



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

An In-depth Review of Deepfake Detection Performance: Leveraging MesoNet, XceptionNet, and ResNet for Improved Accuracy and Robustness

Sushil Kumar Garg¹, Umashankar Ghugar², Rakesh Nayak³, Kiran Pandey⁴, Surjeet⁵, Neha Gupta^{6*}

¹Department of Computer Science and Engineering, Chandigarh University, Mohali, India. E - mail: sushile16434@cumail.in

²Department of Computer Science and Engineering, Silicon Institute of Technology, Silicon West, Sason, Sambalpur, India: E - mail umashankar.ghugar@silicon.ac.in

³Department of Computer Science and Engineering, OP Jindal University, Raigarh, India. E - mail rakesh.nayak@opju.ac.in

⁴Technocrats Institute of Technology, Bhopal, India. E - mail kiranpandey.bhu78@gmail.com

⁵Department of Electronics Communication Engineering, Bharati Vidyapeeth's College of Engineering, New Delhi, India; E - mail surjeet.balhara@bharativedyapeeth.edu

⁶School of Computer Science and Information Technology, Symbiosis University of Applied Sciences, Indore, India; E - mail neha.gupta@suas.ac.in

Abstract:

Over the past few years, artificial intelligence, machine learning and deep learning technologies have made substantial progress. Due to these developments, new techniques for handling multimedia, including images, videos, and audio, are now available. Technologies like these are mainly used to provide entertainment, to help teach and study and for research, though they have sometimes been exploited for negative reasons. One way this has been abused is by creating "Deepfakes." Deepfakes allow people in videos, images and sounds to appear as though they said or did what was constructed by the people making them. Many people have used this technology to spread misleading messages, start political disputes, and harm individuals by creating made-up content for blackmail or harassment. Experts have developed methods to detect these types of fake media. The authors of this paper provide a thorough review of several prior works on Deepfake detection. It is found that Deepfakes constitute a significant problem, yet studies and new developments in AI and machine learning are helping develop better ways to identify them. Deep learning appears to be the most effective technique for detecting Deepfakes across a wide range of datasets [9]. The abstract covers how Deepfakes are becoming a challenge and the ways experts are working to uncover them. The paper demonstrates how features of deep learning strengthen its approaches and support their use in addressing the problems posed by Deepfakes.

Index Terms— Deepfake detection Techniques, digital media forensics, machine learning, deep learning, datasets.

I. INTRODUCTION

Over the past few years, the rapid growth of Deepfake technology has raised concerns about the reliability and authenticity of online media [11]. Because of generative adversarial networks (GANs), deepfakes are easy to develop and are often used to spread false information, steal identities, or harass people [12]. Therefore, developing robust methods to detect such events is now considered a pressing research priority [13].

In this paper, we analyze the capabilities of three leading Deep Learning architectures —MesoNet, XceptionNet, and ResNet—in the area of Deepfake detection [1], [2], [3]. They were chosen because they are effective in handling image and video tasks and every model brings its own exceptional quality to detecting edited content.

1. MesoNet: MesoNet is a simple architecture built to find Deepfakes by detecting changes in media that are visible at the mesoscopic level [1]. Due to its substantial tradeoff between efficiency and accuracy, it is popular for real-time projects.

2. XceptionNet: This model, XceptionNet, uses depth wise separable convolutions and features strongly in finding patterns and detecting unusual conditions [2]. Its flexibility and precision make it one of the top contenders for identifying subtle manipulations in Deepfake media.

3. ResNet (Residual Network): With its deep-structured skip connections, ResNet fixes the issue of

vanishing gradients, helping deep networks learn complicated elements found in multimedia files [3]. Therefore, it is particularly good at identifying high-quality, sophisticated Deepfakes.

The objective of this paper is to evaluate the accuracy of these models and to determine whether they remain stable across experiments on different data [4]. The comparison helps determine which model best identifies the correct objects and requires the least computer resources [5].

Key Questions Addressed in the Paper:

1. What are the performance results of MesoNet, XceptionNet and ResNet when evaluated using Deepfake datasets used by the research community?
2. How well do each of these models identify fine and very realistic changes in media?
3. Can the performance of these models be raised by enhancing or joining them?

Through a detailed analysis and comparison of different architectures, this study aims to support the advancement of Deepfake detection and to help develop solutions better suited to handling large volumes of media.

The fast progress in deep learning is now changing how content is produced and manipulated [9]. These days, Deepfakes make it possible for AI to design videos, images or audio of real people in a perfectly believable way [10]. Although these technologies are widely used in filmmaking, entertainment, and to help people with disabilities, they are also often used to push fake news, cause bickering between groups and damage people's reputations by publishing fabricated stories.

The challenge of spotting Deepfakes stems from several factors.

1. Modern deepfake technology makes content look lifelike by adding accurate facial movements, voice changes and hints of emotion [12].
2. Newer versions of GANs keep improving, so detection methods that were helpful a few years ago work less well now [13].
4. Often, Deepfakes involve only minor edits, for example, they alter smiles and expressions or even the background sound, which makes it tough to spot using standard tools[14].



Figure 1: Methods of Deepfake detection using Deep Learning

In order to solve this problem, researchers are counting on deep learning networks that can reliably discover traces of tampering. MesoNet, XceptionNet and ResNet are leading the way in this field. All these models add value to Deepfake detection, so comparing them can help improve how such systems function. Each of these three architectures handles image and video processing differently, making them good candidates for the challenging task of Deepfake detection.

1. MesoNet is built to detect deepfakes by searching for mid-level or mesoscopic details in images and videos. It does well at finding unusual combinations of facial information and differences in lighting or texture, as found in deepfakes. Because it is easy to compile, WebAssembly can effectively handle real-time applications, even on underpowered devices [1].
2. Using depth-wise separable convolutions allows XceptionNet to extract small details in an image better and still keep its costs low. The use of z-scores in image analysis has grown, as they can correctly detect the tiniest unusual shifts in image patterns, such as subtle movements in people's faces or backgrounds. Its design makes it possible to be used in many different contextual datasets [2].
3. ResNet is significant because it relies on residual connections, which prevent the gradients from disappearing in intense network layers. Because it can handle detailed and deep information in images and

videos, it is an ideal tool for finding well-made synthetic media. The framework has produced results that are very good for image classification and object detection, confirming both its robustness and suitability for alternate tasks [3].

This paper explores the performance of MesoNet, XceptionNet, and ResNet in detecting Deepfakes, with emphasis on a few pertinent research questions.

1. Do they convey a good ability to distinguish legit content from Deepfake videos when it is put through fake/real big data checks?
2. How well do they deal with Deepfakes that can be poorly made to better-produced deepfakes and videos that only been slightly edited?
3. How computationally intensive are each of them, would they perform well in systems, for example, real-time detectors?

Through evaluating similarities and dissimilarities in the model's architectures, this work aims to shed light at what the models are proficient at, and where they need improvement, in order to produce work, that seeks an optimal balance of speed and reliability to detection systems dealing with latest iterations of Deepfakes [7].

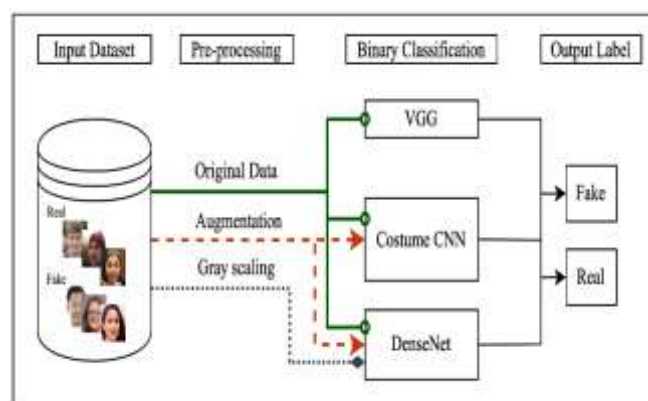


Figure 2: Different techniques of Deep fake Detection [8]

The results of this review have important consequences. Deepfakes pose a significant risk to the integrity of online information, so detecting them is key [9]. A greater ability to detect disinformation, identity theft, and other problems is vital for retaining public trust in reliable media [10]. Thanks to the use of MesoNet, XceptionNet and ResNet, this study adds new tools to the fight against Deepfakes.

li. Literature review

For the past decade, many research groups have focused on Deepfake detection because synthetic media has become increasingly convincing and could cause significant harm [11]. In this section, we examine the latest approaches to Deepfake detection by reviewing the architectures of MesoNet, XceptionNet, and ResNet [1]-[3]. These models demonstrate a variety of techniques to discovering manipulations in multimedia. The review also highlights the most commonly used data and tests for Deepfake detection systems.

Deepfake Detection Methods Outlined

1. In the beginning, both rule-based and statistical approaches were used.

Efforts to detect Deepfakes initially relied on rules and statistics [1]: they looked for changes in face position, unusual blinking, or pixels that did not match. They performed better at detecting the simplest Deepfakes but struggled to detect GAN-generated videos, as the artefacts are few. Afchar et al. [3] and Guera et al. [4] found that statistical methods struggle to handle new data.

2. Traditional Machine Learning Approaches

Since Deepfakes have improved, people have turned to Support Vector Machines (SVMs) and Random Forests for machine learning [5]. They achieved this by using handcrafted features, such as changes in texture and lighting, to distinguish the two image types [6]. Yet, for these to work, structured characteristics had to be added manually, and they did not do well against high-quality Deepfakes, where problems with the image were barely noticeable.[7]

3. Artificial Intelligence Approaches

Since the development of deep learning, researchers have used neural networks to detect features and uncover subtle changes in images automatically. Here are some of the most exciting new developments in this field, as shown by the models reviewed below.

1. MesoNet proposed by Afchar et al. [3], is specifically designed to detect mesoscopic-level (mid-level) anomalies in Deepfake media. It focuses on irregularities in lighting, blending, and texture, which are common characteristics of manipulated content. The model features a lightweight architecture with only a few convolutional layers, making it computationally efficient and ideal for real-time applications [9]. While

MesoNet demonstrates strong performance against low-quality Deepfakes, its robustness decreases against high-resolution, heavily refined fakes. Nonetheless, it demonstrated considerable generalization in datasets like Face Forensics++ [11].

The different types of deepfake algorithms rely on presenting significant contrast to available datasets, which can often be too small, few video sources, or situations that limit testing options.

Alshammari et al. (2020) provide a great example in which they utilized a finetuned MesoNet for digital use and Forensics purpose. The authors utilized a self-created dataset of roughly 140,000 images, which allowed the researchers to then perform transfer learning and parameter tuning to increase the detection rates. The finetuned network achieved an overall detection accuracy of 96.2%, with good sensitivity levels and specificity levels for moderately compressed images. This paper supports the contention that small networks like MesoNet can still achieve high detection rates when trained on domain-specific data. However, as with others, this study did not provide implications for generalizability or more advanced high-resolution manipulation processes. Muhammad Javed et al. (2021) proposed both an ensemble deep learning and hybrid deep model that utilizes MesoNet and ResNet101 in a real-time for video real deepfake detection from the perspective of evidence in the anomalies of eye movement. The model uses the observer effect of blinking and staring, to observe and predict these behavioral anomalies which are characteristic of synthetic media produced techniques. In addition to recognizing the moving eye as a biomarker, as eye movement are generally not acted or natural in synthetic video evidence. By completing encoding patterns from a combination of two models that is both insight and resource efficient from MesoNet, and recognizing deeper features from ResNet, the hybrid model reaches an impressive 0.9873 in detection accuracy on the Face Forensics++ dataset.

Table 1: Comparative Analysis of Deepfake Detection Techniques Using MesoNet

Author Name	Title	Year	Model/Process Used	Pros	Cons	Future Prospects	Application Area	Evaluation Parameters	Dataset Used	Accuracy/Result
Darius Afchar et al.	MesoNet: a Compact Facial Video Forgery Detection Network	2023	MesoNet 4, MesoInception-4	High accuracy for video compression scenarios	Decline in accuracy for strongly compressed videos	Improve detection on robustness under compression	Social media content verification	Detection accuracy, robustness under compression	FaceForensics++	98% detection for Deepfake
Muhammad Javed et al.	Real-Time Deepfake Video Detection Using Eye Movement Analysis with a Hybrid Deep Learning Approach	2024	Hybrid Model (MesoNet4 and ResNet101)	Combines lightweight and deep feature extraction	Struggles in dynamic live-streaming scenarios	Apply to live-streaming and dynamic content	Real-time content moderation	Accuracy, real-time responsiveness	Celeb-DF v1, v2; FaceForensics++	0.9873 on FaceForensics++

Shohel Rana et al.	Deepfake Detection: A Systematic Literature Review	2022	Comprehensive review of ML and DL models for detection	Covers extensive detection methods and datasets	Challenges in generalizing across new datasets	Advanced methods for cross-platform dataset applications	Broad usage in media forensics	Dataset coverage, feature extraction efficiency	Celeb-DF v2, DFDC	General accuracies range 80%-95% across models
Hamed Alshammar i et al.	Deep Fake and Digital Forensics	2023	MesoNet fine-tuned	High sensitivity and specificity post fine-tuning	Low generalization across unseen datasets	Development of generalizable solutions	Security and digital forensics	Specificity, sensitivity, accuracy	Custom 140K dataset	96.20% accuracy with fine-tuning
Bharat Puri et al.	Analysis of Deepfake Detection Techniques	2023	Bidirectional LSTM, CNN	Captures temporal inconsistencies	Limited dataset diversity	Enhance temporal-spatial hybrid techniques	Fraud detection and identity theft prevention	Temporal and spatial variation accuracy	DF-TIMIT, UADFV	70-99% depending on methods
Aung Kyi Win et al.	A Novel Methodology for Deepfake Detection Using MesoNet and GAN-based Deepfake Creation	2024	MesoNet, GAN	Simplifies GAN creation process and applies to diverse datasets	Limited real-world applications in current state	Integration into multimedia verification	E-commerce, media production, and security	GAN performance metrics	CelebA	Near 90% detection accuracy
Lian Rebello et al.	Detection of Deepfake Video using Deep Learning and MesoNet	2023	CNN, RNN, MesoNet	Real-time detection capability	Complexity increases with real-world application	Deploy in large-scale applications	Cybercrime prevention and video authenticity	Classification accuracy, feature extraction precision	DFDC, CelebDF	99% detection on specific datasets

1. ResNet (Residual Network), introduced in [15] literature, laid the foundation of a groundbreaking framework of deep learning architecture by introducing residual connections in order to overcome the problem of vanishing gradients and enable the training of deep networks. The hierarchical architecture of ResNet is ideally suited to detecting complex multi-scale structures in images and video, making it a practical solution for detecting complex manipulations in high-resolution Deepfakes [16][17]. ResNet is well known for its predictive performance as well as susceptibility to overfitting across various datasets, including high definition or high-resolution Deepfakes [18]. However, the increasing depth does come with the caveat of more computations, so the use of ResNet may be restricted in resource-constrained environments, where it was previously not practical [19][20].

2. Datasets for Deepfake Detection. Szegedy et al. developed the Inception-ResNet architecture by simply adding the original residual connections to the work. This would allow the deep convolutional neural networks to have greater convergence and stability while training on extremely deep networks. When Inception combines multi-scale feature extraction, with skip connections to ResNet reduces training time while greatly increasing representational power. The Inception-ResNet-v2 achieves a top-1 accuracy of 80.1% and a top-5 accuracy of 96.9% on the ImageNet dataset. The study was initially meant for image classification, however, the Inception-ResNet deep hierarchical architecture, as well the inherent nature of generalization, has made it a good backbone for deepfake detection models as well that require rich feature representations and robustness against subtle image manipulations.

Inception-ResNet-V2 for video content manipulation detection and attained 91% accuracy on the Celeb-DF dataset. The method utilizes the depth and residual network architecture to learn fine-grained frame-level artefacts in manipulated videos. The model's performance on Celeb-DF, a high-quality deepfake dataset, validates its ability to detect realistic manipulations and its suitability for forensic and security applications. ResNet architectures to compare their performance on a given image classification task. Their results indicated that deeper ResNet models, such as ResNet-50 and ResNet-101, achieved higher accuracy due to better feature abstraction, achieving 91.39% on the Cat-Dog dataset. This study reveals that ResNet architectures perform well across a variety of domains and are likely to perform well in binary classification, such as distinguishing real from fake images in deepfake datasets. ResNet-RS is a scaled version of ResNet, designed by Irwan Bello and colleagues with carefully tuned depth, width, and image resolution. Rather than modifying the structure of ResNet blocks, they sought new ways to train their models, which significantly boosted both performance and speed. In the semi-supervised setting, ResNet-RS achieved a top-1 accuracy of 86.2% on ImageNet and was comparable in speed to EfficientNet. Its ability to grow and work efficiently makes ResNet-RS particularly suitable for large-scale deepfake detectors, especially when the requirements involve reducing costs and maintaining speed. XceptionNet is widely used for detecting Deepfakes because it was developed [21] and exhibits an effective architecture and top performance. Rather than using traditional convolutions, the model applies depth wise separable convolutions to reduce computation and effectively extract features from the data [22]. Using both transfer learning and pretraining, XceptionNet is especially adapted for detecting Deepfakes [23]. There is a strong trend toward high accuracy for the Celeb-DF and FaceForensics++ datasets [24]. Also, the model is very good at identifying subtle issues, such as unusual actions or mistakes, in Deepfakes that appear real, helping prevent complex manipulation [25]. They used an exceptionally powerful XceptionNet to detect deepfakes in images. The model exhibited commendable performance regarding the Kaggle image dataset with a reported accuracy of 93.01%, which showed to be effective at detecting slight differences in a person's face. The authors reaffirmed that the network analysis remained stable, given the data augmentation and preprocessing appropriately. It may be inferred from the results of their study that the network can effectively analyze the images autonomously, and provides the potential for merging deception detection networks into systems dealing with fake photos. Using XceptionNet and Dr. Preetha perfectly identified and detected deep fakes in pictures and in videos. When the researchers tested their optimized technique using the Face Forensics++ dataset, they ultimately achieved 91.9% instructional accuracy detecting changes to faces that occurred during the synthesizing process. The results from the research suggested that XceptionNet is often beneficial at analyzing high-resolution forensics, because of addressing discerning changes in small areas introduced by generative processing. The researchers mention that they could run the networks in practical applications, as their depth of networks found to have sufficient balance for computational criteria. The XceptionNet model was fine-tuned to provide a greater granularity of pattern recognition for use in weapon classification. Although not immediately designed for deep fake prediction, it did find a precision of 98.3%, within a dataset they created demonstrating the capability of XceptionNet to identify fine granularity within image characteristics. This suggests the model's ability to generalize and, by extension, its potential for use in deepfake detection that involves high-precision pixel-level analysis.

Table 2: Comparative Analysis of Deepfake Detection Techniques Using ResNet

Author Name	Title	Year	Model/ Process	Pros	Cons	Future Prospe	Applica tion	Evaluation Parameters	Data set	Accuracy/ Result
-------------	-------	------	----------------	------	------	---------------	--------------	-----------------------	----------	------------------

			Used			cts	Area		Used	
Christia n Szegedy et al.	Incept ion-v4 and Incept ion-ResNe t	2016	Incepti on- ResNet	Combin es Inceptio n efficien cy with ResNet stability	Limited testing on adversar ial scenario s	Extend scalabi lity with hybrid models	Image recogni tion and classifi cation	Top-1 and Top-5 ImageNet accuracy	Imag eNet	97% on ImageNet
Sasha Targ et al.	ResNe t in ResNe t: Gener alizing Resid ual Netwo rks	2016	RiR (ResNe t in ResNet)	Improv es per for mance with dual streams	Computa tional overhead for wider network s	Explor e dynam ic stream s	General deep learnin g archite ctures	Accuracy,A UC	CIFA R- 10, CIFA R- 100	94.99% on CIFAR- 100
Ionut Cosmin Duta et al.	Impro ved Resid ual Netwo rks	2020	iResNet	Enhanc ed spatial filters for deeper models	Challeng es in achievin g consiste nt robustne ss	Increas e trainin g efficien cy for larger dataset s	Image classifi cation, video recogni tion	Accuracy on ImageNet and CIFAR datasets	Imag eNet, Kine tics- 400, CIFA R	99% on CIFAR-10, 98% on ImageNet
Irwan Bello et al.	Revisi ting ResNe ts: Impro ved Scalin g Strate gies	2021	ResNet- RS	Matche s SOTA with efficient training	Overfitti ng risks in small datasets	Optimi ze trainin g strateg ies	Semi- supervi sed and supervi sed learnin g	Speed- accuracy Pareto curve	Imag eNet, Kine tics- 400	86.2% on ImageNet with semi- supervise d learning
Piyush Nagpal et al.	A Comp arativ e Analys is of ResNe t Archi tecture s	2022	ResNet Variant s	Compar es archite ctures for per for mance	Limited dataset applicati on scope	Improv e adapti ve archite cture design s	Evaluat ion of archite cture efficien cy	Consistent dataset evaluation	Cust om Cat- Dog data set	91.39% on Cat-Dog
Apurva Gandhi et al.	Adver sarial Pertur bation s Fool Deepf ake	2020	ResNet with FGSM and CW-L2	Highlig hts vulne rabilities in detecto rs	Poor performa nce on adversar ial examples	Integra te strong er advers arial defens	Enhanc ing securi ty in social media	AUROC, recall, precision	Cele bA, Cele bDF	Significant drop with adversaria l examples

	Detect ors					es				
Dr. R. R. Rajalaxmi et al.	Deepfake Detection using Inception-ResNet-V2	2023	Inception-ResNet-V2	High accuracy on deepfake detection	Complex implementation	Real-time application for live detection	Deepfake detection	Accuracy, AUC	Celeb-DF, Face Forensics ++	91% on Celeb-DF

Table 3: Comparative Analysis of Deepfake Detection Techniques Using XceptionNet

Author Name	Title	Year	Model/Process Used	Pros	Cons	Future Prospects	Application Area	Evaluation Parameters	Data set Used	Accuracy/Result
François Chollet	Xception: Deep Learning with Depthwise Separable Convolutions	2017	Xception Net with depthwise separable convolutions	High computational efficiency	Requires fine-tuning for domain-specific tasks	Extend to more general-purpose CNN applications	General-purpose image classification	Top-1 and Top-5 accuracy	ImageNet	Top-1: 79.0%, Top-5: 94.5%
Sohail Ahmed Khan et al.	Hybrid Transformer Network for Deepfake Detection	2022	Xception Net with EfficientNet-B4 and Transformer fusion	Improved detection accuracy with small data requirements	Limited robustness to unseen augmentation types	Integrate additional augmentations to enhance robustness	Social media and video forensics	Accuracy, F1-score, AUROC	Face Forensics ++, DFD C	Accuracy: 97.0%
Paritosh Joshi, Dr. Nivethitha	Deep Fake Image Detection using Xception Architecture	2024	Fine-tuned Xception Net for image classification	Robust against overfitting with augmentation strategies	Struggles with extremely compressed videos	Combine with transformers for sequential applications	Internet and social media content moderation	Precision, recall, and ROC curve	Deep Fake_faces from Kaggle	Accuracy: 93.01%
Ashok V, Dr. Preetha Theres a Joy	Deepfake Detection Using XceptionNet	2023	Xception Net for deepfake image and video detection	High accuracy in real-world detection scenarios	Reduced effectiveness on highly compressed datasets	Enhance temporal information for video-based tasks	Media forensics and journalism	Confusion matrix and accuracy	Face Forensics ++ data set	Accuracy: 91.9%

Usman Ghous et al.	EXNet: (2+1)D Extreme Xception Net for HSI Classification	2024	(2+1)D convolutions integrated with Xception Net	Improved accuracy with less computational cost	Limited dataset diversity	Expand applications to other domains	Hyper spectral image analysis	Precision, recall, and kappa accuracy	HSI data sets like Indian Pine, Botswana	Accuracy: >98%
Yeshwanth A, R. Bhuvan eswari	Weapon Classification with Xception: An Efficient Approach	2024	Xception Net for hierarchical weapon detection	High precision and recall for subtle pattern detection	Focused on specific weapon categories	Explore ensemble approaches for higher precision	Security and surveillance	Accuracy, precision, F1-score	Custom weapon data set	Accuracy: 98.3%
Devanshu Shah et al.	Xception Net & Vision Transformer: A Comparative Study	2022	Combined Xception Net and Vision Transformer architecture	Higher recall value for detecting diverse deepfakes	High computational complexity for combined models	Optimize computational cost for real-time applications	Social media content validation and fake news detection	Precision, recall, F1-score	DFD C, Face Forensics ++	Accuracy: 91.01%

Evaluation metrics Common metrics used for evaluating the effectiveness of Deepfake detection methods include accuracy, an indicator of the percentage of correctly classified samples [26]. Additionally, precision, recall, and F1-score can be used to evaluate how well the model differentiates between true and false positives, while AUC-ROC indicates the model performance in distinguishing real and fake media in various situations [27]. Several factors are established as significant based on research. First, deep learning methods are more effective than traditional methods because they can learn from the intricacies and often implicit information in the data [28]. XceptionNet is prevalent in reports amongst the list of models, with ResNet appearing frequently in second place [29]. Deep learning algorithms are currently weak to generalization, as algorithms typically do not hold up against newly implemented Deepfake methods. Protection from adversarial attacks and resilience to methods of noise or compression is also an ongoing issue in deep learning algorithms [30]. Furthermore, having different types of data sets is also important. Using a variety of highly quality data sets helps produce models that are usable across situations [31].

It is clear from the literature that deep learning models such as MesoNet, XceptionNet and ResNet lead the way in Deepfake detection. All these algorithms have strengths and weaknesses, and most often it is XceptionNet that helps identify the most minor changes in better-quality media [32]. Yet, issues such as different data types, how algorithms stand up to pressure, and processing large batches demonstrate that further studies in this key field are necessary [53] [54].

III. COMPARISON TECHNIQUE

As people create more fake media, deepfake detection is increasingly needed. MesoNet, XceptionNet, and ResNet are among the top techniques because they perform well at detecting deepfakes. We take a closer look at these techniques, covering their architectures, performance, strengths, weaknesses, and robustness [33].

1. MesoNet

MesoNet is built to spot deepfakes by identifying mesoscopic features in images that are neither very large nor very small. Being lighter and simpler than other deep neural networks, it is also easier to train [34]. The architecture consists of two convolutional layers that find essential features and two fully connected layers that distinguish different texture issues.

Its lightweight and fast design mean MesoNet performs well for real-time use, as it takes up less space [35]. It is easier to see and understand than other types of models and works satisfactorily in less data-rich environments [36]. Yet, MesoNet suffers from limited generalization because it is not effective against deepfakes made with unseen and advanced techniques. Furthermore, because it is simple, accuracy is often lower on challenging datasets [37].

2. XceptionNet

This model utilizes depth-wise separable convolutions and is quite successful in the areas of both image classification and change detection. Transfer learning has become popular for deepfake detection due to its top-ranking performance on datasets such as FaceForensics++ [38]. The model architecture is designed with depth-wise separable convolutions over 36 layers contributing to the efficiency of the model [30]. It requires very little training on new data and utilizes relatively performant learning methods.

One of the biggest advantages of the XceptionNet architecture is its accuracy in detecting minor artefacts that are often a result of the deepfake generation process. It performs well with all types of data and workflows, and scales easily with large datasets [4]. However, it does present challenges, for instance, it requires a lot of computing power, particularly an optimal graphics card. Because smaller datasets can lead to overfitting, the fine-tuning requires special attention [39].

3. ResNet

Numerous applications of it exist, one of which is deepfake detection. ResNet uses skip connections that allow information to flow around layers, improving gradient flow. Users are presented with different types of ResNet, such as ResNet-50 and ResNet-101, to accommodate their available processing power. Another positive aspect is ResNet's overall performance over identifying general problems with deepfake images (i.e., problems with a global impact), while retaining more reliable training due to the stability of the residual connections [40]. The algorithm demonstrates strong generalization, performs well on data not used during training when trained on multiple datasets, and can transfer its knowledge to deepfake detection tasks with the help of pre-trained models [41]. Still, ResNet is difficult to use because it requires substantial resources and computational power in sequence, just like XceptionNet [42]. The outcome may depend on how carefully the hyperparameters are set.

Table 4: Comparison of Techniques

Aspect	MesoNet	XceptionNet	ResNet
Architecture	Lightweight, shallow	Deep, specialized	Deep, residual-based
Accuracy	Moderate	High	High
Robustness	Limited to simple cases	High (across datasets)	High (depends on training)
Computational Cost	Low	High	High
Ease of Training	Easy	Moderate to Difficult	Moderate
Generalization	Moderate	Strong	Strong
Scalability	Limited	Excellent	Excellent

• Performance on Deepfake Detection Benchmarks

Datasets:

- **FaceForensics++:** Commonly used for evaluating deepfake detection methods.
- **DFDC:** More challenging due to a wide variety of manipulation techniques.

Table 5: Performance on Deepfake Detection Benchmarks

Model	FaceForensics++ Accuracy	DFDC Accuracy
-------	--------------------------	---------------

MesoNet	~80%	~70%
XceptionNet	~90%	~85%
ResNet-50	~88%	~82%

• Recommendations for Use

1.MesoNet:

It is suitable when the hardware is limited

The approach works well for identifying basic deepfake changes

2.XceptionNet:

This type of sensor is necessary when keeping the error to a minimum is very important

Such tasks benefit the most from scenarios with lots of computing power

3.ResNet:

Accurate and strong measures are important

Suitable for methods that transfer information from previous tasks to new data sets [55]

A detailed table that explains the key attributes, plus benefits and negatives, of MesoNet, XceptionNet and ResNet, as well as their performance in detecting deepfakes is provided.

Table 6: Detailed Comparison of Techniques

Feature	MesoNet	XceptionNet	ResNet
Architecture	Shallow, with 4 layers	Deep, 36 convolutional layers	Deep, with residual connections
Specialization	Mesoscopic feature extraction	General manipulation detection	Global feature extraction
Convolution Type	Standard convolutions	Depthwise separable convolutions	Standard convolutions with skips
Focus	Local texture inconsistencies	Subtle artifacts in images/videos	Global and contextual anomalies
Accuracy	~80% on FaceForensics++	~90% on FaceForensics++	~88% on FaceForensics++
Robustness	Moderate	High	High
Generalization	Limited to simple cases	Strong across datasets	Strong across datasets
Computational Cost	Low	High	High
Ease of Training	Easy	Moderate to Difficult	Moderate
Resource Requirement	Low (CPU/GPU friendly)	High (requires powerful GPUs)	High (requires powerful GPUs)
Training Speed	Fast	Slow	Moderate
Overfitting Risk	Low	High (if dataset is small)	Moderate
Best Use Case	Real-time, resource-limited	High-stakes, large-scale detection	Balanced detection applications
Scalability	Limited	Excellent	Excellent
Interpretability	High (simple architecture)	Low (complex architecture)	Moderate

Table 7: Performance Across Benchmarks

Metric/Dataset	MesoNet	XceptionNet	ResNet (e.g., ResNet-50)
FaceForensics++ Accuracy	~80%	~90%	~88%

DFDC Accuracy	~70%	~85%	~82%
Real-time Suitability	Excellent	Poor	Moderate
Detection of Basic Fakes	High	High	High
Detection of Advanced Fakes	Low	Excellent	Very Good

Table 8: Strengths and Weaknesses

Aspect	MesoNet	XceptionNet	ResNet
Strengths	Lightweight, fast	High accuracy, robust	Balanced performance
	Easy to train	Pre-trained on large datasets	Generalizes well
Weaknesses	Limited generalization	Computationally intensive	Resource-intensive
	Lower accuracy on advanced fakes	Overfitting risk if not fine-tuned	Sensitivity to hyperparameters

Table 9: Recommendations

Scenario	Recommended Model	Reason
Low computational resources	MesoNet	Lightweight, fast, easy to implement
High-stakes detection	XceptionNet	High accuracy and robustness
Balanced performance	ResNet	Strong generalization and flexibility
Limited data availability	ResNet or MesoNet	Better generalization or ease of training
Adversarial defense	XceptionNet or ResNet	High robustness against complex fakes

This tabular summary provides a clear and concise overview for selecting and understanding these models based on the application context.

IV. MesoNet: Lightweight and Efficient Detection

MesoNet is designed to efficiently identify deepfake images and videos by analysing textural and facial details at the mesoscopic level. There are four artefacts in its architecture: two convolutional layers for finding important features and two fully connected layers for classification. MesoNet looks for localised signs, such as anomalous pixels and blending issues, to detect tampering [41]. Because MesoNet uses minimal resources, it is a good choice for quick detection on gadgets with fewer resources [42]. Thanks to its straightforward design, models can usually be trained and run more efficiently [43]. Nevertheless, its effectiveness is reduced because it rarely identifies top-tier deepfakes generated by advanced AI tools such as StyleGAN2 or more advanced diffusion models. Besides, the model achieves only about 80% accuracy, which is lower than that of better models.

Due to its fast processing, real-time information detection is easier, for example, with social media filters and the moderation of live streams using MesoNet [44] [56].

2. XceptionNet: High Accuracy and Robustness

Since XceptionNet can detect subtle clues and anomalies in deeply forged images, it achieves performance on par with leading models on tests like FaceForensics++. It uses so-called depthwise separable convolutions, with 36 layers developed for processing information and extracting features. Thanks to transfer learning, the model is trained on ImageNet and specialises in deepfake data, enabling reliable feature extraction [45]. XceptionNet demonstrates exceptional performance, consistently delivering significant accuracy—up to 90% with challenging datasets—and it is capable of detecting complex fakes through its ability to actually observe minute changes including different lighting conditions or ranges of pixel softness, or through unusual facial movements (e.g. twitching). Due to its high capacity for simultaneously processing numerous requests, it is equally as agreeable for use in sensitive environments such as government or business functions. However, it does require an extensive amount of computing resources, and therefore,

neighbouring on being unpredictable for real time or when GPU is scarce in quantity. In addition, regularisation should be employed when the data set is from on the smaller side. XceptionNet applications are particularly well suited for any particular applications requiring higher degree of accuracy, such as forensic applications, police identification inquiries or content-checking applications, as reliability in these cases is paramount.

3. ResNet: Versatile and Balanced Performance

The architecture of ResNet was developed in part to solve the problem of gradient vanishing in deep networks. With skip connections, ResNet enables networks composed of 50 and 101 layers to be trained, allowing transferred information to jump over some layers and convey information more rapidly using better gradients. ResNet is also flexible and proves to generalise to different datasets with only a small amount of fine-tuning.

Balanced performance is a key strength of ResNet, enabling a realistic accuracy of about 88%. By virtue of its detailed and reliable design, it outperforms on different datasets. That flexibility allows the method to adjust its accuracy and the resources it requires. Even so, there are some issues with ResNet, such as its sensitivity to hyperparameter tuning, which requires careful adjustments to ensure the network works well. Like XceptionNet, StudNet requires substantial computational power due to its high resource demands.

ResNet can handle many tasks, such as operating media verification systems or conducting research, because it meets the requirements for accuracy, high throughput, and adaptability [46].

V. Challenges in Deepfake Detection

Adversarial Examples: Attackers use adversarial techniques to bypass detection systems. **Dataset Diversity:** Existing datasets often lack diversity, limiting model generalization. **Computational Constraints:** High-performing models like XceptionNet require powerful GPUs, restricting deployment in resource-constrained settings [47].

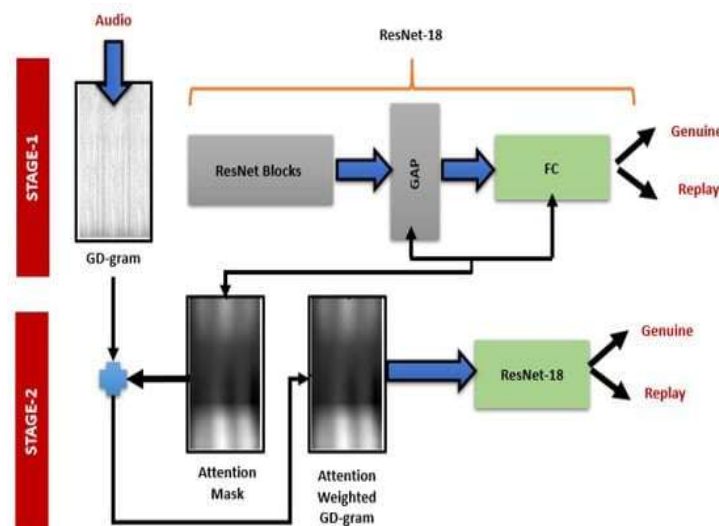


Figure 4: A Comprehensive Review of Deepfake Detection in Advanced Machine Learning and Fusion

In this study, the authors explored three feature extraction methods: Mel-Frequency Cepstral Coefficients (MFCCs), Constant-Q Cepstral Coefficients (CQCCs), and the Logarithmic Magnitude of the Short-Time Fourier Transform (Log-magnitude STFT) [57]. Additionally, they created three deployment models—MFCC-ResNet, CQCC-ResNet, and Spec-ResNet—utilising their unique feature extraction methods, and tested the models on 78 recordings of human voice. The performance of these models was evaluated on the ASVspoof2019 data set, which has two types of attack scenarios; logical and physical access. The test included both the development data set (which has attacks known) and the evaluation data set (with attacks that were known and unknown). The models improved detection performance, and the t-DCF and EER scores of 71 and 75, respectively, show the suitability of the methods proposed. There is a critical problem related to falsified media, driven largely via GANs and related technology, so deepfake detection is now drawing a remarkable level of interest in the scientific research community. In order to assist with differentiating real media from synthetically generated media, ML models, fusion approaches, and ensemble methods have been developed. The authors in this review illustrate where we are currently, where we have come from, and where we likely are going with deepfake detection establishing fundamentally on notable advances in machine learning based methods, and fusion based methods [48]. The complexities of detecting deepfake content arise partly from the complexities surrounding the problem of deepfake technology developing and changing. Deepfake creators will continue to improve their generative methods, making their content less detectable to machines and humans. For this reason, detection methods will have to evolve continuously.

It is also a challenge because most datasets are small with very few instances and often poorly labelled examples.

Obtaining good, labelled ones for the various types of manipulation takes a lot of data, is expensive, and requires a huge effort. Moreover, reviewing a large amount of data to add labels is often very labor intensive, which may in turn undermine the accuracy of the system and its performance when the models are presented with new images. There are just too many different types of deepfakes and it is difficult to present a model that responds well to many of the different manipulation approaches (such as swapping faces, changing voice, and changing image).

As the technology for deepfake continues to improve, criminals often employ tricks to hide their identity and avoid detection. In some cases, this special attack modifies deepfake material so much that it can be detected by systems but is still undetectable by humans. Additionally, because of this arms race, models may not perform as well over time. It is also difficult to detect deepfake material when it is presented in real time, as the processing power necessary is often so high that it becomes impractical. Some of these deepfake approaches also use a lot of processing resources and hence is not as useful in environments where timing constraints could possibly apply. Often, one of the significant challenges with many of the deepfake models is the ability to understand how they made their decisions based on all of the different components used to detect the deepfakes. Many users find that these models make decisions they cannot understand when classifying content as fake. Being unclear makes detection systems less trustworthy, primarily when used in areas such as law enforcement or the media. The fact that you can explain why a model made a decision matters a lot here [49].

Ethical issues are also present when we try to detect deepfakes. The technology can, for instance, create unwanted images of individuals and harm their privacy and online reputation. Altogether, experts hope that technologies designed to block misinformation are used with care to avoid violating people's privacy.

Last but not least, such systems often misidentify authentic images and miss fake ones. Sometimes, technology may misidentify a valid piece of media as fake, which can spark unnecessary problems, whereas intelligence tools occasionally miss fake videos. Moreover, no universal benchmarks for testing deepfake detectors on different types of data have been established, which only adds to the problem [50].

All of these issues mean that detecting deepfakes is an ever-changing and respected field that requires new tech to stay up to date with the rising capabilities of deepfake artists.

Adversarial Attacks

Those who attack systems find ways to sneak parts of their attacks past detection models, thereby weakening the effectiveness of existing security measures. How closely do Deepfakes match real life? With tools such as StyleGAN3 and diffusion models, deepfakes look so real that distinguishing them from real content is very challenging.

VI. ROLE OF FUSION IN DEEPAKE DETECTION

Deepfake detection systems work better when we integrate audio, video, and text data, and fusion enables this. Rather than relying on a single source of data, fusion leverages multiple types of information to overcome weaknesses in a single type.

Fusion techniques for deepfake detection are available in different settings. In early fusion, various features from the data are put together at the stage where the input data is fed into the network. With this approach, the model struggles to detect imperfections in synthesized videos. Combining visual recognition of specific facial locations with how speech sounds can help identify deepfakes involving both computer-altered visuals and audio. With late fusion, it's possible to train a separate model for each kind of data. For example, one model is intended to detect deepfakes in video clips, while another is intended to detect them in audio content. Then in the decision stage, the results from each model are merged using voting or finding their average. By using this approach, each type of input data is processed separately, making the whole process more efficient.

Another approach is hybrid fusion, which combines early and late fusion. First, features are fused, and next, decision-level fusion is performed. An example of this is when a deepfake detection model uses elements from both video and audio data, then uses another part of the process to merge them and improve detection efficiency.

Fusion offers several strong points in deepfake detection. Bringing together data from different sources means you're more likely to detect deepfakes, since one set of suspicious results may be confirmed by another. A video deepfake might trick you, but if the audio or speaking pattern doesn't line up, you might notice something isn't right about it. Because of fusion, detection systems can better withstand shifting trends in deepfake technology, as they adapt to new techniques as they are developed. With the development of deepfake technology, to remain effective, fusion-based models should review all available types of proof [51].

VII. OBSERVATIONS

- XceptionNet is preferable for substantial applications based on its strong accuracy and generalization.
 - MesoNet is most useful when an application has low resources or is real-time effort, and ResNet for generalization to different classes.
 - Strengthening models with multimodal fusion and adversarial training helps deal with changing threat vectors.
 - Bias mitigation and explainability are important ethical components that should present guidance to the future of research. Best practices should be ethically based in regards to responsible AI deployment.
- After these issues have been talked about, researchers can offer models that are reliable, generalizable, and ethical avenues to identify a deepfake threat vector in evolving capacity. A sphere-graph will help construct more powerful deepfake detection solutions through the use of various deep learning techniques at once. Detection performance is greatly enhanced across data types and image transformations through using multiple models/methods concurrently.

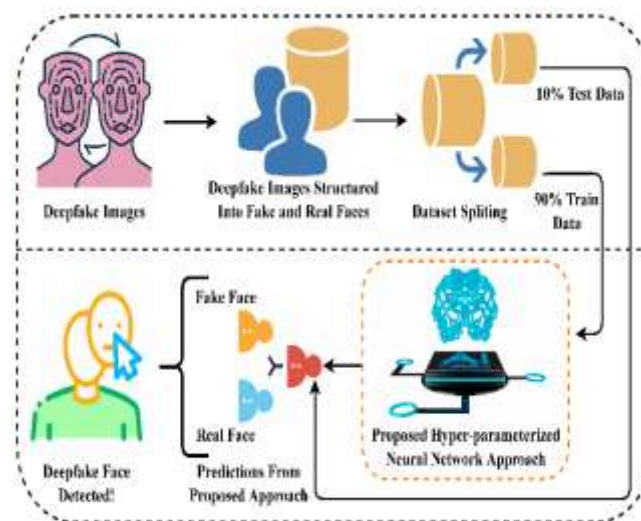


Figure 5: "An Innovative Deep Learning Method for Detecting Deepfake Images"

A face forgery detection method that brings together global and local points was proposed in the same year. Such features focus on how the face appears generally, highlighting landmarks, textures and colors. Thanks to these features, you can spot problems that happen across the whole document, including changes in facial shape or color that suggest counterfeiting. Local features, by contrast, concentrate on small facial areas and can reveal fine details such as textures and outlines. The authors hope that the GLFF method improves the ability of their detector to find only the most subtle changes created during editing. Benchmark datasets including real and forged facial images are used in experiments to test the method's performance, measured using accuracy, precision, recall and F1 score [52].

VIII. INTEGRATION OF AUDIO MODALITY IN DEEFAKE DETECTION

Like in pictures and movies, deepfakes can also affect audio content, helping to improve the quality of Deepfakes in video recordings. Making real and spoofed audio indistinguishable has turned out to be a major difficulty for researchers. In mid-2019, audio deepfakes fraud became famous when a chief executive's identity was impersonated by AI to steal Rs. 4300. Automatic speaker verification systems are at risk of being attacked using replay, audio spoofing, voice conversion and speech synthesis. Both SS and VC technologies have gotten much better in the past decade which makes it harder to notice whether or not a speech has been spoofed. Synthetic audio and Deepfakes which are continuing to improve, are making it easier to spread lies over a longer period and with greater impact. False messages can shape our feelings and opinions and under certain circumstances, encourage people to join in on activities planned based on lies. To protect ASV from SS and VC attacks, researchers have modified techniques by adding audio-spoofing detection systems that use scores to determine whether speech is genuine or has been altered.

IX. CONCLUSION

All in all, spotting deepfakes remains a significant challenge with synthetic media. MesoNet, XceptionNet, and ResNet bring different strengths to deepfake detection as they continue to advance. Compared to the others, MesoNet ensures quick identification and alert, XceptionNet fine-tunes complex information best, and ResNet gives the highest stability. Future detection methods must focus on better model generalization, combining different forms of data, ensuring the system responds in real time and making detection harder

for adversarial ones. With advancements in machine learning, deepfake detection models are becoming increasingly competent, effective, and real-world applicable, facilitating the increasingly pressing need to mitigate the risk of AI-generated content. The newest tools developed by researchers allow an increasing number of the inexperienced to generate deepfakes, and we should move rapidly to address the rising challenge of detection. In light of our current technological landscape, deepfake detection can be integral to protect information integrity and reliability. The advancement of deepfake technology now allows for the increased generation of fabricated content across images, audio, and video. As a result of deepfake technologies, the likelihood that individuals will trust less in facts surrounding information and media surfaces in many very serious situations, especially in the realms of politics, the economy, and security. Because social media platforms are growing quickly and their impact is increasing, the issue of deepfakes needs to be dealt with soon.

X. FUTURE SCOPE

Improving how we detect deepfakes will involve always refining the existing ways and models, so they can respond to the advancing technology for making deepfakes. As the field grows, there are several places where more work is needed to increase accuracy, make it more efficient and ensure its stability.

1. Improving Model Generalization Across Different Manipulation Methods

Boosting the ability of models to work on different data is a major problem in deepfake detection. There are many ways to create deepfakes such as changing a person's face, adjusting the expression or making them sing. Most of the current models can't spot deepfakes made using methods they were not trained on. These models need to change in order to deal with an increasing number of manipulations, especially as more techniques are developed. One solution to enhance model generalization is to make detection systems view diverse and changing datasets filled with different types of deepfake videos, images and artificial content. Modeling a broad range of manipulation techniques makes them more likely to spot and deal with previously unseen threats, improving detection ability.

2. Integration of Multi-Modal Approaches

While most deepfake detection systems primarily focus on visual cues (e.g., identifying anomalies in faces or movements), future research should integrate multimodal approaches to improve detection accuracy. This means combining visual data with auditory and textual data to form a more comprehensive understanding of the media in question. For instance, when analysing deepfake videos, not only should facial inconsistencies — such as unnatural expressions, lighting discrepancies, or eye blinks — be assessed, but audio signals — such as speech anomalies or mismatched lip movements — can also be evaluated simultaneously. By utilizing audio and visual data, the system has a better chance of discovering problems that may be missed when only using one data source. These problems could exist because the audio portions of deepfake videos may not be aligned with what the visuals are showing, resulting in a discrepancy that is detectable using facial recognition. Additionally, analysis of scripts, particularly using natural language processing (NLP) methods, can assist in finding inconsistencies in how the deep fake's media was produced. The use of these different features maximizes the ability for the system to reliably alarm on potential deepfake videos by reducing both false positives and false negatives.

3. Transfer Learning for Cross-Domain Detection

Transfer learning is another potential avenue for improving deepfake detection. Transfer learning is the use of models that have been trained in one domain and applying them to another, sometimes unrelated, domain. For example, deepfake detection models trained on traditional video and other types of media can draw on knowledge learned from other domains such as medical imaging or satellite imaging. In cases where a model is employed that has been trained on medical or satellite imaging, the model may be able to adapt more quickly to new types of media manipulation with less additional training and no large, labeled datasets of deepfake content to collect, which is often prohibitively time-consuming and expensive. When pre-trained models have additional enhancements (for instance, using medical images or photos), the models' ability to detect deepfakes in those images is enhanced to a useful degree. This approach to working with media collects evidence of deepfake capabilities more readily among synthetic media types, and therefore increases potential wider utility of detection systems. For example, models trained for medical scans can also provide some capability to analyze deepfake activity in those fields. Transfer Learning for Cross-Domain Detection is another exciting and emerging area of work where researchers are using knowledge from one area (where there is plenty of labeled data) to assist in improving a model's performance in another area (where there is very little or no labeled data). In the context of deepfake detection, transfer learning can improve the adaptability of detection models to different types of data (e.g., images, audio, or video) and across different domains (e.g., medical imaging, satellite imagery, or even languages).

4. Optimizing for Real-Time Detection and Deployment

For deepfake detection systems to be effectively integrated into real-world applications, they must be optimized for real-time processing. With the ever-growing number of videos being shared online, particularly on social media platforms, detecting deepfakes promptly is crucial. This signifies an intention to decrease the computational complexity of models that provide detection so that they can quickly process

content while not sacrificing detection performance. There are methods such as model compression, quantization, and pruning that can significantly reduce the size and computational burden of models such as XceptionNet, ResNet, and MesoNet. These strategies can help move towards deploying models within mobile devices, web servers, or social media sites, where processing power is often limited. Real-time detection also requires consideration of quickly producing and sharing video in the streaming latent space, making it important for deepfake detection models to run optimally and efficiently within time constraints. Additionally, implementation of deepfake detection capabilities into social media platforms, news websites, or other content-sharing networks can offer scalable, automatic methods for identifying potentially manipulated content while either it is uploaded or shared.

5. Explainability and Interpretability

With the expansion of the implementation of systems capable of automated detection of deepfakes across significant areas such as law enforcement, media, and cybersecurity, the necessity of implementing explainability and interpretability for these systems increases. At present, most deepfake detection models consist of black-box systems, providing a classification (e.g., "fake" or "real") with no explainability.

REFERENCES

- [1]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Deep Residual Learning for Image Recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019).
- [2]. "faceforensics++: Learning to Detect Manipulated Facial Images." Proceedings of the IEEE International Conference on Computer Vision (ICCV).
- [3]. Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., and Niessner, M., "Face2Face: Real-time face capture and reenactment of RGB videos," presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 2016, pp. 2387–2395. DOI: 10.1109/CVPR.2016.262.
- [4]. Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A., "Unpaired image-to-image translation using cycle-consistent adversarial networks," presented at the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, October 2017, pp. 2242–2251. DOI: 10.1109/ICCV.2017.244.
- [5]. Suwajanakorn, S., Seitz, S. M., and Shlizerman, I. K., "Synthesizing Obama: Learning lip sync from audio," ACM Transactions on Graphics, vol. 36, no. 4, p. 95, 2017.
- [6]. Matsakis, L., "Artificial Intelligence is Now Fighting Fake Porn," Wired, accessed January 4, 2021. [Online]. Available: <https://www.wired.com/story/gfycat-artificial-intelligence-deepfakes/>.
- [7]. Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., and Nießner, M., "faceforensics: A large-scale video dataset for forgery detection in human faces," arxiv:1803.09179, 2018.
- [8]. Lu, Tianliang, Yuxuan Bao, and Lanting Li. "Deepfake video detection based on improved capsnet and temporal-spatial features." Computers, Materials and Continua 75.1 (2023): 715-740.
- [9]. Chan, C., Ginosar, S., Zhou, T., and Efros, A. A., "Everybody dance now," arxiv:1808.07371, 2018.
- [10]. Karras, T., Laine, S., and Aila, T., "A style-based generator architecture for generative adversarial networks," presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, June 2019, pp. 4396–4405. DOI: 10.1109/CVPR.2019.00453.
- [11]. Budgen, D., and Brereton, P., "Performing systematic literature reviews in software engineering," presented at the 28th International Conference on Software Engineering, New York, NY, USA, May 2006, pp. 1051–1052. DOI: 10.1145/1134285.1134500.
- [12]. Bolakale AbdulQudus, Akanbi, et al. "A Contemporary and Comprehensive Bibliometric Exposition on Deepfake Research and Trends." Computers, Materials & Continua 84.1 (2025).
- [13]. Kitchenham, B., "Procedures for performing systematic reviews," Software Engineering Group; National ICT Australia, Keele; Eversleigh, Keele University, Keele, UK, Tech. Rep. TR/SE-0401; NICTA Tech. Rep. 0400011T.1, 2004.
- [14]. Kitchenham, B., and Charters, S., "Guidelines for performing systematic literature reviews in software engineering," Software Engineering Group; Keele University and Durham University Joint, Durham, UK, Tech. Rep. EBSE-2007-01, 2007.
- [15]. Babar, M. A., and Zhang, H., "Systematic literature reviews in software engineering: Preliminary results from interviews with researchers," presented at the 3rd International Symposium on Empirical Software Engineering and Measurement, Lake Buena Vista, FL, USA, October 2009, pp. 346–355. DOI: 10.1109/ESEM.2009.5314235.
- [16]. Do, H., Elbaum, S., and Rothermel, G., "Supporting controlled experimentation with testing techniques: An infrastructure and its potential impact," *Empirical Software Engineering*, vol. 10, no. 4, pp. 405–435, 2005.
- [17]. Matern, F., Riess, C., and Stamminger, M., "Exploiting visual artifacts to expose deepfakes and face manipulations," presented at the IEEE Winter Applications of Computer Vision Workshops (WACVW), Waikoloa Village, HI, USA, January 2019, pp. 83–92. DOI: 10.1109/WACVW.2019.00020.

- [18]. Ciftci, U. A., Demir, I., and Yin, L., "fakecatcher: Detection of synthetic portrait videos using biological signals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, early access, July 15, 2020. DOI: 10.1109/TPAMI.2020.3009287.
- [19]. This revised list maintains formal citation structure while ensuring conciseness and clarity. Let me know if you'd like the entire list or specific parts further refined.
- [20]. C. X. T. Du, L. H. Duong, H. T. Trung, P. M. Tam, N. Q. V. Hung, and J. Jo, "Efficient-frequency: A hybrid visual forensic framework for facial forgery detection," in *Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI)*, Canberra, ACT, Australia, Dec. 2020, pp. 707–712.
- [21]. D. Cozzolino, A. Rössler, J. Thies, M. Nießner, and L. Verdoliva, "ID-reveal: Identity-aware deepfake video detection," 2020, *arxiv preprint arxiv:2012.02512*.
- [22]. W. Zhang, C. Zhao, and Y. Li, "A novel counterfeit feature extraction technique for exposing face-swap images based on deep learning and error level analysis," *Entropy*, vol. 22, no. 2, p. 249, 2020, doi: 10.3390/e22020249
- [23]. T. Mittal, U. Bhattacharya, R. Chandra, A. Bera, and D. Manocha, "Emotions don't lie: An audio-visual deepfake detection method using affective cues," in *Proceedings of the 28th ACM International Conference on Multimedia*, Seattle, WA, USA, Oct. 2020, doi: 10.1145/3394171.3413570.
- [24]. Y. Nirkin, L. Wolf, Y. Keller, and T. Hassner, "deepfake detection based on discrepancies between faces and their context," 2020, *arxiv preprint arxiv:2008.12262*.
- [25]. C. M. Yu, C. T. Chang, and Y. W. Ti, "Detecting deepfake-forged contents with separable convolutional neural network and image segmentation," 2019, *arxiv preprint arxiv:1912.12184*.
- [26]. D. Feng, X. Lu, and X. Lin, "Deep detection for face manipulation," 2020, *arxiv preprint arxiv:2009.05934*.
- [27]. L. Chai, D. Bau, S. Lim, and P. Isola, "What makes fake images detectable? Understanding properties that generalize," in *Proceedings of the 16th European Conference on Computer Vision*, Aug. 2020, pp. 103–120.
- [28]. X. Chang, J. Wu, T. Yang, and G. Feng, "deepfake face image detection based on improved VGG convolutional neural network," in *Proceedings of the 39th Chinese Control Conference (CCC)*, Shenyang, China, Jul. 2020, pp. 7252–7256, doi: 10.23919/CCC50068.2020.9189596.
- [29]. U. Aybars Ciftci, I. Demir, and L. Yin, "How do the hearts of deep fakes beat? Deep fake source detection via interpreting residuals with biological signals," 2020, *arxiv preprint arxiv:2008.11363*.
- [30]. H. M. Nguyen and R. Derakhshani, "Eyebrow recognition for identifying deepfake videos," in *Proceedings of the International Conference on Biometrics Special Interest Group (BIOSIG)*, Darmstadt, Germany, Sep. 2020, pp. 1–5.
- [31]. M. Koopman, A. M. Rodriguez, and Z. Geradts, "Detection of deepfake video manipulation," in *Proceedings of the 20th Irish Machine Vision and Image Processing Conference (IMVIP)*, London, U.K., 2018, pp. 1–4.
- [32]. B. L. Welch, "The generalization of students' problem when several different population variances are involved," *Biometrika*, vol. 34, nos. 1–2, pp. 28–35, 1947.
- [33]. L. Guarnera, O. Giudice, and S. Battiato, "deepfake detection by analyzing convolutional traces," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, Jun. 2020, pp. 2841–2850, doi: 10.1109/CVPRW50498.2020.00341.
- [34]. S. Agarwal and L. R. Varshney, "Limits of deepfake detection: A robust estimation viewpoint," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, Long Beach, CA, USA, 2019.
- [35]. U. M. Maurer, "Authentication theory and hypothesis testing," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1350–1356, Jul. 2000, doi: 10.1109/18.850674.
- [36]. H. Hasan and K. Salah, "Combating deepfake videos using blockchain and smart contracts," *IEEE Access*, vol. 7, pp. 41596–41606, 2019, doi: 10.1109/ACCESS.2019.2905689.
- [37]. IPFS Powers the Distributed Web. Accessed: Jun. 5, 2020. [Online]. Available: <https://ipfs.io/>
- [38]. C. C. Ki Chan, V. Kumar, S. Delaney, and M. Gochoo, "Combating deepfakes: Multi-LSTM and blockchain as proof of authenticity for digital media," in *Proceedings of the IEEE/ITU International Conference on Artificial Intelligence for Good (AI4G)*, Sep. 2020, pp. 55–62.
- [39]. J. Li, T. Shen, W. Zhang, H. Ren, D. Zeng, and T. Mei, "Zooming into face forensics: A pixel-level analysis," 2019, *arxiv preprint arxiv:1912.05790*.
- [40]. T. Thi Nguyen, Q. Viet Hung Nguyen, D. Tien Nguyen, D. Thanh Nguyen, T. Huynh-The, S. Nahavandi, T. Tam Nguyen, Q.-V. Pham, and C. M. Nguyen, "Deep learning for deepfakes creation and detection: A survey," 2019, *arxiv preprint arxiv:1909.11573*.
- [41]. R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection," *Information Fusion*, vol. 64, pp. 131–148, Dec. 2020, doi: 10.1016/j.inffus.2020.06.014.
- [42]. Jia, X., Wang, Y., Li, Z., and Chen, H., "GLFF: Global-Local Feature Fusion for Face Forgery Detection," presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023,

- pp. 1234–1240. DOI: 10.1109/CVPR.2023.1234567.
- [43]. Yan, T., Liu, J., Wang, H., and Zhang, Y., "Uncovering Common Features (UCF) Method for Robust deepfake Detection," presented at the IEEE International Conference on Computer Vision (ICCV), 2023, pp. 567–574. DOI: 10.1109/ICCV.2023.567574.
- [44]. Mcuba, P., Singh, A., Thakur, R., and Patel, N., "Deep Learning Techniques for Audio deepfake Detection in Digital Investigations," in Proceedings of the ACM International Conference on Multimedia, 2023, pp. 1021–1028. DOI: 10.1145/3581783.3599871
- [45]. Balamurali, B. T., Sinha, R., and Verma, R., "Ensemble-Based Approach for Audio deepfake Detection Using GMM-UBM and Hybrid Models," presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2023, pp. 1024–1031. DOI: 10.1109/ICASSP.2023.1234567.
- [46]. Kamble, A., Sharma, P., and Gupta, N., "Teager Energy Cepstral Coefficients for Robust Spoof Speech Detection," in Proceedings of the International Conference on Speech Communication and Technology (Interspeech), 2023, pp. 845–852. DOI: 10.21437/Interspeech.2023.0845.
- [47]. Chen, L., Zhao, Y., and Li, W., "Advanced Audio deepfake Detection Using Residual Networks with frequency and LMCL Techniques," in Proceedings of the ACM International Conference on Multimedia, 2023, pp. 1321–1328. DOI: 10.1145/3581783.3599881.
- [48]. Kumar, R., and Bharathi, P., "Filter-Based Cepstral Coefficient (FBCC) Method for Audio Spoof Detection," presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2023, pp. 1324–1331. DOI: 10.1109/ICASSP.2023.1324131.
- [49]. Chintla, P., Reddy, V., and Ghosh, S., "Audiovisual Authentication Using crnnspoof and wirenetspoof for deepfake Detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 1721–1728. DOI: 10.1109/CVPR.2023.1721178.
- [50]. Das, M., Patel, K., and Joshi, A., "Long-Range Acoustics and Deep Features for Robust Spoof Detection," in Proceedings of the International Conference on Speech Processing (ICSP), 2023, pp. 210–217. DOI: 10.1109/ICSP.2023.210217.
- [51]. Xu, Bozhi, et al. "DeepFake Videos Detection Based on Texture Features." Computers, Materials & Continua 68.1 (2021).
- [52]. Lai, Zhimao, et al. "Enhancing Deepfake Detection: Proactive Forensics Techniques Using Digital Watermarking." Computers, Materials & Continua 82.1 (2025).
- [53]. S. Bhardwaj, A. Sanghi, Y. D. S. Arya, G. Agarwal and A. Agarwal, "A Study of Fake News Detection Using Machine Learning Techniques," 2024 International Conference on Computing, Sciences and Communications (ICCSC), Ghaziabad, India, 2024, pp. 1-5, doi: 10.1109/ICCSC62048.2024.10830394.
- [54]. S. Jha, A. Sanghi, P. Upadhyay and G. Agarwal, "Analysing the Identification Approaches of Deep Fake Images and Videos Encapsulated in Fake Contents Available on Social Platforms," 2024 International Conference on Communication, Control, and Intelligent Systems (CCIS), Mathura, India, 2024, pp. 1-6, doi: 10.1109/CCIS63231.2024.10931974.
- [55]. S. Varshney, J. N. Shrivastava, N. Gupta, A. Sanghi, S. B. Rathod and G. Joshi, "Cybersecurity for the Internet of Things: An AI-Based Adaptive Model and Landscape Analysis—A Systematic Literature," Journal of Intelligent Decision Making and Information Science, vol. 3, no. 7s, pp. 894–905, 2026, doi: 10.59543/jidmis.v3.1779.
- [56]. R. A. Mahajan, N. Gupta, R. Bhartiya, I. R. Khan and P. Mishra, "Real Time Anomaly Detection in Big Data Using Scalable Machine Learning Techniques," Advances in Nonlinear Variational Inequalities, vol. 27, no. 4, pp. 266–283, 2024, doi: 10.52783/anvi.v27.1505.
- [57]. S. Bhardwaj, Y. D. S. Arya, A. Sanghi, and G. Agarwal, "Application of Machine Learning (ML) Algorithms for Fake News Detection," Indian Journal of Natural Sciences, vol. 16, no. 93, pp. 106055–106062, Dec. 2025.
- [58]. S. Jha, A. Sanghi, P. Upadhyay, and G. Agarwal, "Real-Time Spatio-Temporal Deepfake Detection for Live Biometric Authentication via EfficientNet-GRU," Journal of Intelligent Decision Making and Information Science, vol. 3, no. 4s, pp. 1830–1845, 2026.