**IMPROVING DEEPFAKE DETECTION IN VIDEOS USING DEEP NEURAL NETWORKS**

**INTRODUCTION**

Deepfakes—synthetic media created using sophisticated machine learning techniques—present serious issues across a range of domains. In order to lessen the detrimental effects of deepfakes, it is essential to comprehend them and develop detection techniques. Through an examination of several models, observations of existing gaps in the literature, and an examination of current methodologies, this research seeks to determine the most effective methods for detecting deepfakes.

1. What are Deepfakes?

Deepfakes are synthetic media made with deep learning algorithms; they are mainly films or images. These algorithms alter or substitute the original information with fake film that is frequently identical to real footage. Because deepfake technology is so sophisticated, there is a chance that it will propagate false information, erode public confidence in the media, and jeopardize security and privacy.

2. Why Is It necessary to detect deepfakes?

The spread of deepfakes intensifies worries about digital manipulation, distortion, and disinformation already present. Deepfake detection is crucial for maintaining information integrity, shielding people from harm, and maintaining the authority of visual content in a variety of fields, such as politics, journalism, entertainment, and security.

3.Research in Deepfakes

One of the most important areas of research regarding deepfakes is the creation and detection of deepfakes. Several studies have shown various techniques for deepfake creation as well as detection. The authors of the study categorized deepfake methodologies into four groups, including deep-learning-based, classical machine-learning-based, statistical, and blockchain-based techniques. And according to their findings deep-learning-based methods perform better than all.

4.Social Impact of Deepfake

Deepfake has an ethical concern with both positive and negative effects on society.

*Positive impact –* Deepfakes are used in creating the best version of videogames, electronic communications, AR/VR oriented clothing and e-commerce are just few applications the holds benefit from deepfake invention. Some instances showing deepfake oriented actions are mentioned below.

* Now businesses can hire phony models and actresses to display fashionable attire on a wide range of models with various height, weight, skin color.
* realistic-sounding and smart-looking assistants and enhanced telepresence in online games and virtual chat environments. This promotes improved online communication and interpersonal relationships.

*Negative Impact*

Deepfake creates numerous amounts of security issues as well as risks to one’s personal space and image. Usually, the targeted audience for deepfakes are famous celebrities and politicians.

This is the reason experts are alerting the public to the threats that deepfakes may pose. Let’s analyzing the few of the negative impacts of deepfake.

* “deepfakes” derives its name from being used to create pornographic material on purpose. To put this into context, VOX research [81–84] indicates that 96% of the deepfake movies in online pornography in 2020 were produced to harm the reputations of their victims.
* Deepfakes can be used by political enemies to sway the people and to foster mistrust. Barack Obama was caught on camera in 2008, 2012, and 2016 stating that individuals in hard-hit areas frequently turned to religion and guns.

**PROBLEM STATEMENT**

Deepfakes pose a significant threat by enabling the creation of highly realistic manipulated videos that can be used to spread misinformation, damage reputations, and erode trust in media. Current deepfake detection methods, while effective to some degree, still face challenges in accurately identifying deepfakes, particularly as deepfake creation techniques continue to evolve.

This research aims to address this critical gap by exploring and evaluating the potential of deep neural networks for improved deepfake detection in videos. The objective is to develop a robust and generalizable deep learning model that can effectively distinguish between real and deepfake videos, even in the face of advancements in deepfake creation methods.

**REVIEW OF THE LITERATURE**

A systematic review is a suitable method for compiling existing studies and identifying, the gaps that can suggest a new area of research. To ensure the significance and impact of this research, we conducted a comprehensive review of existing deepfake detection literature. This critical analysis provided valuable and informed our approach towards developing more robust deepfake detection methods. The review methodology has two main sections, the search phase and the definition of the inclusion and exclusion criteria. The search step entails incite about the search engines, research papers, academic resources that may be used to look for the appropriate research. The below table shows the following search

|  |  |  |  |
| --- | --- | --- | --- |
| TYPE | Paper |  | URL |
| Search Engine | Google Scholar |  | <https://scholar.google.com.au> |
| Digital Library | IEEE Xplore |  | <https://ieeexplore.ieee.org/Xplore/home.jsp> |
| Digital Library | Springer |  | <https://www.springer.com/gp> |
| Digital Library | MDPI |  | <https://www.mdpi.com> |

The following stage involves defining the inclusion criteria (criteria to choose the research paper for the relative study). This improves the quality of research and make the task easy. Moreover, we look at the title and keywords of the resource, to obtain a more targeted search outcome.

|  |  |
| --- | --- |
| Inclusion Criteria number | List of Inclusion Criteria |
| 1. | Must be included in one of the selected databases |
| 2. | Should be related to title searched and must contain |
| 3. | Published in a journal and within last six years (2019-2024) |
| 4. | Should contain the required keywords related to deepfake detection |

**Underlying Deepfake Detection Models**

The existing models in the literature that are being developed to detect deepfake are studied in this section. Based on the underlying detection algorithm, we can broadly categorize these models into two types:

(1) Conventional machine-learning-based models

(2) Deep-learning-based models.

Deep-learning-based models can be further classified into CNN, RNN, and transformer-based models.

1)Conventional machine-learning-based models

Modern machine learning techniques are crucial for comprehending the reasoning behind any choice that may be justified from a human perspective. These techniques provide more control over data and procedures, making them appropriate for the deepfake area. Earlier techniques used tree approach like decision tree and random forest, but now a days Support vector machine (SVM), logistic regression, KNN classifiers and boosting models (XGBoost, ADABoost to identify deepfake)

2)Deep-learning-based models

**CNN-based Models** One of the best deep learning models is CNN. It is well known and directly utilizes pretrained CNN methodologies to learn distinguishing features from each individual frame of the sequences. We discovered that the CNN-based approaches have been employed in deepfake detection studies: Xception, GoogleNet, VGG16, Y-shaped Autoencoder, MesoNet , DeepRhythm.

**RNN-based Models** One of the most used models for sequential data in deep learning is RNN. Numerous RNN models have been found to be employed in the creation and detection of deepfake images and videos. Here are several RNN models that can be used to create and detect deepfakes: BiLSTM, FaceNet, FacenetLSTM, Neural-ODE, CLRNet, CNN+(Bidirectional + entropy RNN), CNN+RNN

**Transformer-based Models** Transformers are highly well-known deep learning models, and they are now used in the field of deepfake generation and detection. The following transformer models can be used to generate and recognize deepfakes: EfficentNet+ViT, M2TR, CViT, ViT, ViT+Distillation.

|  |  |  |  |
| --- | --- | --- | --- |
| Category | Model | Dataset | Accuracy |
| CNN | MesoNet | CelebDF | 57% |
| CNN | MesoInception | CelebDF | 50.73% |
|  | BlazeFace + Efficient Net B5 + SVM | DFDC | 82.46% |
|  | YOLO + Xception + SVM | DFDC | 83.37% |
|  | MTCNN [31,41,43] + InceptionResNetV2 + XGBoost | DFDC | 84.45% |
|  | YOLO + InceptionResNetV2 + XGBoost | DFDC | 90.73% |
|  | YOLO + ResNet152 [38] + SVM | DFDC | 82.50% |
|  | YOLO + ResNet152 [38] + XGBoost | DFDC | 84.88% |
|  | BlazeFace + Efficient Net B5 [43] + XGBoost | DFDC | 84.58% |
|  | BlazeFace + Xception [36] + XGBoost | DFDC | 90.37% |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Gaps and future work analyzed from the already existing resources related to Deepfakes**

**1. Absence of a Centralized Benchmarking** Dataset: Despite the fact that detection accuracy is the main parameter used to assess deepfake detection algorithms, disparities in the reported accuracies of various models are a result of the lack of a centralized benchmarking dataset. The development of detection algorithms is hampered by the lack of consistent datasets, which also makes comparisons unfair.

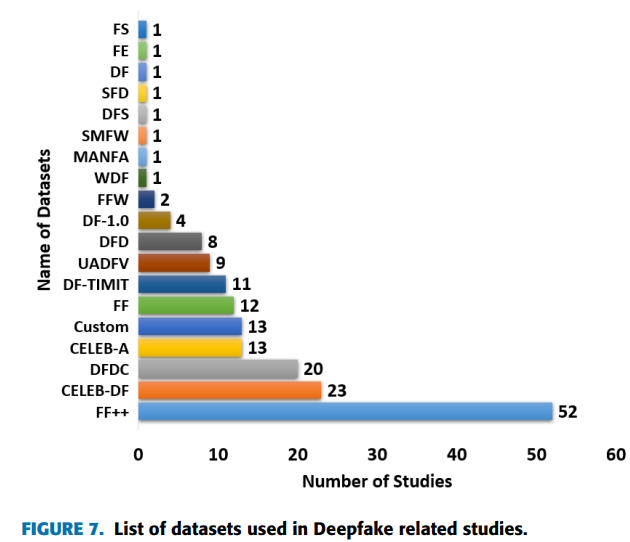
**2. Limited Use of Human Involvement:** Deepfake detection offers unrealized potential for human participation. Contextual awareness, fact-checking, and visual inspection are useful additions to automated detection systems. To date existing research frequently fails to consider the incorporation of human knowledge, undervaluing the advantages of human-in-the-loop methods. Li, L.; Bao, J.; Zhang, T in conference paper “Conference on Computer Vision and Pattern Recognition” introduced a potent method for deepfake detection by exposing the blending boundaries within manipulated images. However, their approach involves a complex trace generator, which results in a high computational workload. In contrast, Durall et al. [21] proposed a straightforward method for deepfake detection based on classical frequency domain analysis followed by a basic classifier.

**3. Challenges with Real-Time Detection**: Because deepfake algorithms are constantly changing and computational limitations exist, real-time detection of deepfakes remains a substantial difficulty. Moreover, there is always a trade-off between accuracy and time latency. Nevertheless, for real-world use of efficient deepfake detection methods, it must be robust, generalizable, quick and computation efficient. To mitigate potential effects and intervene promptly, it is imperative to develop efficient algorithms that can detect deepfakes in real-time circumstances.

**4. Limited Generalizability of Detection Models:** A lot of the deepfake detection models now in use are trained on datasets or manipulation techniques, which restricts their ability to generalize to new or developing deepfake techniques. More flexibility, the ability to learn from a variety of datasets, and the ability to recognize innovative modification techniques are all needed to overcome this constraint in detection models. There is active research in the field of deepfake detection that leverages the frequency domain for various purposes Durall employed the Discrete Fourier Transform (DFT) for deepfake detection. They propose a straightforward approach to detect fake face images by converting the images into the frequency domain using DFT. Afterward, they extract features from a 1D representation of the Fast Fourier Transform power spectrum through logistic regression and Support Vector Machines (SVM). Also, applied frequency domain learning in knowledge distillation to improve low-quality compressed deepfake image detection. **Their methods exhibit a weakness in detection performance for high-quality deepfake images.**

**RESOURCES**

|  |  |  |
| --- | --- | --- |
| RESEARCH TITLE | PAPER | URL |
| An Improved Dense CNN Architecture for Deepfake Image Detection | IEEE  (published on March 2,2023) | [\*An\_Improved\_Dense\_CNN\_Architecture\_for\_Deepfake\_Image\_Detection 2023 p7.pdf](file:///C:\Users\lenovo\AppData\Local\Temp\13d7f1f3-d858-4850-80b5-085a005a1fc1_research%20paper%20deepfake.zip.fc1\research%20paper%20deepfake\An_Improved_Dense_CNN_Architecture_for_Deepfake_Image_Detection%202023%20p7.pdf) |
| Real -Time Advanced Computational Intelligence for Deep Fake Video Detection | MDPI  (published on February 27,2023) | [applsci-13-03095-v2 MDPI 2023 P8.pdf](file:///C:\Users\lenovo\AppData\Local\Temp\68856b90-dc14-482b-b864-87d3df552fd1_research%20paper%20deepfake.zip.fd1\research%20paper%20deepfake\applsci-13-03095-v2%20%20%20MDPI%202023%20P8.pdf) |
| DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern | IEEE  (published on April 20,2020) | [DeepVision\_Deepfakes\_Detection\_Using\_Human\_Eye\_Blinking\_Pattern IEEE P2.pdf](file:///C:\Users\lenovo\AppData\Local\Temp\29036e5a-19e6-4bca-8a71-1aca4fd8960c_research%20paper%20deepfake.zip.60c\research%20paper%20deepfake\DeepVision_Deepfakes_Detection_Using_Human_Eye_Blinking_Pattern%20IEEE%20P2.pdf) |
| Multiclass AI-Generated Deepfake Face Detection Using Patch-Wise Deep Learning Model | MDPI  (published January 2024) | [Multiclass AI-Generated Deepfake Face Detection Using Patch-Wise Deep Learning Model P10.pdf](file:///C:\Users\lenovo\AppData\Local\Temp\1a8401d4-0e30-4c52-86d4-4c57657f5e0e_research%20paper%20deepfake.zip.e0e\research%20paper%20deepfake\Multiclass%20AI-Generated%20Deepfake%20Face%20Detection%20Using%20Patch-Wise%20Deep%20Learning%20Model%20P10.pdf) |
| Assessment framework for deepfake detection in real-world situations | Springer  (published in 2024) | [s13640-024-00621-8 (1)P6.pdf](file:///C:\Users\lenovo\AppData\Local\Temp\adbee269-c788-4aff-ac8b-9f5176657cb8_research%20paper%20deepfake.zip.cb8\research%20paper%20deepfake\s13640-024-00621-8%20(1)P6.pdf) |
| Detection of AI-Created Images Using Pixel-Wise Feature Extraction and Convolutional Neural Networks | MDPI  (published on November 8,2023) | [Detection of AI-Created Images Using Pixel-Wise Feature Extraction and Convolutional Neural Networks P9.pdf](file:///C:\Users\lenovo\AppData\Local\Temp\5eb4d4a9-1ae2-4ce0-8df3-976cd400c767_research%20paper%20deepfake.zip.767\research%20paper%20deepfake\Detection%20of%20AI-Created%20Images%20Using%20Pixel-Wise%20Feature%20Extraction%20and%20Convolutional%20Neural%20Networks%20P9.pdf) |
| An Investigation of the Effectiveness of Deepfake Models and Tools | MDPI  (published on August 4,2023 | [An Investigation of the Effectiveness of Deepfake Models and Tools.pdf](file:///C:\Users\lenovo\AppData\Local\Temp\e059678a-a8d0-43f5-84fc-d176c45c06ef_research%20paper%20deepfake.zip.6ef\research%20paper%20deepfake\An%20Investigation%20of%20the%20Effectiveness%20of%20Deepfake%20Models%20and%20Tools.pdf) |



**AUTHORS-**

SATVIK KARAN (E22CSEU1332)

SARGAM TYAGI (E22CSEU1583)