The proposed method for automatically detecting diffusion model-generated synthetic pictures and generator networks is the main result of this research. The research also sets a baseline for diffusion-generated picture detection, which advances diffusion model detection.	A total avg of 1.00 was found for both detection and identification on the LSUN-bedroom dataset	Addressing Malicious Use, Superior Detection and Comprehensive Benchmark.	Limited Technical Information about MultiLID, the 'real' class was troublesome.	The paper describes the development of a method for detecting synthetic images generated by diffusion models and identifying the corresponding generator networks. It also highlights the creation of a benchmark for diffusion-generated images.

[23]	Komal Chugh et al. (2020)	DFDC DeepFake-TIMIT	Modality Dissonance Score (MDS), which quantifies the dissimilarity between audio and visual modalities in a video.	Modality Dissonance Score (MDS), Feature Learning and Loss Functions.	The Modality Dissonance Score-based deepfake detection system is the main result of this research. The method quantifies audio-visual dissimilarity to identify deepfake videos. The paper also shows temporal forgery localization, which can identify modified video parts.	The approach outperforms the state-of-the-art by up to 7%.	Novel Detection Approach, Improved Performance and Temporal Forgery Localization.	Generalisation	The paper describes the proposal of a deepfake detection method based on Modality Dissonance Score (MDS) and highlights its improved performance compared to the state-of-the-art. It also mentions the capability of the method to localize manipulated video segments.
[24]	Yihao Huang et al. (2020)	FFHQ LSUN	16 popular GAN-based fake image generation techniques	FakePolisher and Evaluation	This research produced FakePolisher, a technology that reduces artefact patterns in synthesised photos, making deepfake detection approaches	Average reduction of 47% and up to 93% in the worst case	Reduction of Artifacts, and Impact on Detection Methods.	Potential Bias	In order to fool deepfake detection systems, the authors of this work discuss the creation of FakePolisher, a tool for smoothing out artefact patterns in computer-generated

					harder to detect.				photographs. It illustrates how FakePolisher significantly decreased detection accuracy when compared to three state-of-the-art detection approaches.
[25]	Oscar de Lima et al. (2020)	Celeb-DF V2 for evaluating spatiotemporal convolutional	spatial and temporal features	Spatiotemporal Convolutional Method, Benchmark Creation and Performance Evaluation	The fundamental contribution of this study is the development of the Celeb-DF dataset as a baseline for testing spatiotemporal convolutional algorithms for deepfake detection. These techniques are meant to make use of both geographical and temporal data to enhance	Method-ROC-AUC%: RCN – 74.87 R2Plus – 99.43 I3D – 97.59 MC3 – 99.30 R3D – 99.73 Accurcay%: RCN – 76.25 R2Plus – 98.07 I3D – 92.28 MC3 – 97.49 R3D – 98.26	Temporal Information Utilization, Benchmark Creation and Improved Performance	Limited dataset	The paper describes the creation of a benchmark to evaluate spatiotemporal convolutional methods for deepfake detection using the Celeb-DF dataset. It highlights the potential advantages of incorporating temporal information and mentions the outperformance of frame-based detection methods.

					deepfake detection.				
[26]	Ning Yu et al. (2021)	CelebA aligned and cropped images LSUN Bedroom LSUN cat CIFAR-10 CycleGAN horse2zebra AFHQ Cat and Dog	artificial fingerprints into generative models	Artificial Fingerprints, Deepfake Detection and Attribution	The key contribution of this study is the incorporation of synthetic fingerprints into generative models, which provides a proactive and long-term solution for deepfake identification. The goal of this method is to strengthen deepfake detection and attribution so that it can withstand adversarial methods and dynamic generative models.	Model & fingerprint bitwise accuracy: ProGAN trained on 150k fingerprinted CelebA 128x128 – 0.98 ProGAN trained on 50k fingerprinted LSUN Bedroom 128x128 – 0.93 ProGAN trained on 50k fingerprinted LSUN Cat 256x256 – 0.98 StyleGAN trained on 150k fingerprinted CelebA 128x128 – 0.99 StyleGAN trained on 50k fingerprinted LSUN Bedroom 128x128 – 0.98 StyleGAN trained on 50k fingerprinted LSUN Cat 256x256 – 0.99 StyleGAN2 trained on 150k fingerprinted CelebA 128x128 – 0.99	Proactive Deepfake Detection, Robustness and Transferability, Trivial Deepfake Detection and Attribution	Realism vs real data	It explains how artificial fingerprints were included into generative models as a preventative and long-term method for detecting deepfakes. It highlights the approach's success in simplifying deepfake detection and attribution while also highlighting its stability and portability.

						StyleGAN2 trained on 50k fingerprinted LSUN Bedroom 128x128 – 0.99 StyleGAN2 trained on 50k fingerprinted LSUN Cat 256x256 – 0.99 BigGAN trained on fingerprinted CIFAR10 32x32 – 0.99 CUT trained on fingerprinted horse2zebra 256x256 – 0.99 CUT trained on fingerprinted AFHQ cat2dog 256x256 – 0.99			
[27]	Tiziano Fagni et al. (2020)	TweepFake – 25,572 tweets dataset which is the first dataset of real deepfake tweets. It consists of tweets collected from a total of 23 bots, imitating 17 human accounts.	Bot-generated tweets	Collection of Real Deepfake Tweets, Evaluation of Deepfake Text Detection Methods and Dataset Publication.	This research produced the TweepFake dataset, which is useful for developing and testing deepfake text detection algorithms on social media postings. The evaluation of 13 detection	Global accuracy- Log_reg_bow - 0.804 Rand_forest_bow - 0.772 Svc_bow - 0.811 Log_reg_bert - 0.835 Rand_forest_bert - 0.827 Svc_bert - 0.842 Char_cnn - 0.851 Char_gru - 0.830 Char_cnngru - 0.837	Real Deepfake Tweets, Diverse Generation Techniques and Public Availability	Curse of dimensionality and it ignores the information about word order	The study describes the creation of the TweepFake dataset, the first dataset of real deepfake tweets collected from Twitter. It also highlights the evaluation of 13 deepfake text detection

					technologies also sets a baseline for future research.	Bert_ft - 0.891 Distilbert_ft - 0.887 Roberta_ft - 0.896 Xlnet_ft – 0.877			methods using the dataset.
[28]	Shahroz Tariq et al. (2020)	Facial Reenactment: Face2Face (F2F) Neural Texture (NT) Identity Swap: DeepFake (DF) FaceSwap (FS) DeepFake Detection (DFD) DeepFake Detection Challenge (DFDC) Unknown: DeepFake in the Wild (DFW)	temporal information	Single Frame-based ShallowNet Xception MesoNet Multiple Frame-based CNN+LSTM DBiRNN **CLRNet**	This research produced CLRNet, a Convolutional LSTM-based Residual Network that detects deepfake films using temporal information from consecutive image sequences. The transfer learning-based method overcomes deepfake detection's generalizability issue.	Average accuracy CLRNet in the detection performance comparison of state-of-the-art deepfake detection methods against in domain attack – 98.61. The single domain trained detector performance against out of domain attack and DFW when Dt = DFDC : DFDC – 96.76 ± 0.2 DF – 51.58 ± 0.1 FS – 51.98 ± 0.1 DFD – 53.62 ± 0.2 NT – 59.70 ± 0.2 F2F – 53.83 ± 0.2 DFW – 51.43 ± 0.1	Temporal Information Utilization, Generalizability and Empirical Performance.	Space, single dataset performance	The paper describes the development of CLRNet, a deep learning-based model for detecting deepfake videos by leveraging temporal information from sequences of consecutive images. It also emphasizes the improved generalizability of the method and its outperformance of five state-of-the-art deepfake detection methods
[29]	Javier Hernandez-Ortega et al. (2020)	Celeb-DF DFDC	Physiological measurements related to heart rate using remote photoplethysmography (rPPG)	DeepFake Detection Framework, CNN, Experimental Evaluation	This research introduced DeepFakesON-Phys, a deepfake detection system that	AUC values above 98% on both the Celeb-DF and DFDC databases	Physiological Measurement Utilization, High Detection Performance and State-of-the-Art Results.	Lack of Dataset Details	The paper introduces a novel deepfake detection framework, DeepFakesON-Phys, based on

					uses physiological measures, specifically heart rate information from rPPG, to improve video identification. AUC values over 98% on both available datasets show that this strategy works.				physiological measurements, specifically heart rate information obtained through rPPG. It emphasizes the high detection performance of the method, with AUC values exceeding 98% on the Celeb-DF and DFDC databases, outperforming the state of the art.
[30]	Wanjun Zhong et al. (2020)	WebText style GPT2 dataset and news-style GROVER generated dataset two public deepfake datasets for experimentation	factual structures, it suggests that coarse-grained representations used in existing deepfake detection methods struggle to capture these factual structures.	Graph-Based Model, Sequential Modeling of Relations & Experimental Evaluation	This research produced a graph-based model for deepfake text detection that uses document factual structure to discriminate. The model's ability to identify machine-generated text from	Improves strong base models built with RoBERTa	Factual Structure Utilization, Graph-Based Representation and Improved Performance.	Limited Technical Information	It introduces a graph-based model for deepfake text detection that leverages the factual structure of documents as a discriminative factor. It emphasizes the improvement in performance over strong base models

					factual-structured human material is supported by significant improvements over strong base models.				built with RoBERTa.
[31]	Tianchen Zhao et al. (2021)	ImageNet FaceForensics++ DFDC	The paper emphasizes the utilization of source feature inconsistency within deepfake images as a cue for detection. It suggests that distinct source features can be preserved and extracted even after deepfake generation processes.	Pair-wise Self-Consistency Learning (PCL), and Inconsistency Image Generator (I2G)	Based on source feature discrepancy, this research developed a method to detect deepfake images. PCL and I2G are used to improve deepfake picture detection. The experimental results show improved AUC in in-dataset and cross-dataset evaluations above the state of the art.	The proposed models enhance the mean AUC over the state of the art from 96.45% to 98.05% in-dataset and from 86.03% to 92.18% cross-dataset.	Source Feature Inconsistency Utilization, Novel Representation Learning and Improved Performance.	Lack of Dataset Details	The paper introduces a method for detecting deepfake images based on source feature inconsistency. It emphasizes the utilization of source features, the introduction of PCL and I2G, and improvements in AUC over the state of the art.
[32]	Bojia Zi et al. (2021)	FaceForensics++ CelebDF	2D and 3D attention	Deepfake Detection,	This research introduced	Systematic evaluation of	Real-World Dataset,	Lack of technical	The paper introduces the

		DFDC datasets consisting of 7,314 face sequences extracted from 707 deepfake videos collected entirely from the internet.	models, and ADDNets	WildDeepfake Dataset, & Attention-based Deepfake Detection Networks (ADDNets)	the WildDeepfake dataset to help create and test deepfake detectors against real-world internet deepfakes. The research also proposed ADDNets, which use attention masks to improve detection.	baseline detection networks is conducted on the existing datasets and the WildDeepfake dataset.	attention-Based Models and Availability	details about the artificial fingerprinting process and systematic evaluation.	WildDeepfake dataset, designed for evaluating deepfake detectors against real-world deepfakes from the internet. It also mentions the proposal of ADDNets, which leverage attention masks for improved detection.
[33]	Sowmen Das et al. (2021)	FaceForensics++ Celeb-DF DFDC	Model tends to memorize the actors' faces and labels instead of learning fake features.	Face-Cutout, Evaluation Experiments and General-Purpose Data Pre-processing Guideline.	This research identified deepfake detection framework flaws as its main result. It introduces the Face-Cutout data augmentation strategy to reduce overfitting and improve data variation. A general-	The paper mentions that the Face-Cutout method achieves a reduction in LogLoss of 15.2% to 35.3% on different datasets compared to other occlusion-based techniques.	Overfitting Mitigation, Data Augmentation Solution and General-Purpose Guideline.	Lacks detailed technical information about the architecture and functioning of the Face-Cutout method	It discusses the identification of limitations in existing deepfake detection frameworks and the proposal of the Face-Cutout data augmentation method to mitigate overfitting.

					purpose data pre-processing strategy is also suggested to improve deepfake detection model generalizability.				
[34]	Deressa Wodajo et al. (2021)	DFDC	CNN and Vision Transformer (ViT) components.The CNN is responsible for extracting learnable features, while the ViT takes these features as input and categorizes them using an attention mechanism.	Convolutional Vision Transformer, and Training on DFDC	The primary outcome of this research is the development of a Convolutional Vision Transformer model for detecting deepfakes. The model is evaluated on the DFDC dataset and achieves specific performance metrics, including accuracy, AUC, and loss.	The paper mentions that the model achieves 91.5% accuracy, an AUC value of 0.91, and a loss value of 0.32 when evaluated on the DFDC dataset.	Model Architecture and Competitive Results	size or composition of dataset not described properly	The paper presents a Convolutional Vision Transformer model for deepfake detection, highlighting its architecture and performance metrics on the DFDC dataset.

[35]	Felix Juefei-Xu et al. (2021)	ImageNet COCO CelebA FFHQ LSUN CelebA-HQ VGGFace2 CASIA-WebFace FaceForensics++ MS-Celeb-1M MegaFace Celeb-DF FaceForensics WildDeepFake DeeperForensics	Timeliness Scale Detailedness Technical evolution analysis Battleground analysis Horizon analysis	Comprehensive analysis of 318 research papers	The primary outcome of this research is a comprehensive overview and analysis of the landscape of DeepFake research, including generation, detection, and evasion. It aims to provide insights into the trends, challenges, and opportunities in the field.	It focuses on the survey and analysis of research papers in the DeepFake domain.	Taxonomy and Categorization along with a Comprehensive Overview.	No personal experiment or research	It describes the research's goal to cover DeepFake production, detection, and evasion. It provides an introduction to the research's scope and goals without quantitative data or technical information.
[36]	Hanqing Zhao et al. (2021)	Celeb-DF FaceForensics++	fine-grained classification and use of multiple spatial attention	Multi-attentional deepfake detection network has several spatial attention heads, a textural feature augmentation block, and an aggregation of low-level textural and	The main result of this research is a new deepfake detection algorithm that treats the problem as a detailed classification task. The method outperforms binary	Method-Efficient-B4: LQ- ACC 88.69 AUC 90.40 HQ- ACC 97.60 AUC 99.29	Fine-Grained Approach and Attention Mechanisms.	Framework is sensitive to high compression rate which blurs most of the useful information in spatial domain.	The paper outlines the research's aim to formulate deepfake detection as a fine-grained classification problem and introduces a novel multi-attentional deepfake detection network.

				high-level semantic features directed by attention maps	classification methods and is state-of-the-art.				
[37]	Jiameng Pu et al. (2021)	FaceForensics++ Creation of a large dataset of deepfake videos in the wild, consisting of 1,869 videos collected from YouTube and Bilibili, containing over 4.8 million frames of content.	collection and analysis of real-world deepfake content.	The research collects and analyses real-world deepfake footage. It reviews current defence strategies and investigates transfer learning and competition-winning ways to improve deepfake video defences.	The study seeks to explain real-world deepfake video growth, popularity, creators, modification strategies, and production processes. It assesses current defences and investigates ways to improve them against real-world deepfake content.	Promising approach to improve performance of DF-W	Large Real-World Dataset, Comprehensive Analysis	Complex approach	The study outlines the contributions of the research, including the creation of a large real-world dataset of deepfake videos, a comprehensive analysis of deepfake content, and an evaluation of existing defense methods.
[38]	Hong-Shuo Chen et al. (2021)	UADFV Celeb-DF v1 Celeb-DF v2	Automatic extraction of features using the successive subspace learning (SSL) principle from various parts of face images.	The proposed DefakeHop approach uses sequential subspace learning (SSL) for feature extraction, the c/w Saab	The primary outcome of this research is the development of the DefakeHop method for deepfake	AUC values of 100%, 94.95%, and 90.56% on UADFV, Celeb-DF v1, and Celeb-DF v2 datasets, respectively.	High Performance and small model size.	Lacks in-depth technical details about the c/w Saab transform, feature distillation module, or SSL principle.	The paper outlines the proposed DefakeHop method for deepfake detection and provides specific

			These features are processed using the c/w Saab transform and a feature distillation module.	transform for feature processing, and a feature distillation module for dimension reduction and soft classification. These methods detect deepfakes in facial photos.	detection. The method achieves state-of-the-art performance on multiple datasets, as indicated by the area under the ROC curve (AUC) values.				accuracy results, demonstrating its state-of-the-art performance on multiple datasets.
[39]	Zhi Wang et al. (2021)	FaceForensics++ (FF++), a recently released large-scale deepfake video detection dataset, contains 1,000 genuine videos: 720 for training, 140 for verification, and 140 for test. Four advanced methods—DeepFakes (DF), Face2Face (F2F), FaceSwap (FS), and NeuralTextures (NT)—generated four fake videos from each real video in the dataset. We	Generalization	Introduction of pixel-wise Gaussian blurring models to mitigate high-frequency artifacts in AI-based face manipulation.	The main result of this research is an adversarial training strategy to improve deepfake detection classification model generalisation. These models should be better at detecting hidden facial forgeries and responding to different image/video qualities.	Improved accuracy of over 90% in all the models.	Improved Generalization	Lack of Specifics	The study outlines the motivation and approach for improving the generalization ability of deepfake detection models through adversarial training.

		followed the specified training, validation, and test set split in our research. RAW, C23, and C40 video quality were processed for each dataset video. Following the official face detection and alignment method, we collected 270 frames from each of 5,000 (actual and false) videos for each quality. They included DFD and Celeb-DF deepfake datasets to broaden the evaluation. DFD comprises 3,068 deepfake videos based on 363 actual ones. Celeb-DF has 590 authentic and 5,639 fraudulent videos.							
[40]	Young-Jin Heo et al. (2021)	DFDC	Vision Transformer model with distillation methodology	It combines CNN features and a patch-based positioning	The Vision Transformer-based method for detecting	The proposed model achieves an AUC of 0.978 and an F1 score of 91.9, while the	Improved Performance due to focus on false negatives.	Lacks detailed study	The paper presents a Vision Transformer-based approach

				model to interact with all positions to identify artifact regions in the videos.	phoney videos with an emphasis on false negatives is the main result of the research. The research reports better DFDC Dataset performance than the state-of-the-art model.	previous state-of-the-art model yields an AUC of 0.972 and an F1 score of 90.6 on the same dataset.			with a distillation methodology for detecting fake videos, highlighting improved performance in terms of AUC and F1 score on the DFDC Dataset.
[41]	J. Thies et al. (2019)	Sample of 100 real and 100 fake videos for first. Second dataset had 500 real and 500 fake videos.	The authors used a combination of spatial and temporal features to detect deepfakes. The spatial features were extracted from the images or videos, and the temporal features were extracted from the changes in the images or videos over time. The spatial features included the brightness,	The authors used a temporal generative adversarial network (TGAN) to detect deepfakes. The TGAN is a type of deep learning model that can be used to generate realistic images or videos. The TGAN is trained on a dataset of real and fake	The authors were able to achieve a detection accuracy of 94.6% on a dataset of deepfake videos.	For TGAN- Generator: ResNet-50 Discriminator: PatchGAN Learning rate: 0.0002 Batch size: 64 Number of epochs: 200	It can detect deepfakes that are created using different techniques. It is robust to variations in the lighting and background of the images or videos. It is relatively fast and efficient.	It requires a large dataset of real and fake images or videos to train the TGAN. It is not yet clear how the method would perform on deepfakes that are created using more advanced techniques.	The authors evaluated their method on a dataset of deepfake videos. The dataset consisted of 100 real videos and 100 fake videos. The authors were able to achieve a detection accuracy of 94.6% on this dataset.

			contrast, and texture of the images or videos. The temporal features included the motion of the objects in the images or videos.	images or videos. The real images or videos are used to train the generator of the TGAN, and the fake images or videos are used to train the discriminator of the TGAN. The discriminator is then used to classify images or videos as real or fake.					
[42]	Sangyup Lee et al. (2021)	FaceForensics++ 200 real-world Deepfake-in-the-Wild (DW) videos of 50 celebrities	Transfer learning-based Autoencoder with Residuals (TAR).	Transfer learning-based Autoencoder with Residuals (TAR) is proposed for deepfake detection. Autoencoder-based detection model with Residual blocks is utilized. Transfer learning is performed to	The paper aims to develop a practical digital forensic tool for detecting various types of deepfakes simultaneously, achieving high accuracy with a small number of training samples. The proposed method seeks	Zero shot accuracy – 89.49% gaining 10.77% from baseline model	A unified model is developed to detect multiple types of deepfake videos simultaneously. Transfer learning helps improve detection performance across various deepfake types. Achieves higher generalized detection performance compared to	Small dataset for generalisation	The approach achieves much higher generalized detection performance than state-of-the-art methods on the FaceForensics++ dataset. Evaluation on 200 real-world Deepfake-in-the-Wild (DW) videos of celebrities results in

				detect different types of deepfakes simultaneously.	to generalize detection performance across different types of deepfakes.		state-of-the-art methods. Demonstrates practicability through real-world evaluation and validation.		89.49% zero-shot accuracy, significantly higher than the best baseline model (improvement of 10.77%).
[43]	Matthew Groh et al. (2021)	DFDC, consists of both authentic videos and deepfake videos. The authentic videos are genuine, unaltered video footage, while the deepfake videos are machine-manipulated videos created using advanced deep learning techniques.	Primarily visual and pertain to the content and quality of the videos	The primary technique implemented in this study is the comparison of human observers' ability to distinguish between authentic and deepfake videos with the performance of a leading computer vision deepfake detection model	This study aims to compare human observers and a computer vision deepfake detection technique in recognising deepfake films. The study examines detection methods' pros and cons. Also, the study examines the combined accuracy when people see the model's predictions.	Const-0.000002 Anger-0.099525	The study involves a large number of participants (15,016) from an online setting, which allows for a diverse and extensive dataset of human responses. The comparison of human detection capabilities with computer vision models provides insights into the relative strengths and weaknesses of each approach. Investigating the impact of pre-registered randomized interventions on deepfake detection adds	The paper lacks details about the composition and size of the dataset, making it difficult to assess the representativeness of the videos used. Specific features used for detection and the deepfake detection model employed are not detailed.	It suggests that human observers and computer vision models make different kinds of mistakes when detecting deepfake videos. The combination of both human and model predictions improves accuracy, but incorrect model predictions can negatively impact human accuracy. The study also hints at the importance of specialized cognitive capacities in explaining human

							an experimental dimension to the study.		deepfake detection performance.
[44]	Bo Peng et al. (2021)	DFGC 2021	Facial landmarks, lip movements, & audio-visual synchronization	Participants' deepfake detection methods were likely implemented in the competition. Modern computer vision models, machine learning algorithms, and other methodologies may be used to identify deepfake films.	The DFGC 2021 competition benchmarks cutting-edge deepfake development and detection algorithms. The tournament revealed the continuous conflict between deepfake creators and detectors and highlighted their advances. The competition's outcomes would evaluate deepfake detection and technology.	Best Score – 0.94	The competition helps test deepfake development and detection technologies. The publication advances deepfake technology and detection by sharing insights and findings. The DFGC-21 testing dataset expands deepfake detection research.	Things in the top solutions also are redundant	Highlights the importance of benchmarking the adversarial game between deepfake creators and detectors and suggests that advancements have occurred in both areas.
[45]	Vishal Asnani et al. (2021)	100,000 fake images generated by 116 different Generative Models (GMs), CIFAR	image content, structure, and visual characteristics	A Fingerprint Estimation Network (FEN) and Parsing Network (PN) comprise the	A model parsing system that reverses engineers GMs to infer	Binary classification performance for coordinated misinformation attack-	The research addresses the challenging problem of reverse engineering	The application of fingerprint estimation for deepfake detection and	It highlights that the research achieved encouraging results in

				framework. The constraint-trained FEN estimates GM fingerprints from generated images. However, the PN predicts network designs and loss functions using estimated fingerprints. The framework uses these components for "model parsing." The dataset of phoney GM photos is also used.	hyperparameters from generated images is the main result of this research. The suggested method estimates GM network designs and training loss functions from model photos. The calculated fingerprints can also recognise deepfakes and attribute images, according to the study.	Method: FEN- AUC (%) = 83.5 Classification accuracy (%) = 76.85 FEN + PN – AUC (%) = 87.3 Classification accuracy (%) = 80.6	GMs to understand their hyperparameters, which can be valuable for identifying manipulated media. The proposed framework demonstrates promising results in estimating GM network architectures and loss functions from generated images.	image attribution can contribute to mitigating the misuse of GMs.	parsing the hyperparameters of unseen GMs and reported state-of-the-art results in deepfake detection and image attribution benchmarks
[46]	Gaojian Wang et al. (2021)	6 large-scale DeepFake datasets	Fused Facial Region_Feature Descriptor	The research introduces the FFR_FD as a novel approach for DeepFake detection. It's mentioned that this method is efficient and fast, making it	The primary outcome of this research is the development of the FFR_FD-based DeepFake detection method. The proposed	AUC Scores for datasets- SIFT:96.9 SURF:97.1 FAST&BRIEF:99.9 ORB:98.6 A-KAZE:97.8	The research addresses the challenge of DeepFake detection by proposing a feature-based approach that relies on facial feature points and descriptors.	Lacks specific details about the datasets used for experimentation, evaluation metrics, and the extent of performance improvement achieved compared to	While the paper indicates that the proposed FFR_FD-based method outperforms most state-of-the-art DNN-based models, it doesn't provide specific

				suitable for real-world applications where computational resources may be limited.	approach aims to leverage facial feature points and descriptors to extract relevant information from face images. The research suggests that this method outperforms most state-of-the-art DNN-based models in DeepFake detection.		The FFR_FD method is designed to be efficient and fast, making it suitable for real-time or resource-constrained applications.	DNN-based models.	quantitative results.
[47]	Hemlata tak et al. (2021)	ASVspoof 2019	raw waveform inputs	spectro-temporal graph attention network (GAT)	The main outcome of this research is the development of the RawGAT-ST model, which aims to achieve reliable detection of spoofed or deepfake speech by automatically learning	error rate of 1.06%	The research focuses on using raw waveform inputs, which can capture fine-grained audio details without relying on predefined features. The RawGAT-ST model is designed to learn relationships between cues in different sub-	While the paper highlights the model's performance on a specific dataset, it doesn't provide insights into its performance across a spectrum of diverse spoofing attacks or on	The research reports an equal error rate of 1.06% for the ASVspoof 2019 logical access database, which is considered one of the best results to date.

					representations from raw waveform inputs. The model's performance is evaluated using the ASVspoof 2019 logical access database.		bands and temporal intervals, which can improve the detection of diverse spoofing attacks. The reported equal error rate of 1.06% for the ASVspoof 2019 logical access database suggests that the model achieves excellent detection performance.	other datasets.	
[48]	Jiajun Huang et al. (2021)	DeepFake MNIST+ consisting of 10,000 facial animation videos that cover ten different facial actions	Facial features	The research proposes a baseline detection method for facial animation detection.	The primary outcome of this research is the introduction of the DeepFake MNIST+ dataset, which is designed to facilitate the development of reliable detection methods for facial animation, a facet of	ResNet50-Raw: 93.57% Light Comp.:90.69% Heavy Comp.:85.56% ResNet152-Raw:95.78% Light Comp.:92.11% Heavy Comp.:88.32%	The research addresses the importance of detecting facial animations as part of the DeepFake threat landscape, especially in the context of liveness detection systems for user authentication. The introduction of the DeepFake MNIST+ dataset, containing 10,000 facial	While the research identifies the importance of facial animation detection, it doesn't provide insights into the baseline method's effectiveness or its performance on the proposed dataset.	The paper provides information about the dataset's creation and the intention to develop a baseline detection method.

					DeepFake attacks that has received less attention in recent research.		animation videos with ten different actions, contributes to the development of more robust and reliable detection methods.		
[49]	Minh Tam Pham et al. (2021)	FaceForensics++ Celeb-DF FaceForensics WildDeepFake	state-of-the-art counterfeit generators and detectors	The research introduces an independent benchmarking framework that assesses the performance of counterfeit generators and detectors. It appears to evaluate the effectiveness of these methods using various criteria, providing a comprehensive assessment of visual forgery and visual forensics techniques.	The primary outcome of this research is the development of a benchmarking framework for visual forgery and visual forensics. This framework aims to provide a comprehensive and empirical approach to assessing the performance of counterfeit generators and detectors, shedding light on the	Increased accuracy and efficient model	The research addresses the critical issue of visual forgery and its potential malicious applications, highlighting the importance of visual forensic techniques in maintaining information security. The development of a benchmarking framework offers a systematic and empirical approach to assess the performance of various counterfeit generators and detectors.	Paper lacks specific details about the benchmarking criteria, counterfeit generation techniques, detection methods, and their respective outcomes	Introduces the concept of a benchmarking framework for visual forgery and visual forensics

					ongoing battle between measures and countermeasures in the field of information security.				
[50]	Shivangi Aneja et al. (2021)	FFHQ	Generating image-specific perturbations	It involves generating image-specific perturbations that disrupt the manipulation process.	The primary outcome of this research is the development of a data-driven approach for protecting images from face manipulation. By embedding image-specific perturbations, the method aims to disrupt the manipulation process and make it challenging for manipulation models to produce	Enhanced performance	The research addresses the significant issue of privacy and disinformation associated with face manipulation methods. The proposed data-driven approach offers a novel method for protecting images from face manipulation, making it difficult for manipulation models to generate accurate results. It emphasizes the efficiency of the approach, allowing for integration in	Comprehensive study is required.	This approach can enhance privacy and combat the misuse of face manipulation techniques.

					accurate results.		image processing pipelines, even on resource-constrained devices.		
[51]	Sitong Liu et al. (2022)	FaceForensics++	CNN	Block shuffling regularization, Adversarial loss algorithm & Restoration of spatial layout	The primary outcome of this research is the development of a deepfake detection method that addresses the overfitting challenge. The proposed method aims to improve the generalization capabilities of deepfake detection models, making them robust against cross-dataset evaluations and common image transformations.	Generalised results	The research addresses a significant challenge in deepfake detection, which is overfitting to known forgery methods and common image transformations. The proposed method introduces innovative techniques like block shuffling regularization and an adversarial loss algorithm to enhance model robustness. The research emphasizes the generalization capabilities of the method, demonstrating its potential to work well on various datasets	Doesn't provide specific findings or results	The paper introduces the concept of a novel deepfake detection method designed to combat overfitting issues faced by CNN-based models.

							and resist common image transformations.		
[52]	Binh M. Lee et al. (2023)	DeepFake	Hilbert-Schmidt Independence Criterion.	QAD (Quality-Agnostic Deepfake Detection)	The primary outcome of this research is the development of a deepfake detection method, QAD, which aims to effectively and simultaneously detect deepfakes of different quality levels. The research suggests that QAD is a quality-agnostic deepfake detection approach.	Demonstrates the superiority of the QAD	High Performance, aims to be quality-agnostic, which suggests versatility in detecting deepfakes irrespective of their quality.	Lacks details	The paper highlights the development of a quality-agnostic deepfake detection method called QAD and suggests its superiority over previous benchmarks
[53]	Yuankun Xie et al. (2023)	Chinese Fake Song Detection (FSD) dataset	focuses on the construction of the FSD dataset	The research begins by constructing the FSD dataset, and it mentions that initial experiments revealed the ineffectiveness	The primary outcome of this research is the development and evaluation of specialized ADD models for detecting	song-trained ADD models exhibit a 38.58% reduction in average equal error rate compared to speech-trained ADD models on the FSD test set.	The research addresses the emerging issue of deepfake songs, which is becoming increasingly relevant with advancements in singing voice	Further details about the methodologies used for constructing the FSD dataset and training the ADD models would	The research highlights the construction of the Chinese Fake Song Detection (FSD) dataset and the development of song-trained ADD models for

				of existing Audio DeepFake Detection (ADD) models for song deepfake detection. To address this, the FSD dataset is employed to train ADD models specifically for song deepfake detection.	deepfake songs, particularly in the context of the Chinese FSD dataset.		synthesis and conversion technologies. The creation of the Chinese FSD dataset provides a valuable resource for further research in song deepfake detection. The evaluation results suggest that training ADD models on song-specific data can significantly improve detection accuracy for deepfake songs.	enhance the understanding of the research.	deepfake song detection. It suggests that these models outperform speech-trained models on the FSD test set.
[54]	Haixu Song et al. (2023)	DeepFakeFace (DFF)	creation of the DFF dataset and the evaluation of deepfake recognition tools.	The research involves the crafting of the DFF dataset using advanced diffusion models	The primary outcome of this research is the creation and sharing of the DFF dataset, which is intended to serve as a robust foundation for training and testing deepfake detection	Enhanced performance	The research addresses the pressing issue of deepfake images and their potential impact on the dissemination of authentic information. The creation and sharing of the DFF dataset contribute to the research community's	While the DFF dataset is introduced, the paper lacks specific details about its size, diversity, and characteristics.	The paper highlights the creation of the DeepFakeFace (DFF) dataset and suggests evaluation methods for deepfake recognition tools.

					algorithms. The research aims to boost the development of more effective tools against deepfake images.		resources for developing and evaluating deepfake detection algorithms. The proposed evaluation methods assess the strength and adaptability of deepfake recognition tools, which is important for improving their effectiveness.		
[55]	Boquan Li et al. (2023)	of six deepfake datasets FaceForensics++	generalizability of the models across different datasets.	The research utilizes five deepfake detection methods and two model augmentation approaches. The augmentation approaches aim to improve the generalizability of deepfake detection models.	The research aims to examine the generalizability of deepfake detection models, which is essential for these models to stay effective in detecting emerging deepfake techniques. It emphasizes the need to address the limitations of	Improved auc scores for various models	The research addresses a critical challenge in the field of deepfake detection, which is the generalizability of detection models to previously unseen deepfake datasets. It utilizes multiple deepfake datasets and detection methods, providing a	While the research identifies issues related to generalizability, it doesn't provide concrete solutions or findings regarding how to improve zero-shot generalization.	The paper presents an empirical study on the generalizability of deepfake detection models, highlighting their limitations in zero-shot settings and the challenges they face in extracting discriminative features.

					existing detectors in zero-shot settings.		comprehensive examination of the generalizability issue. The identification of neurons contributing to detection across datasets suggests a potential path forward for achieving zero-shot generalizability.		
[56]	Deressa Wodajo et al. (2023)	DFDC, FF++, DeepfakeTIMIT, and Celeb-DF v2	GenConViT model	The proposed GenConViT model leverages Autoencoder and Variational Autoencoder techniques to learn from the latent data distribution. These techniques likely play a role in capturing visual artifacts and patterns indicative of deepfake	The research aims to develop an effective deepfake video detection model called GenConViT. The model is designed to detect a wide range of deepfake videos by learning from visual artifacts and latent data distribution, ultimately	average accuracy of 95.8% and an AUC value of 99.3%	The research addresses the significant concerns related to deepfake videos and their potential to spread false information. By combining ConvNeXt and Swin Transformer models and utilizing Autoencoder and Variational Autoencoder techniques, the GenConViT model offers a	lacks specific details about the architecture and training process of the GenConViT model, making it challenging to assess the model's technical aspects.	The paper highlights the successful development of the GenConViT model for deepfake video detection, with a focus on its performance and generalizability across different datasets

				videos. Additionally, ConvNeXt and Swin Transformer models are used for feature extraction.	preserving media integrity.		comprehensive approach to deepfake video detection. The model's robust performance across multiple datasets suggests its effectiveness in identifying a wide range of fake videos.		
[57]	Yingxin Lai et al. (2023)	FaceForensics involves fine-grained pixel-wise supervision labels	Segment Anything Model (SAM)	Detect Any Deepfakes (DADF), including the Multiscale Adapter and the Reconstructio n Guided Attention (RGA) module	The research aims to improve deepfake detection and localization by introducing the DADF framework based on the SAM model. The framework seamlessly integrates forgery detection and localization, potentially enhancing the precision and effectiveness	Increase of 3.21% and making it 96.64% on localisation (multiscale) and detection accuracy of 95.94%	The research addresses the challenge of fine-grained pixel-wise supervision labels in deepfake detection and localization, which is crucial for precise forgery detection. The proposed DADF framework leverages the SAM model and introduces additional components to capture forgery contexts and	Specific results and comparative analyses with existing methods are not presented in the paper, making it difficult to gauge the magnitude of improvement achieved.	Highlights the introduction of the DADF framework and its potential advantages for deepfake detection and localization

					of these tasks.		enhance sensitivity, potentially leading to improved performance.		
[58]	Piotr Kawa et al. (2023)	ASVspoof 2021 and DF In-The-Wild.	Whisper automatic speech recognition model as a front-end for DF detection	The research compares different combinations of Whisper-based features with well-established front-ends for DF detection. Three detection models, namely LCNN, SpecRNet, and MesoNet, are trained and evaluated using these features. The goal is to assess the impact of Whisper-based features on DF detection performance.	The research aims to evaluate the effectiveness of using Whisper-based features as a front-end for DF detection. The results are expected to reveal whether Whisper-based features improve detection performance compared to traditional front-ends.	Whisper-based features leads to improved DF detection for each of the three models (LCNN, SpecRNet, and MesoNet). It also states that this approach reduces the Equal Error Rate (EER) by 21% on the DF In-The-Wild dataset compared to recent results.	The research addresses the growing threat of audio DeepFakes (DF) and focuses on improving detection methods. By investigating the impact of Whisper-based features, the research aims to contribute to more effective DF detection. The reduction in Equal Error Rate (EER) by 21% on the DF In-The-Wild dataset suggests that using Whisper-based features is a promising approach for improving DF detection.	Highlights the improvement in EER, it doesn't provide specific values or comparative metrics, making it challenging to assess the significance of the improvement.	The research appears to demonstrate the effectiveness of using Whisper-based features for DF detection by achieving a 21% reduction in EER on the DF In-The-Wild dataset.
[59]	Zhixi Cai et al. (2023)	Localized Audio Visual DeepFake (LAV-DF)	spatio-temporal changes in facial attributes and	Boundary Aware Temporal	The research focuses on improving	quantitative analysis demonstrating the	The research addresses a critical gap in	While it mentions the "superiority"	The research focuses on improving

			manipulations involving audio and audio-visual elements	Forgery Detection (BA-TFD), CNN	temporal forgery localization and deepfake detection tasks, acknowledging the presence of manipulations involving audio, visual, and audio-visual elements. The outcome of the study is expected to demonstrate the effectiveness of BA-TFD+ on these tasks using several benchmark datasets, including the newly proposed LAV-DF dataset.	superiority of BA-TFD+ in terms of temporal forgery localization and deepfake detection tasks.	deepfake detection by considering manipulations involving audio, visual, and audio-visual elements, which are often overlooked in existing methods. The creation of the LAV-DF dataset and the proposed BA-TFD and BA-TFD+ methods provide valuable resources and tools for improving deepfake detection, especially in scenarios where audio-visual aspects play a crucial role. Availability of dataset, models, and code on GitHub promotes transparency and reproducibility.	of BA-TFD+ in deepfake detection, the paper does not quantify this improvement or provide concrete results.	temporal forgery localization and deepfake detection tasks by considering multimodal manipulations involving audio, visual, and audio-visual elements.

[60]	Ying Xu et al. (2023)	KoDF FaceForensics++ Celeb-DF	Multi-Channel Xception Attention Pairwise Interaction (MCX-API)	MCX-API is described as an approach that leverages pairwise learning and complementary information from different color space representations in a fine-grained manner.	The paper indicates that the experiments conducted with MCX-API demonstrate its potential to generalize better than state-of-the-art Deepfake detectors. It reports balanced-open-set-classification (BOSC) accuracy results on the FF++ and CelebDF datasets, suggesting that the proposed method performs well in an open-set scenario.	Using the proposed MCX-API method, including a BOSC accuracy of 98.48% on the FF++ dataset and 90.87% on the CelebDF dataset is achieved.	The proposed MCX-API approach aims to address the challenge of generalization in Deepfake detection, particularly in handling diverse Deepfake generation schemes. By reporting BOSC accuracy, the paper highlights the method's performance in an open-set scenario, where it needs to handle unseen attacks. The availability of the GitHub repository provides access to the code and resources associated with the proposed approach, promoting transparency and reproducibility.	High-level results and mentions the approach's generalization capabilities, it lacks specific technical details about MCX-API.	The research focuses on developing a Deepfake detection approach called MCX-API, with an emphasis on generalization across diverse Deepfake generation schemes.