IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, April 2026
Impact Factor: 8.2

Development of AI/ML Based Solution for Detection

Anushka Singh[1], Snehal Bhosarkar[2], Pradnya Jambhale[3]
Students, Department of Computer Engineering[1,2,3]
SNKSITs, Lonavala, Maharashtra, India

Abstract: *Deep learning has proven effective in a variety of tough issues, including computer vision, human-level control, and large data analytics. However, as deep learning technology advanced, software was developed that jeopardized national security, democracy, and privacy. Deepfake is a new technology that uses deep learning to create fake photos and videos that look very real. It's important to have tools that can automatically detect and check the quality of these AI-created images and videos. These systems help us quickly tell if a picture or video is real, edited, or fake, and they ensure that the quality is good and not misleading. An investigation of the strategies used to construct the most significant deepfakes, as well as the approaches proposed in the literature for detecting them. We provide a complete examination of the difficulties highlighted by deepfake technology, as well as recommendations for future and upcoming research opportunities. It also supports creating new and more reliable ways to handle deepfakes as they become more complex.*

Keywords: Deep Learning, CNN, Pre-Processing, Feature Extraction, Face Detection and Face Recognition

I. INTRODUCTION

With easy internet access and fast-growing technology, people and businesses can now communicate through social media. Lifelike digital content (text, video, and audio) may now be created using breakthroughs in generative artificial intelligence (AI). "Deepfakes" are fake content like images, videos, or sounds that look and sound real, made using advanced AI and machine learning. [1] Over the past few years, big improvements in AI and machine learning have led to new tools that help create and edit media, such as photos and videos. While the majority of people use technology for education and entertainment, some dishonest people abuse it for destructive or illegal objectives. To disseminate fake information, provoke political unrest or violence, or even threaten and control individuals, extremely realistic bogus audio, video, or image content has been generated. The stunning, lifelike, and expertly edited movies are now known as "Deepfake." [2]

Recent methods for detecting deepfakes focus on looking at each frame of a video and checking for signs that it might be fake. Some more advanced systems also look for unusual changes over time in the video, but most research mainly focuses on what's happening in each frame [3]. A human- centered approach to detecting forgeries in facial images use dynamic prototypes as visual explanations. Right now, most deepfake detection methods use "black-box" algorithms that check videos frame by frame, without considering changes that happen over time in the video. The presence of temporal artifacts in deepfake movies makes them easier for supervisors to detect and explain [4].

Deepfakes are videos where someone's face is replaced with another person's face using artificial intelligence (AI), have become increasingly problematic. These videos have the potential to jeopardize privacy and encourage fraud. Detecting high- quality Deepfake films may be challenging for human vision [5]. Deepfakes use generative adversarial networks to generate fake video content. Deepfake technology has raised questions about its potential impact on society, particularly electoral bias. Researchers have focused on building detecting algorithms to mitigate the detrimental impact of deepfakes [6].

DOI: 10.48175/IJARSCT-32680

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, April 2026 Impact Factor: 8.2

Recently, a free machine learning-based software application simplified the process of creating convincing face swaps in videos, resulting in "deepfake" videos that require minimal editing. Realistic bogus movies can be used to incite political unrest, blackmail, or stage terrorist acts. This article talks about a method that can automatically detect deepfake videos by paying attention to changes over time in the video.[7]

II. LITERATURE SURVEY

In 2023 et.al Yogesh Patel [1] have Researched Customers may now easily access multimedia content in smart communities thanks to social media. Deep learning and machine learning models are new technologies that have come from recent progress in teaching computers to see and understand images, as well as process and understand language. Using generative adversarial networks (GAN), it can now recreate voice, pictures, or video streams while adapting to changing settings based on a person's unique visual and aural characteristics.

Deepfakes are thus actively used on social media sites to disseminate misinformation and propaganda that harms an individual's or company's image. Recently, many studies have looked into how deepfake audio and video are made and how to detect them. Finding deepfake material is the major objective of most surveys conducted today, despite.

In 2022, MD Shohel Rana explained that in the last twenty years, big improvements in AI, machine learning, and deep learning have created new ways to edit and make media, like photos and videos. Although these technologies are mostly used for positive things, like teaching and entertainment, they also make it easier to create realistic edits and effects, dishonest individuals have abused it for evil or illegal purposes. For example, fake audio, video, or images that look real have been created to spread false information, stir up hate or political problems, or even threaten and blackmail people. The term "deep fake" refers to heavily duplicated, realistic, and recently popularized altered videos. Since then, numerous solutions for dealing with Deepfake's problems have been extensively researched in the literature.

In 2020, John K. Lewis and others researched how important it is to verify digital media. The rise of Generative Adversarial Networks has made it harder to spot fake media. Deepfakes, which are fake videos that use altered faces or voices of real people, threaten online privacy and trust. They can be used to spread false information, damage the reputations of famous people, or push political agendas. Even though deepfakes have some flaws, many people can't tell the difference between real and altered videos or images.

As a result, automated systems that can properly and quickly identify whether digital content has been approved are critical. Recent deep fake detection systems look at individual frames of video and use spatial information to determine authenticity.

In 2021, Michael Tsang [4] and others did research on detecting fake facial images. They introduced a new method that focuses on understanding the human side of the process, using dynamic prototypes to explain the results. Most deepfake detection methods use "black-box" models that check each frame of a video one by one. However, only a few methods look at how things change over time in the video. Nonetheless, these temporal anomalies are required in deep fake movies in order to recognize and explain them to a human supervisor. To do this, we developed the Dynamic Prototype Network (DPNet), which use prototypes, or dynamic representations, to provide a coherent and accessible explanation for temporal artifacts observed in deep fake systems.

In 2021, Kaihan Lin [5] and others studied how deepfakes can harm people and society if used wrongly. Researchers have been working hard to find ways to detect deepfakes to reduce their negative impact. Although there has been progress, there's still not a full understanding of how deepfakes are made and detected because researchers have different goals. This study gives an updated overview of how deepfakes are created and detected, and organizes the various methods and data used in deepfake research. The researchers believe this will be a helpful resource for future studies on detecting deepfakes.

A. Challenges with Current Technology

Deepfakes are often made using a technology called GANs (Generative Adversarial Networks). These networks have two parts: the **generator**, which makes fake images or videos, and the **discriminator**, which tries to figure out if the

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, April 2026
Impact Factor: 8.2

content is real or fake. Both parts work together and improve as they are trained. The generator makes fake data, and the discriminator tries to tell if it's real or fake. As they keep working together, the generator gets better, and eventually, the fake data (like deepfakes) becomes so realistic that it's hard to tell the difference from real data.[8]

Challenges in Detection Technologies: Deepfake generating tools (such as GANs) improve their ability to elude existing detection methods [9]. CNNs are widely used in image-based deepfake detection. They are extremely adept at detecting visual anomalies in deepfakes, such as minor pixel deviations, unusual textures, and lighting variations [10].

Image Processing and Real-World Challenges:

This study investigates how Generative Adversarial Networks (GANs) are utilized as the cornerstone for deepfake production, particularly in facial faking techniques [11]. Advanced image processing enables incredibly realistic and misleading face swaps, making it difficult for detection systems to identify tiny visual irregularities [12]. Deepfakes are created by using advanced image editing techniques, like changing faces, adjusting lighting, and blending features together [13]. Motivation for exploring deep learning technologies in deepfakes creation and detection

We need to do more research on deepfake creation methods to improve what we can do with fake media. This will help create more realistic and engaging applications in the future.[1] Strongly needed to protect society from the detrimental effects of deepfakes are scalable and dependable detection systems, particularly in the domains of politics, economics, and private life. Research on deepfake detection systems that use image processing and machine learning is crucial to reducing these threats [2]. The increasing ethical and legal need to give people and organizations tools to guarantee media authenticity and guard against identity theft is driving research into technology detection [3].

B. Research Gap :

Lack of Generalized Detection Models

Current detection algorithms consistently outperform specific datasets or types of deepfakes, but they struggle to generalize across numerous deepfake generation processes, unknown data, and real-world circumstances [14]. This is because to the rapid improvement and variety of deepfake production methods, which produce distinct and dynamic patterns that detection models trained on insufficient or outdated data cannot consistently recognize [15].

Cross-Domain Deepfake Performance

This is because the models have a propensity to overfit to certain features or artifacts in the training data, making them less successful at generalizing to new or emerging deepfake approaches, which might vary substantially in terms of modification methods and visual aspects [16].

Spectral Analysis in Real-World Applications

This is because the models have a tendency to overfit to certain characteristics or artifacts in the training data, making them less effective at generalizing to new or developing deepfake approaches, which may differ significantly in terms of modification methods and visual components [17].

Limited Dynamic Adaptability of Prototypes Current detection approaches usually fail to respond to the dynamic nature of deepfake creation processes [18].

C. ISSUE OF DEEPFAKE DETECTION WITH SOLUTION

Exceptionally reliable deepfakes

Issue: As deepfake technology has improved, it's now much simpler to make fake images and videos that look very real. Most of these deepfakes only have small visual errors, making it hard for regular methods to spot the changes.

Solution: Deep learning models, like Convolutional Neural Networks and Recurrent Neural Networks, can be taught to spot small visual mistakes or unusual behaviors in videos and images.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, April 2026
Impact Factor: 8.2

Difficulty in Real-Time Detection

Problem: Real-time deepfake detection is computationally intensive since enormous volumes of visual and auditory data must be processed quickly.

Solution: Edge computing techniques can be utilized to perform detection closer to the data source, which results in lower latency and faster detection.

Dataset Limitations

Problem: Training models to detect all kinds of deepfakes is difficult because there aren't enough high-quality and varied datasets available for deepfake detection.

Solution: Weaknesses in generalization can be found by employing cross-dataset evaluation techniques, which involve training and testing models over numerous datasets.

III. EXISTING SYSTEM

Today's deepfake production and detection systems rely largely on advanced machine learning and image processing techniques. These methods aren't perfect, but they've made substantial strides in making extremely realistic deepfakes and detecting modified material. Deepfake detection systems may be generalized, making them successful for specific types but ineffective for newer or rarer versions.

Furthermore, the rapid spread of deepfake technology that exceeds detection skills has generated in an ongoing arms race between producers and detectors. Furthermore, many detection algorithms need significant computing power, rendering them unsuitable for general public use.

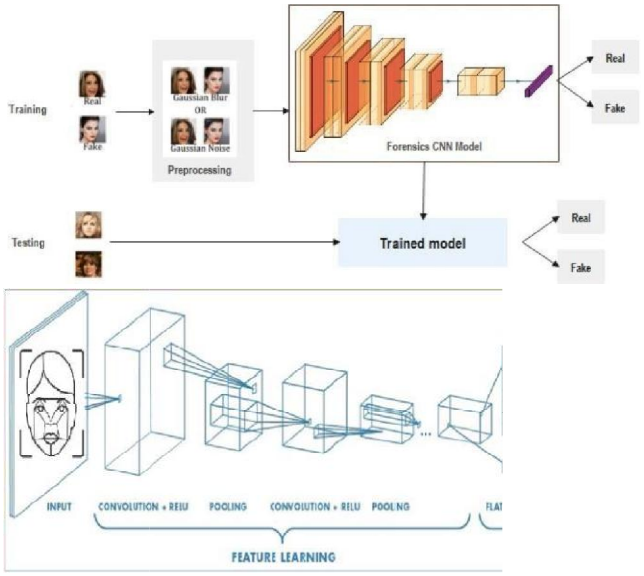

Fig: Proposed Architecture

The proposed method's enhanced Convolutional Neural Network (CNN) architecture addresses problems in prior maps by deepfake detection methods. The CNN's ability to detect minute abnormalities and fine-grained spatial properties in and images and videos enables it to generalize more reliably across various types of deepfakes. CNN is trained on a vast collection of detect distinct patterns of manipulation. Furthermore, multi- numbers scale feature extraction inside the CNN architecture increases takes deepfake detection accuracy even with high-resolution straight line while also increasing processing efficiency for greater accessibility.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, April 2026 Impact Factor: 8.2

Convolutional Layers: These layers filter the input data and apply convolution to extract valuable properties. Each filter from the previous layer to the next. They classify based on the (CNNs) have played a key role in improving Deep Learning for computer vision. CNNs are special computer systems that learn from data and are particularly good at understanding images and videos. CNNs are great at analyzing visual data, like images and videos, because they automatically find recognizes unique patterns or features in data.

Pooling Layers: These layers decreases size of the feature making the data smaller after the convolution layers.

Two common methods for doing this are maximum pooling and average pooling, which simplify the information by across picking the most important values or averaging them.

Flatten: Flattening is when data is turned into a long list of multi- numbers so it can be passed to the next step in the model. It increases takes the output from the previous step and arranges it into one straight line of features. images,

Fully Connected Layers: Dense layers connect every neuron features obtained from previous levels.

IV. ALGORITHM

·**PROPOSED SYSTEM CNN (Convolution Neural Network) :** Convolutional Neural Networks (CNNs) have played a key role in improving Deep Learning for computer vision. CNNs are special computer systems that learn from data and are particularly good at understanding images and videos. CNNs are great at analyzing visual data, like images and videos, because they automatically find important details in the data without needing to be told what to look for. The main layers in a CNN are:

SYSTEM ARCHITECTURE

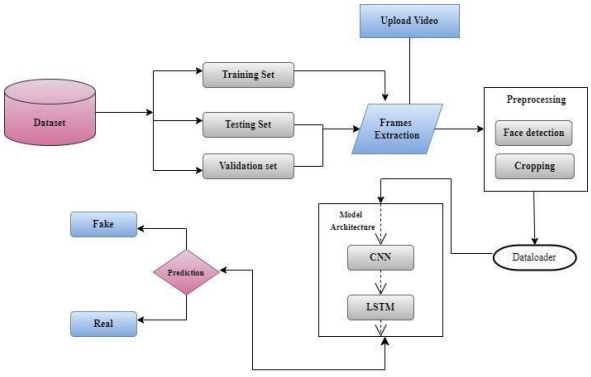

Fig: System Architecture

Data Collection and Preprocessing: The dataset contains both real and fake videos.

Feature Extraction: The video is broken down into individual pictures (frames). Then, faces are found in each of these pictures.

Model Architecture: This part of the network is responsible for picking out important details from the cropped face images. **Training:** The model is taught using a set of training data that has the important features. The goal is for the model to learn the differences between real and fake videos by recognizing patterns in them.

Output: The trained model is used to determine whether a new video is real or fake. The model looks at the features taken from the new video to make this prediction.

V. APPLICATION

Social Media: On social media, deepfakes that spread misleading information or are used to harass and vilify people are frequently seen.

Financial Fraud Prevention: The banking sector utilizes deepfakes to impersonate well-known executives in voice-activated phishing scams and video interactions in order to perpetrate fraud.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, April 2026 Impact Factor: 8.2

Media Integrity and Journalism: The public's trust and the integrity of the journalism profession are jeopardized when deepfakes are exploited to create false information and fake news.

DISCUSSION

Comparison With Existing System

Deep learning models, such as CNNs, outperform human feature extraction techniques in automatically extracting complex patterns, such as odd lighting, texture variations, and strange facial motions, from images and videos. Because hostile instances and high-quality deepfakes rely on outdated techniques and unstable feature sets, they are highly effective at fooling current image processing and machine learning systems.

Implementation in Challenge

To detect all kinds of deepfakes, detection systems need to be trained with large collections of both real and fake images and videos. Deepfakes can be made in a variety of ways, therefore detectors must be exposed to a diverse set of false information, which is not always available. Because of their various features, detection approaches, particularly CNN hybrids for video-based detection, require large processing resources. Real-time detection complicates the situation, especially in videos.

Future Development Opportunity

Concentrating on real-time detection methods and creating more sophisticated machine learning models that can adapt to new deepfake creation strategies. Constructing models that shed light on their decision- making processes, supporting clients in understanding the functioning of deepfake detection, and increasing their technological trust. Creating user-friendly applications to increase awareness and educate the public about deepfakes and the technology.

Real World Impact on User

Misinformation: Deepfake technology makes it difficult to tell whether content, like videos or images, is real or fake because it looks so convincing. This is risky because it can help spread false information and fake news. As a result, trust may be lost in personal contacts, social media, and the internet. Security and Privacy Risks: Deepfakes, which change people's voices, faces, or behaviors without their knowledge, can be used for identity theft, fraud, and cyberbullying. This raises significant privacy and security problems, especially in political or legal circumstances.

VI. CONCLUSION

Deepfakes have become more popular because there's so much content on social media, and tools to create deepfakes are now easier to get. Social media also makes it simple to share fake videos. One way to tell if a video is real or a deepfake is by using AI called a neural network. This AI uses something called CNN (Convolutional Neural Network) to look at the video and decide if it's real or fake, with a high level of confidence in the result.

FUTURE SCOPE

Machine learning-based deep fake detection systems require continual algorithm and model modifications to keep up with advanced techniques.AI technology enables real-time identification of deep fakes in live video feeds, blocking their propagation.

We can integrate deepfake detection system within social media platforms. This system will automatically analyze images and videos before they are posted, verifying their authenticity and determining whether they are fake or real content. By implementing this solution, social security media can enhance security, reduce the spread of misinformation, and promote trust among users. We can also integrate Multi- layered verification process to enhance privacy and security.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, April 2026
Impact Factor: 8.2

REFERENCES

[1]. Patel, Yogesh, Sudeep Tanwar, Rajesh Gupta, Pronaya Bhattacharya, Innocent Ewean Davidson, Royi Nyameko, Srinivas Aluvala, and Vrince Vimal. "Deepfake generation and detection: Case study and challenges." IEEE Access (2023).

[2]. Rana, Md Shohel, Mohammad Nur Nobi, Beddhu Murali, and Andrew H. Sung. "Deepfake detection: A systematic literature review." IEEE access 10 (2022): 25494-25513.

[3]. Lewis, John K., Imad Eddine Toubal, Helen Chen, Vishal Sandesera, Michael Lomnitz, Zigfried Hampel-Arias, Calyam Prasad, and Kannappan Palaniappan.

[4]. "Deepfake video detection based on spatial, spectral, and temporal inconsistencies using multimodal deep learning." In 2020 4. IEEE Applied Imagery Pattern Recognition Workshop (AIPR), pp. 1- 9. IEEE, 2020.

[5]. Trinh, Loc, Michael Tsang, Sirisha Rambhatla, and Yan Liu. "Interpretable and trustworthy deepfake detection via dynamic prototypes." In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 1973-1983. 2021.

[6]. S. P and S. Sk, "DeepFake Creation and Detection:A Survey," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2021, pp. 584-588, doi: 10.1109/ICIRCA51532.2021.9544522.

[7]. D. Pan, L. Sun, R. Wang, X. Zhang and R. O. Sinnott, "Deep fake Detection through Deep Learning," 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT), Leicester, UK, 2020, pp. 134-143, doi: 10.1109/BDCAT50828.2020.00001.

[8]. A. Malik, M. Kuribayashi, S. M. Abdullahi and A. N. Khan, "DeepFake Detection for Human Face Images and Videos: A Survey," in IEEE Access, vol. 10, pp. 8757-18775, 2022,doi: 10.1109/ACCESS.2022.315118.