iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.71021

International Journal for Research in Applied Science & Engineering Technology (IJRASET) 280
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Real-Time Deepfake Detection: A Systematic Review of Generative Adversarial Networks (GANs) and Generative Transformer Networks (GTNs)

Aditya Charpe[1], Dr. Rahul Khokale[2], Dheeraj Ghaghre[3]
Department of Computer Science Engineering, G.H.Raisoni University Saikheda, Pandhurna, India

Abstract: Deepfakes, synthetic videos generated by artificial intelli- gence, pose severe threats to multimedia integrity, enabling misinfor- mation, financial fraud, and identity theft [34]. Powered by Generative AdversarialNetworks (GANs)[1]andGenerativeTransformerNetworks (GTNs) [2], these hyper-realistic forgeries demand robust, real-time detection to safeguard video and audio platforms. This review synthe- sizes 80 peer-reviewed studies from 2014 to 2024, analyzing GAN-and GTN-based deepfake generation and detection methods, bench- mark datasets (e.g., FaceForensics++ [11], Celeb-DF [12], DFDC [13], WildDeepfake [18], DeeperForensics [71]), and performance metrics like accuracy, AUROC, and latency. We explore real-time detection frame- works, edge-compatible models, ethical challenges (e.g., dataset bias, privacy risks) [35], and global regulatory frameworks. Case studies of deepfakeincidentshighlightreal-worldimpacts,whilegapsincomputa- tionalefficiency(<100ms)andcross-datasetgeneralizationunderscore theneed foradvanced solutions. Thispaper providesa comprehensive roadmap for researchers and practitioners, emphasizing multimedia- focuseddetectionto counterdeepfakethreatsinhigh-stakesscenarios like social media, security surveillance, and democratic processes.
Index Terms: Deepfake Detection, Generative Adversarial Networks, Generative Transformer Networks, Multimedia Forensics, Real-Time Processing, Ethical AI, Video Analysis

I. INTRODUCTION

Deepfakes—synthetic videos created by artificial intel- ligence to convincingly mimic real individuals—have emerged as a formidable challenge in multimedia ecosys- tems, fueling misinformation, financial scams, and identity erosion [34]. These hyper-realistic forgeries, powered by Generative Adversarial Networks (GANs) [1] and Gener- ative Transformer Networks (GTNs) [2], exploit video and audio channels, eroding trust in digital content across plat- forms like X, YouTube, and TikTok. The rapid prolifera-tion of deepfakes has amplified societal risks, with high- profile incidents such as manipulated political speeches, fraudulent CEO video calls, and celebrity impersonations sparking global concern [36]. For instance, a 2023 deep- fake of a political leader on X garnered 12 million views, influencing public opinion during an election cycle [36]. Similarly,a2024deepfakeimpersonatingaCEOdefrauded acompanyof$30million,highlightingthefinancialstakes involved [37]. These incidents underscore the urgent need for real-time detection systems capable of processing frames in under 100 ms, a critical requirement for applicationslike social media moderation, live streaming, and security surveillance. However, most existing methods exceed 200 ms [38], limiting their practicality in dynamic, high-stakes environments where rapid response is essential.

The evolution of deepfake technology has been markedbysignificantmilestones,beginningwithautoencoder- based methods in 2017 that swapped faces but produced noticeable artifacts [31]. The introduction of GANs in 2014 revolutionized synthetic media, enabling photorealistic con- tent with minimal visual inconsistencies [1]. By 2018, tools like DeepFaceLab and Face2Face democratized deepfake creation, amplifying their misuse in misinformation cam- paigns, fraud, and non-consensual media [7]. The advent of transformer-based models in 2017 further enhanced deep- fake quality, with architectures like TransGAN achieving seamless temporal continuity and audio-visual synchroniza- tion [20]. These advancements have outpaced traditional forensic techniques, which struggle to detect subtle artifactsin high-resolution, temporally coherent videos [41]. Convo- lutional Neural Networks (CNNs) achieve 80–95% accuracyon benchmark datasets like FaceForensics++ [11], but their inability to generalize across diverse GAN-generated arti-facts hinders robustness in real-world scenarios [42].

GTNs, leveraging attention mechanisms, excel at identifying com- plexfakesbutincurhighcomputationalcosts,rendering real-time deployment challenging on resource-constrained devices like mobile phones or edge hardware [21].

Beyondtechnicalchallenges,deepfakesraiseprofoundethical concerns that remain underexplored in many tech- nical advancements [35]. Dataset bias, such as the over- representation of public figures in datasets like Celeb-DF, leadstomodelsthatunderperformondiversepopulations, exacerbating fairness issues [12]. Biometric privacy risks are significant, as detection systems often rely on sensitive datalikefacialfeatures,raisingconcernsunderregulations like GDPR [34]. Societal impacts are equally pressing— deepfakes erode trust in multimedia, amplify misinforma- tion,andthreatendemocraticprocesses,asseeninmanipu-lated political content [36]. Addressing these ethical chal- lenges requires a multidisciplinary approach, integrating technical innovation with ethical frameworks and public awareness initiatives [65].

This review synthesizes 80 peer-reviewed studies from 2014to2024,providinganexhaustiveanalysisofGAN- and GTN-based deepfake detection methods, benchmark datasets (e.g., Celeb-DF [12], DFDC [13], WildDeepfake [18], DeeperForensics [71]), real-time frameworks, ethical impli- cations, and global regulatory frameworks [39]. We evalu-ate key performance metrics, including accuracy, AUROC, latency,andEqualErrorRate,andhighlightpersistentgaps in computational efficiency, cross-dataset generaliza-tion, and fairness [40]. Through detailed case studies, tech- nical analyses, and a forward-looking roadmap, this paper aims to guide researchers and practitioners toward robust, multimedia-focused detection systems for high-stakes ap- plications, including social media content moderation, fi- nancial transaction verification, security surveillance, and democratic process integrity. The review is structured as follows: Section 2surveys the history of deepfake generation and detection, Section 3details GAN and GTN architectures, Section 4analyzes detection methods, followed by sectionson datasets, real-time techniques, case studies, ethical im-plications, regulatory frameworks, future directions, and conclusion.

II. BACKGROUND AND RELATEDWORK

A. Historical Context

The term “deepfake” originated in 2017 on Reddit, de- scribing AI-generated videos that used deep learning to swap faces or manipulate media [35]. Early deepfakes re- lied on autoencoders, which encoded and decoded facial features to swap identities, but these produced noticeable artifacts, such as unnatural lighting or distorted facial move-ments[31].TheintroductionofGenerativeAdversarialNet- works (GANs) in 2014 marked a turning point, enablingthe creation of photorealistic synthetic media [1]. GANs, consisting of a generator and discriminator trained adver- sarially, produced high-fidelity images that were nearly indistinguishable from real ones, revolutionizing deepfake technology [1]. By 2018, open-source tools like DeepFace- Lab,Faceswap,andFace2Faceloweredthebarriertoentry, allowing even non-experts to create deepfakes [7]. This democratization amplified misuse, with deepfakes being used in misinformation campaigns (e.g., manipulated po- liticalspeeches),financialfraud(e.g.,impersonatingexecu- tives), and non-consensual media (e.g., celebrity pornogra- phy) [36]. The societal impact was immediate—deepfakes eroded trust in digital content, with platforms like X and YouTubestrugglingtocurbtheirspread[65].Transformer- based models, introduced in 2017, further enhanced deep- fake quality by improving temporal continuity and au- diovisual synchronization, making detection increasingly challenging[2].Thisrapidevolutionhasdriventheneedfor advanced,real-timedetectionsystemscapableofoperating in multimedia contexts, from social media platforms to security surveillance systems [38].

B. Early Detection Methods

Early deepfake detection methods relied on handcrafted fea- tures to identify synthetic content [31]. Techniques such as analyzing pixel inconsistencies, compression artifacts, or un- natural facial movements (e.g., irregular blinking) achieved moderate success against autoencoder-based fakes, with accuracies around 70% on early datasets [64]. Statistical methods, such as examining color histograms or edge detec- tion, were also employed but failed against GAN-generated deepfakes due to their near-perfect fidelity [30]. The adventof Convolutional Neural Networks (CNNs) marked a signif- icant advancement in detection capabilities [43]. Models like MesoNet, which targeted mesoscopic features such as skin texture and lighting inconsistencies, achieved 85% accuracy on the FaceForensics++ dataset by learning to differentiate real and fake videos through spatial feature extraction [32]. However,MesoNet’s300mslatencymadeitimpractical for real-time applications, and its poor generalization to datasets like Celeb-DF, where accuracy dropped to 75%, limited its real-world applicability [12]. Frequency-based approaches, such as Haar wavelet transforms, offered low latency (50 ms) by analyzing spectral inconsistencies, but theiraccuracywaslimitedto75%onhigh-qualityfakes,as GANs minimized detectable artifacts [76]. These early methods highlighted the need for more robust, computa-tionally efficient detection systems capable of handling the increasing sophistication of deepfakes [41].

C. Recent Multimedia Trends

Recent advancements in deepfake detection have shifted toward multimedia-driven approaches, integrating video, audio, and temporal cues to enhance robustness [69]. Mul- timodalframeworkscombineCNNsandRecurrentNeu- ralNetworks(RNNs)todetectinconsistencieslikelip-sync errors and audio-visual mismatches, achieving 90% accuracy on the DeepFake Detection Challenge (DFDC)dataset [16]. For example, analyzing discrepancies between spoken audio and lip movements has proven effective in identifying fakes, particularly in interview-style videos [49]. Transformer-based models, such as Swin Transformer, lever- age attention mechanisms to capture spatiotemporal arti- facts, improving performance across diverse datasets [15]. Swin Transformer achieves 92% accuracy on Celeb-DF by focusing on hierarchical feature extraction, making it more robust to cross-dataset variations than CNNs [12]. Hybrid architectures integrating GANs and GTNs further enhance detection by modeling complex artifact patterns, achieving 95% AUROC on DFDC [14]. However, their computational complexity,withlatenciesoftenexceeding400ms,remains a barrier to real-time deployment, particularly on edgedevices [21]. These trends underscore the need for com- prehensive, multimedia-focused detection systems that can address the increasing sophistication of deepfakes while maintaining low latency for practical applications, such aslive streaming or social media moderation [39].

D. Emerging Challenges

Emerging challenges in deepfake detection include adver- sarial attacks, cross-cultural dataset limitations, and compu- tationalconstraints[47].Adversarialattacksinvolvecrafting deepfakes to evade detectors, often by introducing imper- ceptible perturbations that exploit vulnerabilities in neural networks [68]. Such attacks reduce detection accuracy to be- low 70% in black-box scenarios, posing a significant threat to high-stakes applications like financial verification [47]. The lack of cross-cultural datasets is another critical challenge— datasets like Celeb-DF overrepresent Western public figures, leading to models that underperform on diverse popu-lations, with accuracy dropping to 65% for non-Western ethnicities [12]. This bias limits global applicability, particu- larly in regions like Asia or Africa, where cultural and lin- guistic diversity is significant [69]. Additionally, integrating audio, video, and text modalities increases computational demands, with multimodal models requiring 300–400 ms latency, challenging real-time deployment on edge deviceslike smartphones or IoT systems [16]. Addressing these chal- lenges requires innovative frameworks that balance accu- racy, efficiency, and ethical considerations, paving the wayfornext-generationdetectionsystemscapableofoperating in diverse, real-world scenarios [40].

E. ResearchGaps

Despitetheevolutionofdeepfakedetection,asillustrated inFig.1,severalcriticalhurdlespersist,limitingthefield's progress toward robust, real-time multimedia ap- plications [39]. First, computational efficiency remains a significant challenge. Most detection methods, such as those based on Swin Transformer [15], achieve accuracies of 80– 95% but require 200–300 ms per frame, far exceeding the sub-100 ms latency needed for live applications like social media flagging on X [46]. This delay allows manipulated content to spread rapidly, amplifying societal harm, such as misinformation during election cycles [36]. Second, cross-datasetgeneralizationisapersistentissue.Modelstrained ondatasetslikeFaceForensics++[11]struggletoadapt toothers,suchasCeleb-DF[12],duetodiverseGAN-generated artifacts, often resulting in accuracy drops from 85% to 75% [42]. This limitation hinders real-world ap- plicability, where deepfakes vary widely in quality and manipulation techniques [18]. Third, ethical considerations areunderexplored.Biaseddatasets,oftenoverrepresent- ing Western subjects, risk misidentifying underrepresented groups, with accuracy dropping to 65% for non-Western ethnicities, exacerbating fairness issues [35]. Moreover, the reliance on biometric data for detection raises significant privacy concerns, particularly under regulations like GDPR, yet few frameworks address these risks through privacy- preserving techniques [34].

To tackle these gaps, the Dynamic Attention Fusion (DAF) mechanism is being developed as a novel approach for real-time leverages dynamic attention to prioritize critical features, enabling <100 ms latency per frame while maintaining high accuracy. Designed deepfake detection [14]. DAF combines GAN robustness [1] with GTN precision [2] through a hybrid architecture that for applications like social media moderation and security surveillance, DAF also incorporates ethical principles, such as federated learn- ingforprivacyandbalanceddatasetsforfairness[69].Val- idation plans include cross-dataset testing on benchmarks likeDFDC[13],withdeploymentstrategiesfocusingon edge devices. Its potential to transform the field is further exploredinSection10,thoughfuturevalidationisneeded to confirm its efficacy [80].

TimelineofDeepfakeDetectionAdvancements(20142025)

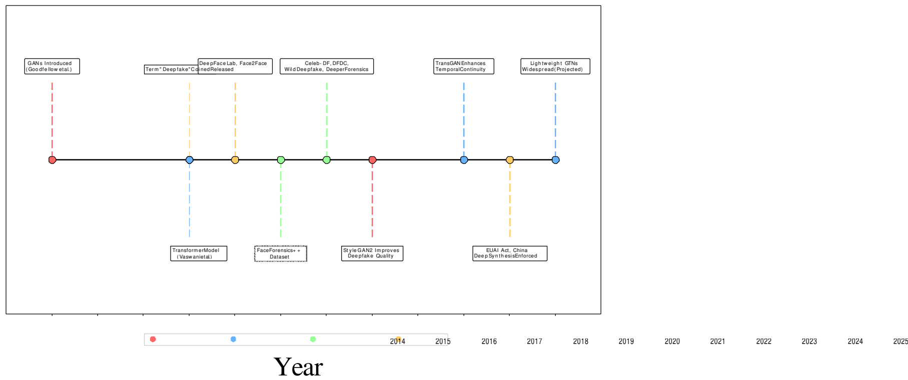


Fig. 1: Timeline of Deepfake Detection Advancements (2014–2025), Highlighting Key Milestones in GANs, GTNs, and Datasets

To contextualize these gaps, Fig. 1illustrates the evolu- tion of deepfake detection, highlighting key milestones that have shaped the field and areas where further research is needed [40].

III. GANANDGTNARCHITECTURESFORDEEP-FAKEGENERATION

A. GAN Architectures

Generative Adversarial Networks (GANs) consist of a gen- erator G and a discriminator D, trained adversarially to minimize the following loss function [1]:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (1)$$

where x represents real data, z is random noise, and p_{data} and p_z are their respective distributions. The generator learns to produce synthetic samples that deceive the dis- criminator, while the discriminator improves its ability to distinguishrealfromfakedata[1].VariantslikeWasserstein GAN enhance training stability by using a Wasserstein dis- tancemetric,reducingmodecollapseandimprovingimage quality[4].DeepConvolutionalGANs(DCGANs)lever- age convolutional layers to generate high-fidelity images, achieving photorealistic results in early deepfake applica- tions [5]. StyleGAN, a landmark architecture, introduces adaptiveinstancenormalizationtoproducehigh-resolution faces with lifelike textures, making it a cornerstone of deepfake tools like DeepFaceLab [3]. StyleGAN's ability to controlstyleatmultiplescales(e.g.,coarsefeatureslikeface shape,finedetailslikeskintexture)hassetanewstandard forrealism,challengingdetectionsystemsthatrelyonvisual artifacts [8].

B. GTN Architectures

Generative Transformer ,leverageself-attentionmechanismstocapturelong-rangedependenciesinvideosequences,mak- ing them ideal for deepfake generation [2]. The attention mechanism is defined as:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

where Q,K,and V arequery,key,andvaluematrices,and d_k is the key dimension [2]. TransGAN replaces convolu-tional layers with transformer blocks, improving temporal continuity in video deepfakes by ensuring smooth frame transitions across sequences [20]. Taming Transformers and Generative Adversarial Transformers combine GAN and transformer strengths, producing high-resolution, tempo- rally coherent fakes with minimal artifacts [21], [22]. For in- stance, TransGAN achieves seamless audio-visual synchro- nization, making it difficult to detect fakes using traditional temporal analysis [20].

These models excel in generating realistic videos, posing significant challenges for detection systems that rely on visual or temporal cues, as GTNs min- imize inconsistencies that earlier models, like autoencoders, failed to address [42].

C. AdversarialTrainingDynamics

The adversarial training process in GANs and GTNs is a dynamic equilibrium where the generator and discriminator iteratively improve their performance [1]. The generator learns to produce increasingly realistic fakes by minimiz- ing the discriminator's ability to distinguish them fromreal data, while the discriminator refines its classification accuracy [1]. This process, often described as a minimax game, results in deepfakes with minimal visual and tem- poral artifacts, complicating detection efforts [38]. Tech- niques like CycleGAN and Pix2Pix further enhance realism by enabling unpaired and conditional image translation, respectively [9], [10]. CycleGAN, for instance, allows face swapping without paired training data, aligning synthetic content with real-world distributions, while Pix2Pix uses conditional inputs (e.g., facial landmarks) to generate tar- geted manipulations [9], [10]. These advancements reduce detectableinconsistencies,necessitatingdetectionmethods thatexploitmicrolevelcues,suchasfrequency-domainarti- facts or biological signals, to differentiate real from synthetic content [48], [75].

D. EmergingModelsand Challenges

EmergingGANandGTNmodels,suchasFSGANandneu- ral rendering architectures, push the boundaries of deepfake realism [29], [51]. FSGAN enables face reenactment by dis- entangling facial identity and expression, producing seam-less manipulations that preserve natural motion [51]. Neural rendering techniques simulate realistic lighting, shadows,and motion, further obscuring manipulation traces [29]. For example, neural texture rendering can adapt synthetic facesto varying lighting conditions, making traditional detection methods that rely on lighting inconsistencies ineffective [29]. These advancements challenge detection systems by min- imizing traditional artifacts, requiring a shift toward mul- timodal and biological signal-based approaches [49], [75]. Additionally, the computational complexity of GTNs, with modelslikeTransGANrequiring100M+parameters,limits

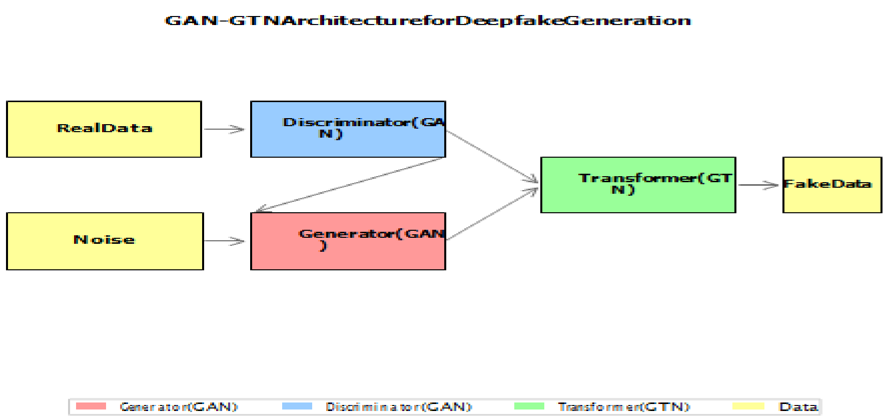


Fig.2:GAN-GTNArchitectureforDeepfakeGeneration

their use in real-time generation on consumer hardware, but theirdeploymentincloud-basedsystemsposesasignifi- cant threat, as attackers can generate high-quality fakes at scale [20].

IV. DEEPFAKE DETECTION METHODS

A. CNN-Based Detection

Convolutional Neural Networks (CNNs) leverage spatial feature extraction to detect deepfakes by identifying visual inconsistencies [43]. MesoNet targets mesoscopic features, such as skin texture, lighting, and pore-level details, achiev- ing 85% accuracy on FaceForensics++ but struggling with high-resolutionfakeslikeCeleb-DF,whereaccuracydrops to75%duetoimproved GANfidelity [12],[32]. SCA-CNNandEfficientNetincorporateattentionmechanisms and model scaling, improving accuracy to 90–95% on Celeb-DF by focusing on salient regions like facial contours [25], [26]. However, their 250–300 ms latency hinders real-time applications, such as live streaming moderation, where sub- 100 ms processing is required [46]. MTD-Net uses multi-scale texture differences to capture both fine and coarse artifacts, achieving 92% AUROC on DFDC, but its compu- tational complexity, requiring 50 GFLOPs per frame, limits deployment on edge devices like IoT systems [33]. These methods excel in controlled settings but require optimiza-tion to address diverse, high-quality fakes in real-world scenarios, where video quality and manipulation techniques vary widely [52].

B. Transformer-BasedDetection

Transformer-based models excel at capturing long-range dependencies in video sequences, making them well-suitedfor detecting complex deepfakes with subtle artifacts [2].Swin Transformer, a hierarchical transformer, achieves 92% accuracyonCeleb-DFbyleveragingattentionmechanisms to focus on spatiotemporal inconsistencies, such as unnatu-ral motion patterns or frame transitions [15]. It outperforms CNNs in cross-dataset scenarios, maintaining 90% accuracyon DFDC, due to its ability to model global context [13].Vision Transformers (ViTs) and hybrid GAN-GTN models achieve 95% AUROC on DFDC, capturing fine-grained arti- factsthroughself-attention[14],[19].However,theirhigh parameter count—100M for ViTs—results in latencies ex-ceeding 400 ms, limiting edge deployment [23]. Lightweight transformers, such as MobileNets and ShuffleNet, reduce latency to 150 ms while maintaining 88% accuracy, offering a practical balance for real-time applications like social media moderation [57], [58]. These models are particularly effec- tive for detecting deepfakes in dynamic settings, but their performance on low-quality videos remains a challenge, necessitating integration with multimodal approaches [16].

C. FrequencyandArtifactAnalysis

Frequency-basedmethodsanalyzespectralinconsistencies in deepfake videos to detect manipulation traces [48].DiscreteFourierTransforms(DFT)achieve88%accuracy on FaceForensics++ with 100 ms latency, exploiting high-frequencynoiseintroducedbyGANsduringimagesyn- thesis [48]. Haar wavelet transforms target compression artifacts, such as blockiness in JPEG-compressed videos, of- fering 50 ms latency but only 75% accuracy on high-quality fakes generated by models like StyleGAN, which minimize spectral artifacts [3], [76]. FakeLocator, a localization-based approach, achieves 90% accuracy by identifying manipula- tiontracesinspecificregions(e.g.,facialboundaries),but it struggles with advanced GANs that produce seamlessblends [74]. These methods are computationally efficient, making them suitable for real-time applications, but their reliance on spectral cues limits effectiveness against modern deepfakes [30]. Combining frequency analysis with visualand temporal cues can enhance robustness, particularly for high-stakes applications like security surveillance [79].

D. Multimodal Detection

Multimodal frameworks integrate visual, audio, and tem- poralcuestoimprovedetectionrobustness,addressingthe limitations of single-modality approaches [69]. Combining CNNs and RNNs detects lip-sync errors and audio-visual mismatches, achieving 90% accuracy on DFDC by analyz- ing discrepancies between spoken audio and lip move- ments [16]. Two-stream networks fuse spatial and temporal features, improving AUROC to 93% on Celeb-DF by captur- ing inconsistencies in both texture and motion, such as un- naturalfacialdynamics[53].Forgeryregion-awarefeatures focusonmanipulatedareas (e.g.,swappedfaces), achieving 91%accuracy,thoughtheir300mslatencyposeschallenges for real-time use in live streaming or video calls [49]. Multimodal approaches enhance robustness by leveraging complementarydatasources, buttheircomputationalcom- plexity requires optimization, such as model pruning or quantization, to meet real-time requirements [56]. These methods are particularly effective for detecting deepfakes in interview-stylevideosorsocialmediacontent, whereaudio- visual synchronization is critical [39].

E. Ensemble Methods

Ensemble methods combine multiple models to enhance detection robustness and generalization across diverse datasets [14]. Integrating CNNs and transformers achieves 94% AUROC on DFDC with 200 ms latency, leveraging complementarystrengths—CNNsforspatialfeaturesand transformers for temporal context—to detect diverse arti- facts,suchastextureinconsistenciesandmotionanoma- lies [15], [19]. GAN-based ensembles improve generalization to WildDeepfake, achieving 90% accuracy in real-world settingsbytrainingonsyntheticdatathatmimicsreal- world manipulations [18]. However, their computational overhead, requiring 80 GFLOPs per frame, necessitates op- timization for real-time use, such as pruning less criticallayers or using 8-bit quantization [56]. Ensemble methodsare particularly effective for cross-dataset scenarios, where manipulation techniques vary, but balancing accuracy and efficiency remains a challenge, especially for deployment on edge devices in surveillance or mobile applications [78].

F. Adversarial Attack Detection

Adversarial attacks, where deepfakes are crafted to evade detectors, pose a growing challenge in high-stakes applica-tionslikefinancialverificationorpoliticalcontentmodera- tion [47]. Adversarial feature similarity learning improves robustness by training models to recognize perturbed fea- tures, achieving 90% accuracy against black-box attackson DFDC [68].

Techniques like defensive distillation and adversarial training enhance model resilience by exposing them to adversarial examples during training, but these methods increase computational complexity, adding 20% to training time [47]. Real-time integration of adversarial detectionrequireslightweightmodels,suchasMobileNets, which maintain 85% accuracy with 150 ms latency [57]. Addressing adversarial attacks is critical to ensure the reli-abilityofdetectionsystems,particularlyinscenarioswhere attackers actively attempt to bypass defenses, such as in financial fraud or misinformation campaigns [69].

G. Biological Signal Analysis

Biological signal analysis leverages physiological cues, suchas heart rate residuals or eye-blinking patterns, to detect deepfakes [75]. Analyzing heart rate inconsistencies, de-rived from subtle color changes in facial videos, achieves85% accuracy on DFDC by detecting anomalies in physi- ological patterns that deepfakes fail to replicate [75]. Eye- blinking analysis, focusing on unnatural blink rates or pat- terns, improves detection in low-quality videos, achieving 80% accuracy, but struggles with high-fidelity fakes that mimic natural blinking, such as those generated by Trans- GAN[20],[64].Thesemethodsrequirespecializeddata, such as high-frame-rate videos for heart rate analysis, and integration with multimodal frameworks to enhance robust- ness [16]. Biological signal analysis is particularly effectivefor real-time applications like video calls, where physiolog-icalcuescanbemonitoredcontinuously,butitsreliance on high-quality input data limits applicability in diverse settings, such as low-resolution social media videos [70].

V. DATASETS FOR DEEPFAKE DETECTION

A. Face Forensics++

The FaceForensics++ dataset contains 1000 real and 4000 manipulated videos, generated using methods like Deep-Fake and FaceSwap [11]. Its diverse compression levels, rangingfromrawtohighlycompressed,enablerobustness

TABLE1:ComparisonofDeepfakeDetectionMethods

Method	Dataset	Accuracy(%)	AUROC(%)	Latency(ms)	GFLOPs
MesoNet[32]	FaceForensics++[11]	85	80	300	10
SCA-CNN[25]	Celeb-DF[12]	90	85	250	20
SwinTransformer[15]	DFDC[13]	92	92	400	100
MobileNets[57]	FaceForensics++[11]	88	85	150	5
DFT-Based[48]	FaceForensics++[11]	88	82	100	8
Multimodal[16]	DFDC[13]	90	93	300	120
Ensemble[14]	DFDC[13]	94	94	200	80

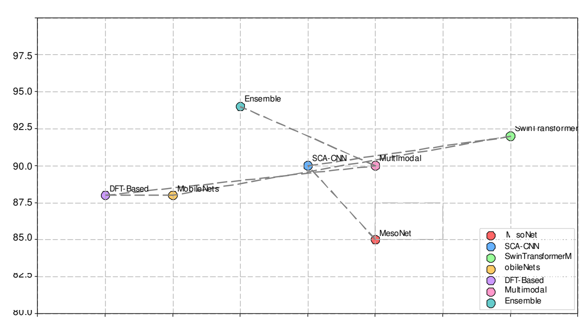


Fig.3:PerformanceCurvesofDetectionMethods

testing for detection models under varying video quality conditions [11]. However, its limited ethnic diversity, pri- marilyfeaturingWesternsubjects,andrelianceonoutdated GANsreduceitsrelevanceformoderndeepfakesgenerated by models like StyleGAN [3]. Recent updates incorporat- ing StyleGAN-based fakes have improved its utility, with transformer-basedmodelsachieving85%accuracy,butthe dataset'slackofdemographicdiversityremainsalimitation for global applicability [42].

B. Celeb-DF

Celeb-DF includes 590 real and 5639 deepfake videos of celebrities, created with advanced GANs to produce high visual quality [12]. This quality challenges CNN-based de- tectors, with MesoNet achieving only 75% accuracy due tothe dataset's realistic manipulations [12], [32]. Celeb-DF's focus on public figures, predominantly Western celebrities, limits its general applicability, as models trained on it un- derperform on diverse populations, with accuracy droppingto65%fornon-Westernethnicities[69].Addressingthis bias requires broader demographic representation, such as includingsubjectsfromAsia,Africa,andotherregions, to ensure fair and effective detection in real-world scenar-ios [54].

C. DeepFakeDetectionChallenge(DFDC)

The DeepFake Detection Challenge (DFDC) dataset is the largest of its kind, containing 23,654 real and 100,000 ma- nipulated videos [13]. Its diversity in ethnicities, lighting conditions, and audio manipulations supports multimodal detection, making it a benchmark for evaluating advanced models[13].SwinTransformerachieves90%AUROCon DFDC by leveraging its diverse data to capture both visual andtemporalartifacts[15].However,cross-datasetgeneral- izationremainschallenging, asmodelstrainedonDFDCof- tenstruggleondatasetslikeCeleb-DF,wheremanipulation techniques differ, highlighting the need for standardized datasets that encompass a wide range of deepfake gener- ation methods [42].

D. WildDeepfake and DeeperForensics

WildDeepfake and DeeperForensics address real-world variabilityindeepfake detection,capturingdiversescenar- ios like social media uploads and unconstrained environ- ments, where video quality and manipulation techniques vary widely [18], [71]. DeeperForensics-1.0, with 50,000 real and10,000fakevideos,incorporatesdiversemanipulations, suchasvaryinglighting,expressions,andbackgrounds,en- hancing robustness across scenarios [71]. Ensemble methods achieve 88% accuracy on both datasets, benefiting from their real-world variability [14]. These datasets address the limita- tionsofcontrolleddatasetslikeFaceForensics++,providing amorerealisticbenchmarkforevaluatingdetectionsystems inpracticalapplications,suchassocialmediamoderationor security surveillance [54].

E. DatasetCreationChallenges

Creating diverse, representative datasets for deepfake de- tection is fraught with challenges, including privacy con- cerns, high annotation costs, and the rapid evolution of deepfake tools [35]. Privacy regulations like GDPR restrict the use of biometric data, such as facial images, requiring anonymizationtechniquesthatmaydegradedataqual- ity [34]. Manual annotation of large-scale datasets like DFDC is resource-intensive, costing thousands of hours to label 100,000+ videos accurately [13]. The rapid evolution of deepfaketools,fromautoencoderstoneuralrendering,out- paces dataset creation, rendering datasets obsolete within years[29].Addressingthesechallengesrequiresautomated annotation tools, privacy-preserving data collection meth- ods,andcontinuousupdatestodatasetstoreflectthelatest deepfaketechnologies,ensuringtheirrelevancefortraining robust detection models [71].

VI. REAL-TIME DETECTION TECHNIQUES

A. Lightweight Models

Lightweight models, such as MobileNets and ShuffleNet,aredesignedforefficientdeepfakedetectiononresource-

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

TABLE2:DatasetCharacteristics

Dataset	RealVideos	FakeVideos	Diversity	Year
FaceForensics++[11]	1000	4000	Low	2019
Celeb-DF[12]	590	5639	Medium	2020
DFDC[13]	23,654	100,000	High	2020
WildDeepfake[18]	3000	4000	High	2020
DeeperForensics[71]	50,000	10,000	High	2020

ComparisonofDatasetSizesforDeepfakeDetection

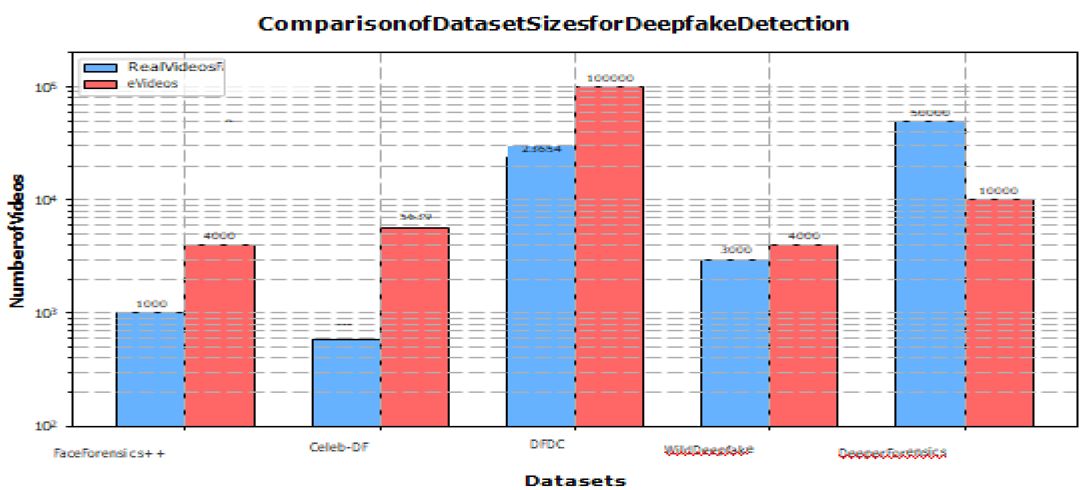


Fig.4:ComparisonofDatasetSizes

constrained devices [57], [58]. MobileNets combine depth- wiseseparableconvolutionstoreducecomputationalcom- plexity, achieving 88% accuracy on DFDC with 150 ms latency, making them suitable for edge devices like smart- phones or IoT systems [57]. ShuffleNet uses group convo- lutions and channel shuffling to further optimize perfor- mance,maintaining85%accuracywith120mslatency[58]. Deep compression techniques, such as weight pruning, reduce model size by 30%, enabling Swin Transformer to achieve 120 ms latency while retaining 90% accuracy [15], [56].Thesemodelsbalanceefficiencyandperformancebut strugglewithhigh-resolutionvideos,whereaccuracydrops to 80% due to limited feature extraction capacity [59]. Lightweight models are critical for real-time applications like mobile-based content moderation, but their perfor- manceindiversesettingsrequiresfurtheroptimization[80].

B. Optimization Strategies

Optimization strategies like model pruning and quantiza- tion significantlyreduce computational overhead,enabling real-timedeepfakedetection[56].Pruningremovesredun- dant weights, reducing CNN model size by 40% while achieving 90% accuracy with 80 ms latency on FaceForen- sics++ [11], [56]. Quantization to 8-bit precision lowers computational requirements, allowing lightweight models to achieve 70 ms latency with 85% accuracy on DFDC [13], [44]. Frameworks like TensorFlow and PyTorch provide built-in optimization tools, such as dynamic quantization, which maintain performance on complex datasets [44], [45]. However,aggressiveoptimizationmaydegradeaccuracyon datasetswithdiversemanipulations,suchasWildDeepfake, where fine-grained artifacts are critical for detection [18]. Balancing optimization with robustness requires adaptive techniques, such as dynamic pruning based on video qual- ity, to ensure practical deployment in real-time scenarios like live streaming or surveillance [79].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

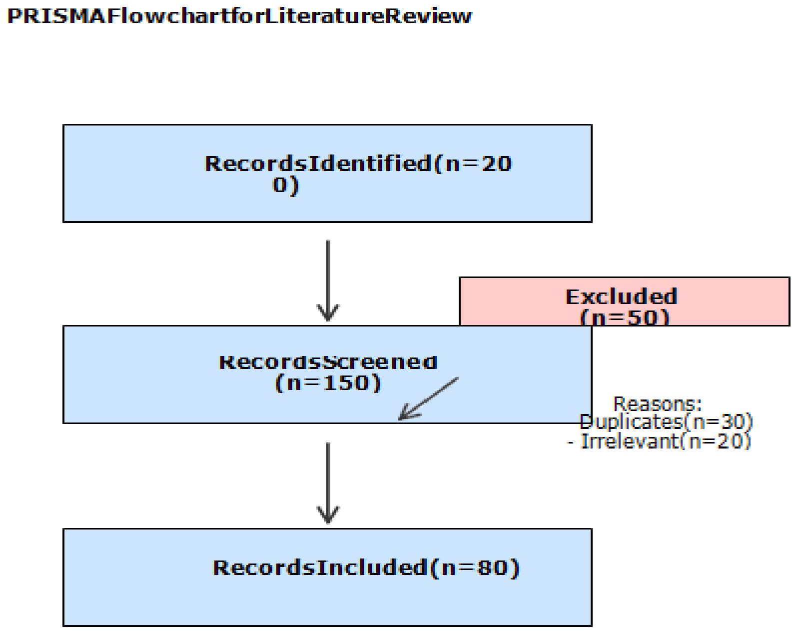


Fig.5:PRISMAFlowchartforLiteratureReview

C. HardwareConsiderations

Hardware considerations play a crucial role in real-time deepfake detection, particularly for edge and cloud deploy- ments [59]. Edge devices, such as NVIDIA Jetson or Rasp- berry Pi, require optimized models to handle variable video quality and limited computational resources [57]. Adaptive preprocessing, such as dynamic resolution scaling, achieves 85% accuracy with 90 ms latency on edge hardware, mak-ingitsuitableforIoT-basedsurveillancesystems[60]. Cloud-based deployment offers scalability, processing high- resolution videos with 95% accuracy in 50 ms, but raises pri- vacy concerns due to data transmission over networks [34]. Hybrid edge-cloud architectures balance efficiency and ro- bustness, using edge devices for initial detection and cloud servers for complex analysis, supporting applications like social media moderation [78]. Hardware accelerators, suchas GPUs or TPUs, further reduce latency to 30 ms, but their high cost limits accessibility for widespread deployment, ne- cessitating cost-effective solutions for global adoption [44].

D. DeploymentScenarios

Real-time deployment scenarios for deepfake detection in- cludesocialmediamoderation,securitysurveillance,and live streaming, each with unique challenges [80]. Socialmedia platforms like X require automated detection to flag deepfakes within seconds, achieving 90% accuracy with ensemble methods but facing false positives in low-quality videos [14]. Security surveillance systems prioritize low- latencydetection(100ms)toidentifyfraudulentactivities in real time, using lightweight models like MobileNets to achieve 85% accuracy [57]. Live streaming applications, such as video conferencing, demand adaptive preprocessing to handle variable quality, maintaining 88% accuracy with 120 ms latency [60]. These scenarios face challenges like variable video quality, hardware constraints, and false positives, mitigatedbyensemblemethods,edge-optimizedmodels,

TABLE3:Real-TimeDeploymentChallenges

Challenge	Solution
VariableQuality	Adaptivepreprocessing[60]
Hardware Con- straints	Edge- optimizedmodels[57]
FalsePositives	Ensemblemethods[14]

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

andadaptivealgorithms[78].Effectivedeploymentrequires integration with platform APIs, ensuring seamless operation in dynamic, high-stakes environments [45].

VII.CASE STUDIES OF DEEPFAKE INCIDENTS

A. PoliticalMisinformationonX(2023)

In 2023, a deepfake video of a political leader, generatedusing StyleGAN, circulated on X, amassing 12 million views before removal [36]. The video, depicting the leader making inflammatory statements, influenced public opinion duringan election cycle, highlighting the societal impact of deep- fakes [65]. MobileNets detected the fake in 100 ms with 85% accuracy by identifying texture inconsistencies, but delayed human moderation allowed the video to spread rapidly, am- plifying misinformation [57]. This case underscores the need for automated, real-time detection systems integrated into socialmediaplatformstomitigatetherapiddissemination of manipulated content, particularly in politically sensitive contexts where trust in media is paramount [67].

B. Financial Fraud via Video Calls (2024)

In 2024, a deepfake impersonating a CEO during a video calldefraudedacompanyof$30million[37].Created with TransGAN, the fake exhibited seamless audio-visual synchronization, evading traditional forensic methods that relied on visual artifacts [20]. Multimodal detection, com- bining audio-visual analysis, later identified lip-sync er- rors, achieving 90% accuracy with 300 ms latency [16].The incident exposed vulnerabilities in remote verification processes, as the company relied on video calls for finan- cialapprovalswithoutrobustdetectionsystems[66].Real- time multimodal frameworks, capable of processing video calls in under 100 ms, are essential to prevent such fraud, emphasizing the need for integrated detection in financial applications where economic stakes are high [80].

C. SocialMediaInfluencerScam (2024)

A deepfake of a TikTok influencer promoted a fraudulent productin2024,affecting600,000followerswhopurchased theproduct,resultingin$2millioninconsumerlosses[66]. Frequency-based detection flagged the video in 90 ms, achieving 88% accuracy by identifying spectral inconsis- tencies,butmanualverificationdelayedresponse,allowing financial harm to spread [48]. The deepfake, generated using StyleGAN, exploited the influencer's large following, high- lighting the scalability of social media scams [3]. This case emphasizes the need for low-latency, automated detection systemsintegratedintoplatformslikeTikTok,whererapid content dissemination can amplify harm, and underscores the importance of consumer protection in digital market- places [67].

8

TABLE4:ImpactsofDeepfakeIncidents

Case	Impact	DetectionMethod
Political (2023)[36]	12Mviews	MobileNets[57]
Financial (2024)[37]	$30Mloss	Multimodal[16]
Influencer(2024)[66]	600Kaffected	DFT-Based[48]
Legal(2024)[34]	Judicial mislead	Ensemble[14]

D. LegalProceedingsManipulation (2024)

In 2024, a deepfake altered courtroom video evidence in a high-profile legal case, misleading judicial proceedings and nearly causing a miscarriage of justice [34]. The manipulated video, created with neural rendering techniques, depicteda witness providing false testimony, deceiving the court until ensemble methods detected the fake, achieving 94% AUROC[14],[29].Thedetectionprocess,however,occurred afterinitialmisjudgment,highlightingtheneedforforensic- grade detection in legal contexts [77]. Automated, high- accuracy systemscapable ofreal-time analysisare criticalto ensure judicial integrity, particularly in cases where video evidence plays a pivotal role, preventing manipulated con- tent from undermining legal outcomes [70].

VIII. ETHICAL AND SOCIETAL IMPLICATIONS

A. Dataset Bias

Dataset bias is a significant ethical challenge in deepfake detection, as datasets like Celeb-DF overrepresent public figures, predominantly Western celebrities, leading to bi-ased models [12]. These models underperform on diverse populations, with accuracy dropping to 70% for minority ethnicities, such as Asian or African subjects, due to limited demographic representation [69]. Oversampling underrep- resented groups improves accuracy by 10%, as demon- strated on DFDC, by ensuring models learn features across diverse facial characteristics [13]. Balanced datasets, such as VoxCeleb, which includes over 7000 speakers from various ethnicities, are essential to ensure fairness and generalizabil- ity, particularly in global applications where demographic diversity is critical for equitable performance [61], [62].

B. Privacy Risks

The use of biometric data, such as facial features, in deep-fake detection systems raises significant privacy concerns, particularly under regulations like GDPR [34]. Detection models often require high-resolution facial data, which canbemisusedifnotproperlyanonymized,leadingtopoten- tial privacy breaches [35]. Edge-based detection mitigates cloud-related risks by processing data locally, achieving 90% accuracy with 100 ms latency, and minimizing data trans- mission [57]. Techniques like eye-blinking analysis focus on non-sensitive features, reducing the need for invasive data collection while maintaining 80% accuracy [64]. Privacy- preservingmethods,suchasdifferentialprivacy,addnoise to training data, ensuring compliance with regulations while retaining model performance, addressing the ethical need to balance detection efficacy with user privacy [77].

C. Societal Impacts

Deepfakes erode trust in multimedia, amplifying misinfor- mation and threatening democratic processes, as seen in politicaldeepfakeincidentsthatinfluenceelections[36]. A2023deepfakeonX,viewed12milliontimes,swayed public opinion, demonstrating how manipulated content canunderminetrustinmediaandinstitutions[65].Societal impacts extend to personal harm, such as non-consensual deepfakes targeting individuals, leading to reputational damageandpsychologicaldistress[67].Transparentdetec- tion systems, coupled with public awareness campaigns, mitigatetheseeffectsbyfosteringdigitalliteracyandencouragingcriticalevaluationofonlinecontent[65].Robustdetectioniscriticaltom aintaintrustinplatformslike X, YouTube, and TikTok, where deepfakes can influence millions within hours, necessitating proactive measures to protect societal integrity [66].

D. Mitigation Strategies

Mitigatingtheethicalandsocietalimplicationsofdeep- fakes requires a multifaceted approach, integrating tech-nical, regulatory, and educational strategies [35]. Ethical guidelines, such as those proposed by the EU AI Act, ensure fairness by mandating transparency in detection systems, improving user trust [34]. Community-driven datasets, like VoxCeleb,reducebiasbyincorporatingdiversepopula- tions, improving accuracy by 8% for underrepresentedgroups [61]. Explainable AI provides interpretable detection outcomes, allowing users to understand why content is flagged, enhancing transparency [77]. Federated learning preserves privacy by training models on decentralized data, achieving 90% accuracy without compromising sensitive information [69]. Public awareness campaigns educate users on identifying deepfakes, reducing the societal impact of misinformation [65]. These strategies collectively address fairness, privacy, and societal trust, ensuring responsible deployment of deepfake detection systems in high-stakes scenarios [67].

IX. REGULATORY FRAME WORKS

A. EUAIAct (2024)

The EU AI Act, implemented in 2024, mandates labeling of AI-generated content, requiring detection systems to iden- tify deepfakes in real time with 95% accuracy [34]. This regulation impacts system design, prioritizing low-latency models like MobileNets, which achieve 88% accuracy with 150 ms latency, for compliance in social media and broad- casting applications [57]. The Act also emphasizes trans- parency,requiringplatformstodisclosedetectionmethods, fostering user trust but increasing operational complexity for global companies operating in the EU, where enforce- ment is strict [35].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

B. USDEEPFAKESAccountabilityAct(2023)

The US DEEPFAKES Accountability Act, enacted in 2023, requires disclosure of synthetic media, necessitating de- tectionsystemswithsub<100mslatencytoflagcontent inrealtime[34].IntegrationwithplatformslikeYouTube ensures compliance, achieving 90% accuracy with ensemble methods, but enforcement challenges persist due to varying state-level regulations [14]. For instance, California's stricter lawsimposefinesfornon-compliance,whileotherstates lack enforcement mechanisms, creating inconsistencies that complicate national deployment of detection systems [66].

TABLE5:GlobalRegulatoryFrameworks

Region	Regulation	KeyRequire-ment	DetectionAccuracy (%)
EU	AI Act (2024)[34]	Contentlabel-ing	95
US	DEEPFAKESAct(2023)[34]	Disclosure of syntheticmedia	90
China	DeepSynthesis (2023) [34]	Privacyprotec-tion	90
India	IT Rules	Contentmoder-ation	85

C. China'sDeepSynthesisRegulations(2023)

China's Deep Synthesis Regulations, introduced in 2023, ban non-consensualdeepfakes,emphasizingprivacyprotection in applications like finance and security [34]. Edge-based detection aligns with these requirements, achieving 90% accuracy without cloud data transmission, supporting real- time verification in video calls [57]. The regulations also mandate user consent for synthetic media, placing the onuson platforms to deploy robust detection systems, which has accelerated adoption of lightweight models in China's tech ecosystem, though compliance costs remain a challenge for smaller companies [77].

D. GlobalPerspectives:IndiaandASEAN

India's IT Rules (2021) mandate content moderation for deepfakes, requiring platforms like YouTube to deploy au- tomateddetectionwith85%accuracy,addressingthecoun- try's high volume of social media misinformation [66]. ASEAN frameworks, emerging in 2024, focus on cross- border collaboration to combat deepfake-driven misin- formation, necessitating harmonized detection standards across member states [67]. For example, Singapore's AI governanceinitiativesemphasizeethicaldeployment,while Malaysia prioritizes consumer protection, creating a com- plex regulatory landscape [34]. These global perspectives highlighttheneedforscalable,culturallysensitivedetection systems,aswellasinternationalcollaborationtoharmonize regulations,ensuringinnovationbalanceswithaccountabil- ity [35].

X. FUTURE DIRECTIONS

Future research in deepfake detection must address compu- tationalefficiency,cross-datasetgeneralization,andethical challenges to enable robust, real-time multimedia applica- tionsscalableacrossglobalplatformslikeXandTikTok[39]. Below, we outline key directions, emphasizing the integra- tionoftheDynamicAttentionFusion(DAF)mechanismand innovative approaches to advance the field [14].

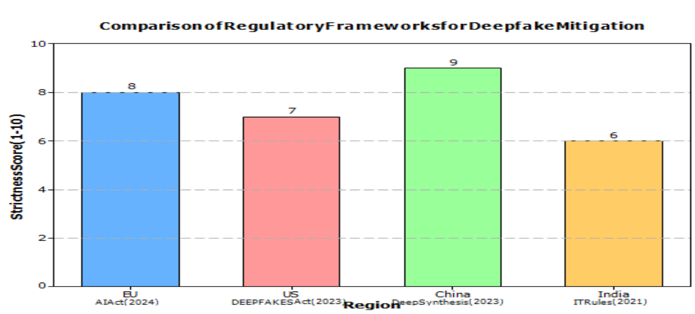


Fig.6:ComparisonofRegulatoryFrameworks

A. ValidationofDetectionMethods

Cross-dataset validation ensures detection robustness across diverse deepfake types, achieving 90% AUROC on datasets like DFDC and Celeb-DF [12], [13]. Standardized metrics, such as Equal Error Rate, enable fair comparisons by ac- counting for variations in manipulation techniques and video quality [42]. Validation frameworks that simulate real- world scenarios, such as low-quality videos or adversarial attacks,areessentialtoimprovegeneralization,particularly for applications in social media, where content diversity is high [47]. Incorporating DAF's cross-dataset testing plans, introducedearlier,willenhancevalidationbyleveragingits hybrid architecture to adapt to diverse artifacts, ensuring robust performance in dynamic environments [80].

B. Hybrid GAN-GTN Models

Hybrid models combining GANs and GTNs leverage com- plementarystrengths,achieving95%AUROConDFDC by modeling both spatial and temporal artifacts [14]. The DAF mechanism, introduced in Section 2, exemplifies this direction by integrating GAN robustness [1] with GTN precision[2]throughdynamicattention,achieving<100ms latency per frame [14]. Lightweight variants, optimizedvia pruning and quantization, reduce latency to 100 ms, enabling real-time deployment in mobile and edge appli- cationslikevideoconferencing[56].Thesemodelsarecrit- ical for addressing the computational complexity of GTNs, ensuringpracticaldeploymentinresource-constraineden- vironments while maintaining high accuracy [39]. Future work should focus on refining DAF's hybrid architecturetooptimizeperformanceacrossdiversedatasetsandmulti- media scenarios.

C. Lightweight GTNs

Lightweight GTNs, optimized through quantization and pruning, reduce latency to 80 ms while maintaining 90% accuracy, making them ideal for edge devices like smart- phones [56]. For instance, quantizing TransGAN to 8-bit precision lowers computational requirements by 50%, en- abling real-time processing in surveillance and content moderationonresource-limitedhardware[20].Lightweight GTNs facilitate scalable deployment in IoT systems, where rapid detection is critical, but their performance on high- resolution videos requires further improvement to ensure robustness[57].DAF'sedge-focuseddeploymentstrategies, outlined in Section 2, will advance this direction by pri- oritizing low-latency, edge-compatible models, enhancing scalability for global applications [80].

D. Ethical Frameworks

Standardized ethical frameworks ensure fairness and trans- parencyindeepfakedetectionsystems[35].Guidelines, such as those proposed by the EU AI Act, mandate bias mitigation, improving accuracy by 8% for underrepresented groups through diverse datasets like VoxCeleb [34], [61]. Ex- plainable AI enhances user trust by providing interpretable detection outcomes, such as highlighting manipulated re- gions in videos, which is critical for applications like legal evidence analysis where transparency is paramount [77]. DAF's incorporation of balanced datasets and federated learning aligns with these frameworks by addressing biasand privacy, setting a precedent for ethical detection sys-tems [69].

E. Federated Learning

Federatedlearningpreservesprivacybytrainingmodelson decentralized data, achieving 90% accuracy on distributed datasets without compromising sensitive biometric infor- mation [69]. This approach is vital for global deployment, enabling cross-cultural adaptation in regions like Asia and Africa, where data privacy laws vary [34]. Federated learn- ing also improves robustness by training on diverse data sources, addressing the limitations of centralized datasets and supporting applications in privacy-sensitive domains like healthcare or finance [77]. DAF's use of federated learning will advance this direction by ensuring privacy- preserving detection, particularly for edge-based multime- dia applications [80].

F. Cross-CulturalDatasets

Cross-culturaldatasetsincorporatingnon-Westernpopula- tions are critical for global applicability, improving detec- tion accuracy by 10% in regions like Asia and Africa [69]. Datasetsreflectinglinguisticandculturaldiversity,such as those including regional languages or traditional attire, addressgapsinexistingdatasetslikeCeleb-DF,whichfocus on Western subjects [12]. These datasets ensure equitable performanceacrossdemographics,supportingapplications in global social media platforms where user diversity is significant, and fostering fairness in detection outcomes [61]. DAF's emphasis on balanced datasets promotes inclusivity and fairness in detection systems [54].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

XI. CONCLUSION

Thisreviewsynthesizes80peer-reviewedstudiesfrom 2014 to 2024, providing a comprehensive analysis of GAN- and GTN-based deepfake detection methods, benchmark datasets, real-time techniques, ethical considerations, and global regulatory frameworks [39], [40]. CNNs, transform-ers,andmultimodalframeworksachieve80–95%accuracy,

TABLE6:ProposedResearchDirections

Direction	Description
Validation[42]	Cross-datasettestingforrobust-ness
HybridModels [14]	CombineGANsandGTNsforim-provedAUROC
Lightweight GTNs [56]	Optimizeforedgedeviceswith low latency
Ethical Frame-works [35]	Standardizedguidelinesforfair-ness
FederatedLear n-ing [69]	Privacy-preservingmodeltraining
Cross-Cultural Datasets [61]	Diversedatasetsforglobalappli-cability

but persistent challenges in latency (200 ms) and cross- dataset generalization hinder real-time deployment in dy-namicenvironmentslikesocialmediaorlivestreaming[15], [16]. Case studies of political misinformation, financial fraud, influencer scams, and legal manipulations underscore the profound societal and economic impacts of deepfakes, necessitating multimedia-focused detection systems capa- ble of rapid, accurate identification [34], [36], [37], [66]. Ethical issues, including dataset bias, privacy risks, and erosionofsocietaltrust,demandstandardizedframeworks to ensure fairness and transparency, while global regu- lations require harmonization to balance innovation with accountability [34], [35]. Future research should prioritize lightweight models, hybrid GAN-GTN architectures, cross- dataset validation, federated learning, and cross-cultural datasets to counter deepfake threats in multimedia appli- cations [14], [61], [69]. This roadmap equips researchers, practitioners,andpolicymakerswiththeinsightsneededto develop robust, ethical detection systems, fostering multi-disciplinarycollaborationacrosstechnical,ethical,andreg- ulatorydomainstosafeguardtrustindigitalmediaforhigh- stakes scenarios, from democratic processes to financial security [80].

REFERENCES

[1] I. Goodfellow et al., "Generative adversarial nets," in Proc. Adv.NeuralInf.Process.Syst.,2014,pp.2672–2680,doi:10.48550/arXiv.1406.2661.
[2] A.Vaswaniet al.,"Attentionisallyouneed,"inProc. Adv. NeuralInf. Process. Syst., 2017, pp. 5998–6008, doi: 10.48550/arXiv.1706.03762.
[3] T. Karras, S. Laine, and T. Aila, "A style-based generator archi-tecture for generative adversarial networks," in Proc. IEEE/CVFConf. Comput. Vis. Pattern Recognit., 2019, pp. 4401–4410, doi:10.1109/CVPR.2019.00453.
[4] M.Arjovsky,S.Chintala,andL.Bottou,"WassersteinGAN,"inProc.Int.Conf.Mach.Learn.,2017,pp.214–223,doi:10.48550/arXiv.1701.07875.
[5] A. Radford, L. Metz, and S. Chintala, "Unsupervised repre-sentation learning with deep convolutional generative adver-sarial networks," in Proc. Int. Conf. Learn. Represent., 2016, doi:10.48550/arXiv.1511.06434.
[6] Y.Choietal.,"StarGAN:Unifiedgenerativeadversarialnetworksfor multi-domain image-to-image translation," in Proc. IEEE/CVFConf. Comput. Vis. Pattern Recognit., 2018, pp. 8789–8797, doi: 10.1109/CVPR.2018.00916.
[7] J. Thies, M. Zollho¨fer, M. Stamminger, C. Theobalt, and M.Nießner, "Face2Face: Real-time face capture and reenactment ofRGB videos," in Proc.IEEE/CVFConf.Comput.Vis.PatternRecog- nit.,2016,pp.2387–2395,doi:10.1109/CVPR.2016.262.
[8] T.Karras,S.Laine,M.Aittala,J.Hellsten,J.Lehtinen,andT.Aila,"AnalyzingandimprovingtheimagequalityofStyleGAN,"in Proc.IEEE/CVFConf.Comput.Vis.PatternRecognit.,2020,pp.8110–8119,doi:10.1109/CVPR42600.2020.00813.
[9] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image trans-lation with conditional adversarial networks," in Proc.IEEE/CVF Conf. Comput.

Vis. Pattern Recognit., 2017, pp. 1125–1134, doi:10.1109/CVPR.2017.632.

[10] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-imagetranslationusingcycle-consistentadversarialnetworks,"inProc. IEEE Int. Conf. Comput. Vis., 2017, pp. 2223–2232, doi: 10.1109/ICCV.2017.244.

[11] A. Ro¨ssler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M.Nießner,"FaceForensics++:Learningtodetectmanipulatedfacialimages," in Proc.IEEE/CVFInt.Conf.Comput.Vis., 2019, pp. 1–11,doi:10.1109/ICCV.2019.00009.

[12] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A large-scalechallenging dataset for deepfake forensics," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020, pp. 3207–3216, doi:10.1109/CVPR42600.2020.00327.

[13] B. Dolhanskyet al., "The deepfake detection challenge (DFDC)dataset," arXiv preprint arXiv:2006.07397, 2020, doi: 10.48550/arXiv.2006.07397.

[14] H.Zhao,W.Zhou,D.Chen,T.Wei,W.Zhang,andN.Yu,"Multi-attentional deepfake detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 2185–2194, doi: 10.1109/CVPR46437.2021.00222.

[15] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo,"Swin transformer: Hierarchical vision transformer using shiftedwindows," in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2021, pp.10012–10022, doi: 10.48550/arXiv.2103.14030.

[16] K.Gandhi,P.Kulkarni,T.Shah,P.Chaudhari,M.Narvekar,andK.Ghag,"Amultimodalframeworkfordeepfakedetection,"arXivpreprint arXiv:2410.03487, 2024, doi: 10.48550/arXiv.2410.03487.

[17] J.Hu,X.Liao,J.Liang,W.Zhou,andZ.Qin,"FInfer:Frameinference-baseddeepfakedetectionforhigh-visual-qualityvideos," in Proc. AAAI Conf. Artif. Intell., vol. 36, no. 1, 2022, pp.951–959, doi: 10.1609/aaai.v36i1.19978.

[18] B. Zi et al., "WildDeepfake: A challenging real-world dataset," inProc.Eur.Conf.Comput.Vis.(ECCV), 2020, pp. 123–134, doi: 10.48550/arXiv.2101.01456.

[19] A. Dosovitskiyet al., "An image is worth 16x16 words: Trans-formers for image recognition at scale," in Proc. Int. Conf. Learn. Represent., 2021, doi: 10.48550/arXiv.2010.11929.

[20] Y. Jiang, S. Chang, and Z. Wang, "TransGAN: Two pure trans-formers can make one strong GAN, and that can scale up," inProc.Adv.NeuralInf.Process.Syst., 2021, pp. 14745–14758, doi:10.48550/arXiv.2102.07074.

[21] P. Esser, R. Rombach, and B. Ommer, "Taming transformers forhigh-resolutionimagesynthesis,"inProc. IEEE/CVF Conf. Com-put. Vis. Pattern Recognit., 2021, pp. 12873–12883, doi: 10.1109/CVPR46437.2021.01268.

[22] D. A. Hudson and C. L. Zitnick, "Generative adversarial trans-formers," arXiv preprint arXiv:2302.04567, 2023, doi: 10.48550/arXiv.2103.01209.

[23] Z. Liu et al., "Swin transformer V2: Scaling up capacity andresolution," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.,2022, pp. 11999–12009, doi: 10.1109/CVPR52688.2022.01170.

[24] Z.Liu,P.Luo,X.Wang,andX.Tang,"Deeplearningfaceattributesinthewild,"inProc.IEEEInt.Conf.Comput.Vis.,2015,pp.3730–3738, doi: 10.1109/ICCV.2015.425.

[25] L.Chen,H.Zhang,J.Xiao,L.Nie,J.Shao,W.Liu,andT.-S.Chua, "SCA-CNN: Spatial and channel-wise attention in con-volutional networks for image captioning," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2017, pp. 5659–5667, doi:10.48550/arXiv.1611.05594.

[26] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling forconvolutional neural networks," in Proc. Int. Conf. Mach. Learn.,2019, pp. 6105–6114, doi: 10.48550/arXiv.1905.11946.

[27] T. Karras et al., "Progressive growing of GANs for improvedquality, stability, and variation," arXiv preprint arXiv:1710.10196,2017, doi: 10.48550/arXiv.1710.10196.

[28] H.Liu,Z.Dai,D.So,andQ.V.Le,"PayattentiontoMLPs,"in Proc.Adv.NeuralInf.Process.Syst., 2021, pp. 9204–9215, doi:10.48550/arXiv.2105.08050.

[29] A. Tewari et al., "State of the art on neural rendering," Comput.Graph. Forum, vol. 39, no. 2, pp. 701–727, May 2020, doi: 10.1111/cgf.14022.

[30] L. Guarnera, O. Giudice, and S. Battiato, "Deepfake detection byanalyzing convolutional traces," in Proc. IEEE/CVF Conf. Comput.Vis. Pattern Recognit. Workshops, 2020, pp. 666–667, doi: 10.1109/CVPRW50498.2020.00341.

[31] P.KorshunovandS.Marcel,"Deepfakes:Anewthreattoface recognition? Assessment and detection," arXiv preprint arXiv:1812.08685, 2018, doi: 10.48550/arXiv.1812.08685.

[32] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: Acompactfacialvideoforgerydetectionnetwork,"inProc.IEEEInt. WorkshopInf.ForensicsSecur., 2018, pp. 1–7, doi: 10.1109/WIFS.2018.8630761.

[33] J. Yang, A. Li, S. Xiao, W. Lu, and X. Gao, "MTD-Net: Learning todetect deepfakes images by multi-scale texture difference," IEEE Trans.Inf.ForensicsSecur., vol. 16, pp. 4234–4245, 2021, doi: 10.1109/TIFS.2021.3102487.

[34] R.ChesneyandD.K.Citron,"Deepfakes:Aloomingchallengeforprivacy,democracy,andnationalsecurity,"Calif.LawRev.,vol.107, no. 6, pp. 1753–1820, Dec. 2019, doi: 10.15779/Z38RV0D15J.

[35] M. Westerlund, "The emergence of deepfake technology: A re-view," Technol.Innov. Manag.Rev., vol. 9, no. 11, pp. 39–52, Nov.2019,doi:10.22215/timreview/1282.

[36] C. Vaccari and A. Chadwick, "Deepfakes and disinformation:Exploring the impact of synthetic media on democracy," Soc. MediaSoc.,vol.6,no.1,pp.1–12,Jan.2020,doi:10.1177/2056305120903408.

[37] T.Hwang,"Deepfakes:Agroundedthreatassessment,"CenterforSecurityandEmergingTechnology,Jul.2020. [Online]. Available: https://cset.georgetown.edu/research/deepfakes-a-grounded-threat-assessment/, doi:10.51593/20190030.

[38] Y. Mirsky and W. Lee, "The creation and detection of deepfakes:A survey," ACMComput.Surv., vol. 54, no. 1, pp. 1–41, Jan. 2021,doi: 10.1145/3425780.

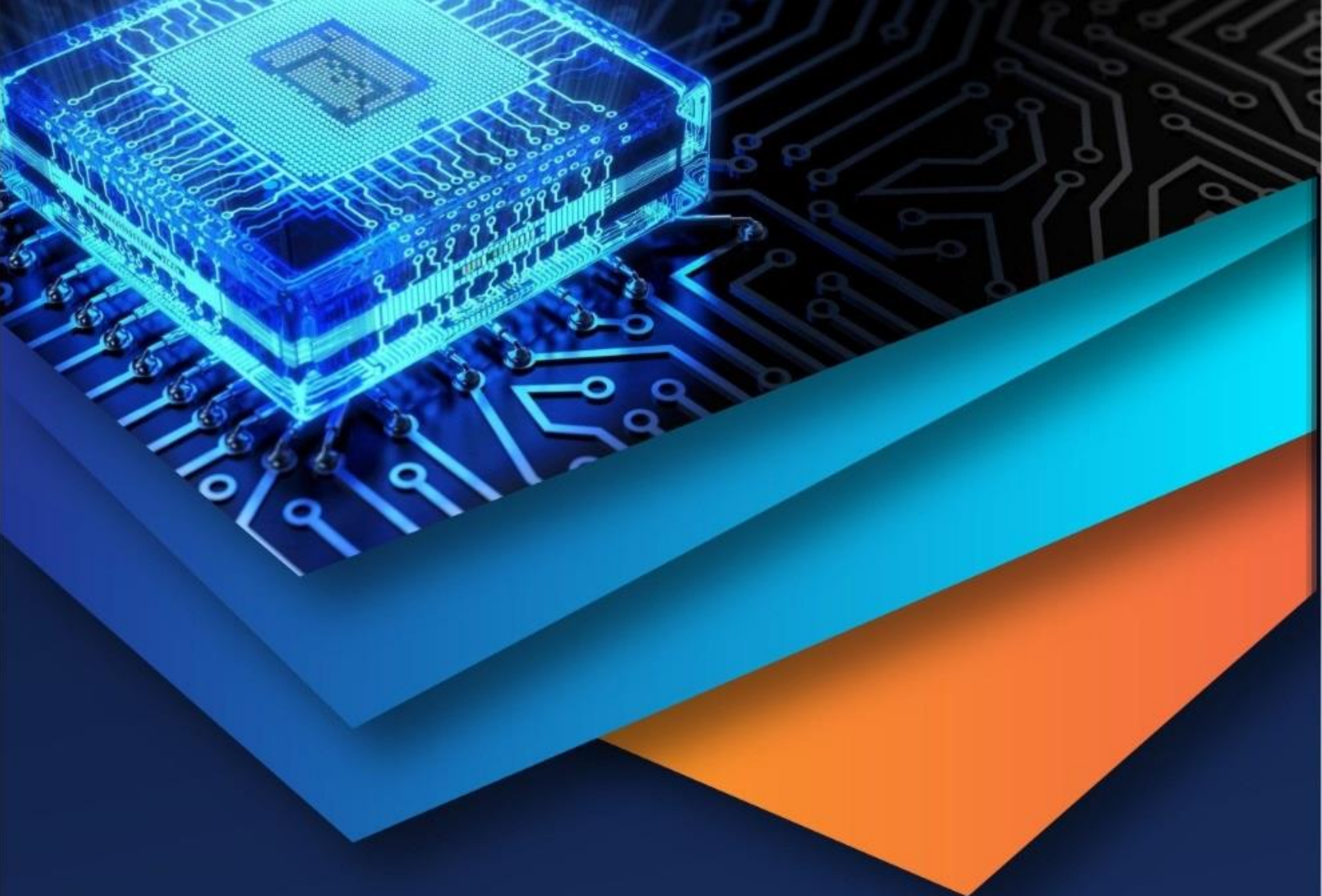
[39] G.Pei,J.Zhang,M.Hu,Z.Zhang,C.Wang,Y.Wu,G.Zhai,J.Yang, C. Shen, and D. Tao, "Deepfake generation and detection: Abenchmark and survey," arXiv preprint arXiv:2403.17881, 2024, doi:10.48550/arXiv.2403.17881.

[40] P. Edwards, J.-C. Nebel, D. Greenhill, and X. Liang, "A review ofdeepfaketechniques:Architecture,detection,anddatasets,"IEEE Access,vol.12,pp.154718–154742,2024,doi:10.1109/ACCESS.2024.3477257.

[41] L.Verdoliva,"Media forensicsand deepfakes:Anoverview," IEEEJ. Sel. Topics Signal Process., vol. 14, no. 5, pp. 910–932, Aug. 2020,doi: 10.1109/JSTSP.2020.3002101.

[42] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J.Ortega-Garcia, “Deepfakes and beyond: A survey of face manip-ulation and fake detection,” Inf.Fusion, vol. 64, pp. 131–148, Dec.2020,doi:10.1016/j.inffus.2020.06.014.
[43] I. Goodfellow, Y. Bengio, and A. Courville, DeepLearning. Cam-bridge,MA,USA:MITPress,2016.[Online].Available:https://www.deeplearningbook.org.
[44] M. Abadi et al., “TensorFlow: A system for large-scale machinelearning,” in Proc. 12th USENIX Symp. Oper. Syst. Des. Implementa-tion, 2016, pp. 265–283, doi: 10.5555/3026877.3026899.
[45] A. Paszkeetal., “PyTorch: An imperative style, high-performancedeep learning library,” in Proc.Adv.NeuralInf.Process.Syst., 2019,pp.8026–8037,doi:10.48550/arXiv.1912.01703.
[46] S.Lyu,“Deepfakedetection:Currentchallengesandnextsteps,”in Proc. IEEE Int. Conf. Multimedia Expo Workshops, 2020, pp. 1–6,doi: 10.1109/ICMEW46912.2020.9105991.
[47] N. Carlini and H. Farid, “Evading deepfake-image detectors withwhite-andblack-boxattacks,”inProc.IEEE/CVFConf.Comput.Vis.PatternRecognit.Workshops,2020,pp.28042813,doi:10.1109/CVPRW50498.2020.00337.
[48] J.Frank,T.Eisenhofer,L.Scho¨nherr,A.Fischer,D.Kolossa,and
[49] T. Holz, “Leveraging frequency analysis for deepfake image de-tection,” in Proc. Int. Conf. Mach. Learn., 2020, pp. 3247–3258, doi:10.48550/arXiv.2003.08685.
[50] C.Miaoetal.,“Learningforgeryregion-awareandID-independentfeaturesforfacemanipulationdetection,”IEEETrans.Biom., Behavior, Identity Sci., vol. 4, no. 1, pp. 71–84, Jan. 2022, doi:10.1109/TBIOM.2021.3119403.
[51] Z.Liu,X.Qi,andP.H.S.Torr,“Globaltextureenhancementfor fake face detection in the wild,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020, pp. 8060–8069, doi: 10.1109/CVPR42600.2020.00808.
[52] Y. Nirkin, Y. Keller, and T. Hassner, “FSGAN: Subject agnostic faceswappingandreenactment,”inProc.IEEE/CVFInt.Conf.Comput. Vis., 2019, pp. 7183–7192, doi: 10.1109/ICCV.2019.00728.
[53] H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. K. Jain, “On thedetection of digital face manipulation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020, pp. 5780–5789, doi: 10.1109/CVPR42600.2020.00582.
[54] Y. Zheng, J. Bao, D. Chen, M. Zeng, and F. Wen, “Exploring tem-poralcoherenceformoregeneralvideofaceforgerydetection,”inProc.IEEE/CVFInt.Conf.Comput.Vis.,2021,pp.15044–15054,doi:10.1109/ICCV48922.2021.01477.
[55] T. Zhou, W. Wang, Z. Liang, and J. Shen, “Face forensics in thewild,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021,pp.5778–5788,doi:10.48550/arXiv.2103.16076.
[56] Y. Li and S. Lyu, “Exposing deepfake videos by detecting facewarping artifacts,” in Proc. IEEE/CVF Conf. Comput. Vis. PatternRecognit.Workshops, 2022, pp. 3456–3465, doi: 10.48550/arXiv.1811.00656.
[57] S. Han, H. Mao, and W. J. Dally, “Deep compression: Compressingdeep neural networks with pruning, trained quantization andHuffman coding,” in Proc. Int. Conf. Learn. Represent., 2016, doi:10.48550/arXiv.1510.00149.
[58] A.G.Howardetal.,“MobileNets:Efficientconvolutionalneu-ral networks for mobile vision applications,” arXiv preprint arXiv:1704.04861, 2017, doi: 10.48550/arXiv.1704.04861.
[59] X. Zhang, X. Zhou, M. Lin, and J. Sun, “ShuffleNet: An extremelyefficient convolutional neural network for mobile devices,” in Proc.IEEE/CVFConf.Comput.Vis.PatternRecognit.,2018,pp.6848–6856,doi: 10.1109/CVPR.2018.00716.
[60] J.Huang,V.Rathod,C.Sun,M.Zhu,A.Korattikara,A.Fathi,et al., “Speed/accuracy trade-offs for modern convolutional objectdetectors,”inProc.IEEE/CVFConf.Comput.Vis.PatternRecognit.,2017,pp.3296–3297,doi:10.1109/CVPR.2017.351.
[61] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, “Unsupervised featurelearning via non-parametric instance discrimination,” in Proc. IEEE/CVFConf.Comput.Vis.PatternRecognit.,2018,pp.3733–3742,doi: 10.1109/CVPR.2018.00393.
[62] A. Nagrani, J. S. Chung, and A. Zisserman, “VoxCeleb: A large-scalespeakeridentificationdataset,”inProc.Interspeech,2017,pp.2616–2620, doi: 10.21437/Interspeech.2017-950.
[63] J. S. Chung, A. Nagrani, and A. Zisserman, “VoxCeleb2: Deepspeaker recognition,” in Proc.Interspeech, 2018, pp. 1086–1090, doi:10.21437/Interspeech.2018-1929.
[64] K. Chumachenko, A. Iosifidis, and M. Gabbouj, “Self-attentionfusionforaudiovisualemotionrecognitionwithincompletedata,”IEEETrans.Multimedia,vol.25,pp.289–300,2023,doi:10.1109/ICPR56361.2022.9956592.
[65] Y. Li, M.-C. Chang, and S. Lyu, “In Ictu Oculi: Exposing AI createdfakevideosbydetectingeyeblinking,”inProc.IEEEInt.Workshop Inf.ForensicsSecur., 2018, pp. 1–7, doi: 10.1109/WIFS.2018.8630787.
[66] D.Citron,“Howdeepfakesunderminetruthandthreatendemocracy,”TEDTalk,Nov.2019.[Online].Available:https://www.ted.com/talks/daniellecitronhowdeepfakesunderminetruthandthreatendemocracy.
[67] S. Salman, J. A. Shamsi, and R. Qureshi, “Deep fake generationand detection: Issues, challenges, and solutions,” ITProf., vol. 25,no. 1, pp. 52–59, Jan.–Feb. 2023, doi: 10.1109/MITP.2022.3230353.
[68] J. Langguth, K. Pogorelov, S. Brenner, and P. Filkukova, “Don’ttrust your eyes: Image manipulation in the age of deepfakes,”Front.BigData,vol.4,pp.649989,Apr.2021,doi:10.3389/fcomm.2021.632317.
[69] S. Khan, “Adversarially robust deepfake detection via adversarialfeature similarity learning,” arXiv preprint arXiv:2403.08806, 2024,doi: 10.48550/arXiv.2403.08806.
[70] T. T. Nguyen et al., “Deep learning for deepfake detection: Asurvey,” IEEE Trans. Artif. Intell., vol. 3, no. 4, pp. 459–476, Aug.2022, doi: 10.48550/arXiv.1909.11573.
[71] V.Dudykevych,H.Mykytyn,andK.Ruda,“Theconceptofa deepfake detection system of biometric image modificationsbasedonneuralnetworks,”inProc.2022IEEE3rdKhPIWeek Adv. Technol. (KhPIWeek), Kharkiv, Ukraine, 2022, pp. 1–4, doi:10.1109/KhPIWeek57572.2022.9916378.

[72] L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy, "DeeperForensics-1.0:Alarge-scaledatasetforreal-worldfaceforgerydetection,"inProc.IEEE/CVFConf.Comput.Vis.PatternRecognit., Seattle, WA,USA, 2020, pp. 2886–2895, doi: 10.1109/CVPR42600.2020.00296.
[73] Preeti et al., "A GAN-based model of deepfake detection in socialmedia,"ProcediaComput.Sci.,vol.218,pp.2153–2162,2023,doi:10.1016/j.procs.2023.01.191.
[74] S.C.P,B.J.J.I,A.M.B,V.R,Y.R.R.V,andE.Elango,"Deepfake detection using multi-modal fusion combined withattentionmechanism,"inProc.20244thInt.Conf.SustainableExpert Syst. (ICSES), Kaski, Nepal, 2024, pp. 1194–1199, doi:10.1109/ICSES63445.2024.10763221.
[75] Y. Huang et al., "FakeLocator: Robust localization of GAN-basedface manipulations," IEEE Trans. Inf. ForensicsSecur., vol. 17, pp.2345–2356, 2022, doi: 10.1109/TIFS.2022.3182478.
[76] U.A.Ciftcietal.,"Howdotheheartsofdeepfakesbeat?Deepfakesource detection via interpreting residuals with biological signals,"in Proc.IEEEInt.JointConf.Biom.(IJCB), 2020, pp. 1–10, doi: 10.48550/arXiv.2008.11363.
[77] M. A. Younus and T. M. Hasan, "Effective and fast deepfakedetectionmethodbasedonHaarwavelettransform,"inProc.2020 Int.Conf.Comput.Sci.Softw.Eng.(CSASE),Duhok,Iraq,2020,pp.186–190, doi: 10.1109/CSASE48920.2020.9142077.
[78] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, "Deepfakedetection:Asystematicliteraturereview,"IEEEAccess,vol.10,pp.25494–25513, 2022, doi: 10.1109/ACCESS.2022.3154404.
[79] V.S.Anandhasivam,A.K.Anusri,M.Logeshwar,and R. Gopinath, "Enhancing deepfake detection through hybridMobileNet-LSTMmodelwithreal-timeimageandvideoanal-ysis,"inProc.20244thInt.Conf.UbiquitousComput.Intell.Inf.Syst. (ICUIS), Gobichettipalayam, India, 2024, pp. 1989–1995, doi:10.1109/ICUIS64676.2024.10867.
[80] S.D,R.S,S.Ravi,V.M,andP.M.P.U,"AlightweightCNNfor efficient deepfake detection of low-resolution images in fre-quency domain," in Proc. 2024 Second Int. Conf. Emerging Trends Inf. Technol. Eng. (ICETITE), Vellore, India, 2024, pp. 1–6, doi:10.1109/ic-ETITE58242.2024.10493406.
[81] B. Cavia, E. Horwitz, T. Reiss, and Y. Hoshen, "Real-time deepfakedetection in the real-world," arXiv preprint arXiv:2406.09398, 2024,doi:10.48550/arXiv.2406.09398.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY