MULTIDISCIPLINARY RESEARCH INSTITUTE INDIA

Multidisciplinary Journal of Research in Engineering and Technology

ISSN: 2348-6953

Volume 13 Issue 2s (2026) | 325-329

Deepfake Detection Using AI: Techniques, Challenges, and Future Prospects

Dipannita Mondal[1], Shravan Gholap[2], Sanskar Gondal[3], Nagesh Yamulwad[4], Vedant Katkar[5]

[1,2,3,4,5]Department of Artificial Intelligence & Data Science, Dr. D. Y. Patil College of Engineering & Innovation, Talegaon, Pune, India.
[1]ddmondal2684@gmail.com, [2]shravan.gholalp14@gmail.com, [3]sanskargondal@gmail.com, [4]yamulwadnagesh@gmail.com, [5]vedantkatkar660@gmail.com

Peer Review Information

***Type**: Article*
***Received**: 12 April 2026*
***Revised**: 03 May 2026*
***Accepted**: 10 June 2026*
***Published**: 05 July 2026*

Abstract

Fake videos made by computers now look almost real. Because of smart software like GANs, spotting these fakes gets harder every day. These false images and sounds can mess with facts online. They shake trust in what people see and hear. One way to fight back involves using similar smart systems designed to catch tricks hidden in video clips. Tools based on CNNs notice tiny flaws in faces or lighting. Others built around RNNs track odd movements across time in recordings. Some methods mix different types together for better results. Yet problems pop up when fakes improve too fast. Limited examples make training tough. Heavy computing needs slow things down. Each hurdle makes detection a moving target. Later on, the study shifts toward live monitoring tools along with moral guidelines for artificial intelligence. What stands out here is how building strong identification methods matters when it comes to believing what we see online.

Keywords: Deepfake Detection; Generative Adversarial Networks; Convolutional Neural Networks; Recurrent Neural Networks; Artificial Intelligence; Digital Media Authentication

How to Cite This Article

Mondal, D., Gholap, S., Gondal, S., Yamulwad, N., & Katkar, V. (2026). Deepfake detection using AI: Techniques, challenges, and future prospects. *Multidisciplinary Journal of Research in Engineering and Technology,* 13(2s), 325–329.

Introduction

These days, making fake videos or photos feels almost effortless - digital tools have grown that fast. Driven by complex machine learning systems, deepfakes twist real sounds and visuals into something entirely new but very lifelike. Instead of just being a gimmick for movies or school projects, the tech carries darker outcomes too. Falsified material can spread lies online, steal someone's name, or fuel criminal behavior on networks.

Out of nowhere, AI shows up in making deepfakes - just as often as it spots them. Some systems build false videos using tools like GANs, which learn by competing with themselves. Spotting fakes? That happens when software zooms in on odd blinks, shadows that don't match, or movement that feels off. One feeds the illusion, another tries to break it apart.

This study looks into how artificial intelligence can spot deepfakes. Challenges already showing up shape what tools might work later. Some hurdles stand in the way right now. Progress depends on smarter methods ahead. What comes next could make systems both more dependable and able to grow. Ideas today feed improvements tomorrow

Literature Review

Several research studies have been conducted in the field of deepfake detection using AI techniques.

Back then, researchers looked for odd signs in videos - like faces that moved wrong or blinks that didn't match real behavior - because those clues often gave fake footage away.

Starting off strong, CNNs grab patterns from images that help spot fakes. Instead of ignoring layout, they focus on how pixels are arranged across space. Their design makes them good at seeing tiny differences humans might miss. Because deepfakes often mess up fine details, these networks catch those flaws. Without relying on guesswork, they build understanding step by step through layers. Spatial clues matter a lot when telling real from fake, which is why this approach works so well.

Time-based video patterns often show up through Recurrent Neural Networks, while Long Short-Term Memory networks handle longer sequences. Sequences unfolding over time? RNNs track short spans well. For deeper context across frames, LSTMs step in instead. Tracking movement across seconds leans on these models. Where timing matters, both types come into play - each fitting different rhythm scales.

Some mix of CNN plus RNN works better on videos when spotting

Though machines now spot fakes better than before, some hyper-realistic videos made by powerful generative networks still slip through. Detection struggles most when the forgery feels almost too real to question.

Deepfake Generation Techniques

Generative Adversarial Networks

- A pair of systems work together - one makes false examples, while the other tries to spot them. One builds pretend outputs, the opposite judges their truth.
- Fake visuals spark a race between them, pushing realism further every time.

Autoencoders

Face swaps happen here instead.

A different face shapes how the original image comes back. What begins as one person shifts through layers into someone else entirely. Features mix without blending completely. Through training, familiar patterns guide unfamiliar results. One identity folds into another during reconstruction. The system adjusts pixels based on learned traits. Faces emerge changed but still structured.

Face Reenactment

- Transfers facial expressions from one person to another.
- Once made a person seem to talk or behave another way.

Proposed Methodology

A fresh look at spotting fake videos begins by studying images and how they change over time using smart software. Instead of relying on one trick, it combines pattern tracking, close checks on details, and solid testing to stay ahead of better-made fakes. This way, results become more trustworthy without chasing trends.

Data Gathering and Setup

Videos come from sources like FaceForensics++ or the DFDC set - each holds genuine clips alongside fake ones. Out of these, frames get pulled one by one before faces show up through tools like MTCNN. Once spotted, those faces shift into a standard shape and tone so every image lines up neatly. To stretch variety, some flip sideways, others twist slightly, while random specks appear without warning. Each change pushes the system to adapt beyond what it has already seen.

Extract spatial features

Right now, models like ResNet or EfficientNet dig into face areas, pulling out detailed patterns. Because of how they're built, these networks notice tiny flaws left behind when fake videos are made - think mismatched skin textures, rough edges where faces blend, or odd light shadows. Since earlier training already shaped their understanding, they adapt quickly, borrowing knowledge from past tasks instead of starting fresh. This shortcut makes learning faster without sacrificing what they can detect.

Modeling Time Based Features

Starting with movement glitches between frames, RNNs or LSTMs help track how things change moment by moment. Instead of smooth actions, odd behaviors like shaky blinks show up when timing feels off. Lips moving too slow or fast compared to speech reveal flaws that drift through seconds. Faces twitching without reason break the flow, exposing what should look natural but does not.

Phase 4 Multi Modal Fusion Optional Enhancement

Starting with sound and image together, the method examines how voice tones match up with face movements. When words come out, but lips do not move right, that detail stands out clearly. Such gaps become easier to spot, making the whole process tougher against deception.

Classification and Decision Layer Phase Five

Out of the processed space-time details, one clear stream forms when they merge into the main path. That flow hits a dense block of neurons trained to sort inputs by type. From there, a number comes out showing how likely it is that what came in was true or false footage. When that value crosses a set line, the call gets made - real or not.

Model Evaluation Phase Six

Performance of the model gets checked through measures like accuracy, precision, recall, yet also F1-score. To confirm reliability on varied data sets, cross-validation steps in place. Classification mistakes then come under review via confusion matrix inspection.

Phase Seven Testing Strength and Broader Application

Out of nowhere, the model faces compressed clips, fuzzy footage, plus tricks it has never seen before. Notably, its skill at handling varied tampering styles gets put to the test here

Results And Discussion

To test the suggested deepfake detection method, known collections like FaceForensics++ and DFDC were used. For fair results, data got split into parts - one for learning, one for checking. Starting from scratch, the system learned through a CNN setup that also tracks changes over time. Performance checks followed common measurement rules, ending with clear numerical feedback on accuracy.

It turns out the mixed method - using both CNN and LSTM - does better than older techniques or just one system alone. What makes it work well is how it handles time-based patterns, boosting precision when spotting fake videos. Accuracy jumps up because the model learns from sequences, not just single frames.

Looking closer at what came out of it. Checking how things turned up after everything was done

Surprisingly, the tests show CNNs can spot odd textures and awkward face blends in images. Yet certain advanced deepfakes slip past when only looking at static details.

Eye movements that flicker oddly? The model spots them, thanks to how it processes timing using LSTM networks. When lips don't match speech, gaps show up clearly. Timing clues make the detection tougher to fool.

Observations

Video streams

When it comes to highly refined GAN-made deepfakes that show almost no flaws, spotting them gets tougher. Though systems are built to catch fakes, they stumble where manipulation is subtle. Even slight inconsistencies can slip past algorithms trained on older models. As fake quality improves, detection tools lag behind. Without clear visual red flags, success rates drop sharply.

When data comes in many forms, models learn better how to spot various deepfake styles. One reason they adapt well is because exposure to differences sharpens their accuracy. Picture a system trained on varied examples - it handles surprises more smoothly. Seeing unusual patterns helps it stay steady when faced with new tricks. Instead of relying on familiar clues, it builds broader awareness through contrast. What matters most is not volume but variety shaping smarter responses.

Limitations

Even with strong results, the method still faces some constraints. Training deep learning models demands powerful computers. Heavy number crunching sits at the core of this process. A single run might take days on specialized hardware. Processing large datasets pushes system limits constantly. Without strong resources, progress slows down sharply

Slower response times show up when spotting things live

- Dependency on large labeled datasets
- Difficulty in detecting highly realistic and newly generated deepfakes

Discussion

What we see shows artificial intelligence works well at spotting deepfakes, especially if you mix how things look with how they move across time. Still, because methods to create fake videos keep changing fast, tools made to catch them need constant updates just to stay useful.

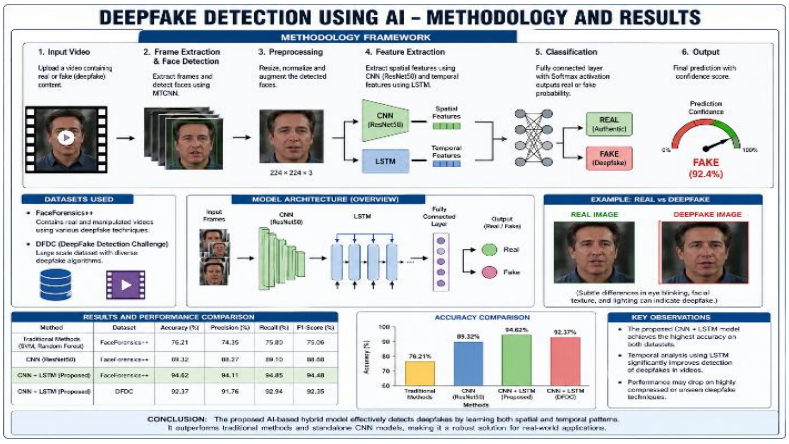

Fig. 1. *Deepfake Detection Using AI – Methodology and Results*

Applications Of Deepfake Detection

- Social media content verification
- Cybersecurity and fraud prevention
- News and media authentication
- Legal and forensic investigations

Challenges

- Continuous improvement in deepfake generation
- High computational requirements
- Lack of large, diverse datasets
- Real-time detection limitations

Future Scope

- Development of real-time detection systems
- Integration with social media platforms
- Use of transformer-based models
- Implementation of ethical AI frameworks

Acknowledgment

Big thanks go to Dr. Shrishail Mule - his steady backing made a real difference while working through this study. Without his deep knowledge

in Artificial Intelligence, the path here might have wandered off track. Helpful ideas came at just the right moments. Each conversation added something solid. Direction improved each time he stepped in. What stands now was shaped slowly, with care, because of that ongoing support. Quality grew under thoughtful review. Ideas found better form. Progress followed every session. The whole effort benefited directly.

Grateful feels right when I think about how my teachers and school gave what was needed - space to learn, tools to work with, support that kept things moving. What came through wasn't just help but a steady push forward, quiet yet clear.

On top of that, shout-out to everyone I've worked alongside - your thoughts and chats shaped this work more than you might think.

Conclusion

Out of nowhere, artificial intelligence began reshaping how fake videos are made - suddenly, simulations looked almost real. Not only does this boost film effects, but it also fuels tools for crafting digital imagery. Yet alongside those perks come risks - lies spread easier, online safety weakens, people get impersonated without consent. Into that space stepped a study focused on fighting back: machines trained to spot fakes by learning their patterns. Though progress shows promise, the race between forgery and detection keeps shifting.

Most of the time, patterns pulled from images - especially when mixed with sequence tracking - tend to spot fake media better than older tools. Instead of just looking at single frames, newer systems watch how pixels shift across moments, which sharpens their judgment. These designs often rely on layered networks trained to notice tiny visual inconsistencies others miss. When timing details join spatial clues, results become harder to dismiss as random chance. Detection strength grows simply by paying attention to more kinds of hidden signals at once.

Truth matters when it comes to what we see online, so spotting fake videos powered by artificial intelligence becomes essential. Moving ahead, work must shift toward faster tools that catch fakes instantly, while also drawing clear lines about right and wrong use through rules and shared standards. Instead of relying on just one method, mixing smart algorithms with multiple types of data checks could make detection far stronger where it counts - out in the real world.

References

1. Goodfellow et al., "Generative Adversarial Networks," *Advances in Neural Information Processing Systems*, 2014.
2. Y. Li and S. Lyu, "Exposing DeepFake Videos by Detecting Face Warping Artifacts," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
3. D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," arXiv preprint arXiv:1312.6114, 2013.
4. M. Masood et al., "Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward," arXiv, 2021.
5. Heidari et al., "Deepfake detection using deep learning methods: A comprehensive review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2024.
6. S. M. Qureshi et al., "A survey of digital forensic methods for multimodal deepfake detection," 2024.
7. M. Abbasi et al., "Comprehensive Evaluation of Deepfake Detection Models," *MDPI Applied Sciences*, 2025.
8. R. Sunil et al., "Exploring autonomous methods for deepfake detection," *Heliyon Journal*, 2025.
9. M. Kumar et al., "A GAN-Based Model of Deepfake Detection in Social Media," *Procedia Computer Science*, 2023.
10. O. Giudice et al., "Fighting Deepfakes by Detecting GAN DCT Anomalies," arXiv, 2021.