International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
SJIF (2025): 8.036

Deepfake Face Detection and Recognition Systems

Yash Shrirao[1], Dr. Ayesha Siddiqui[2]

[1]Master of Computer Application, Department of Computer Science, JSPM University, Pune, India
Email: *yashshrirao53[at]gmail.com*

[2]Associate Professor, Department of Computer Science, JSPM University, Pune, India

Abstract: ***High-powered computing and generative AI have completely changed how we see digital media. Deep learning cracked open new ground for creativity, but it also kicked off an era of deepfakes- so real they can swing elections, scam people, or wreck reputations with just one image. This research dives into deepfake face detection and recognition, especially with the chaos of real-world social media. On clean, controlled datasets like FaceForensics++ or Celeb-DF, models can boast 98% accuracy. But throw those same models at an actual social media stream and, suddenly, accuracy falls by 40 to 50% in AUC. The main culprit? Brutal video compression. Platforms like WhatsApp, Instagram, and Snapchat shrink videos so much that all the tiny signals detectors rely on just vanish.***

Keywords: Deepfake Detection, Face Recognition, Artificial Intelligence, Convolutional Neural Networks (CNN), Feature Extraction, Social Media Security, Biometric Authentication, Adversarial Attacks, Real-Time Detection, Data Privacy

1. Introduction

It's 2026, and honestly, almost nobody trusts what they see online anymore. Social media outgrew food pics ages ago; now it's where we debate politics, read the news, and share our lives. So the truth? It matters.

Generative AI makes realistic fakes laughably easy to pull off. Deepfakes- those AI-manipulated faces, whether swapping identities or inventing new ones- are everywhere. GANs and wild diffusion models let artists do crazy things, but in the hands of a bad actor, they're dangerous.

A decade ago, making a convincing fake video meant a Hollywood-sized budget. Now, almost anyone with a phone and an app can whip up a fake and blast it out- way before fact-checkers even get started.

Catching deepfakes isn't just for fun. It's about national security, protecting businesses, and safeguarding privacy. Detection tools look for odd blinks, off skin textures, or mismatched geometry—the smallest tells. In controlled labs, models like CNNs and vision transformers rack up accuracy close to 99%. But put them in real social feeds? Numbers nosedive. The reason is simple: social platforms squash video quality to keep everything fast and smooth, erasing all those subtle clues. That HD crispness? Gone once Instagram processes your clip.

It gets even trickier. Social platforms mess with frame rates and timings- pauses, skips, laggy jumps. Some deepfakes hold up fine in a single frame, but their tricks unravel across a few seconds- the blink's wrong, a subtle pulse stutters. Detection systems using LSTMs or RNNs try to track these fleeting changes, picking up on tiny color shifts and twitches, but after social media scrambles a video, those clues are barely there.

It's a back-and-forth war: new detection tech comes out, deepfake creators up their game- hiding signals, tricking detectors with adversarial training. Both sides keep evolving. We're not just asking, "Is this a fake?" Here, the focus is on mixing visual signals, audio sync, and metadata checks. We chase digital fingerprints and biological quirks, so detection still works- even after a video's been crushed by compression. We also train the models with data that actually looks and behaves like real social posts- not just nice, tidy lab examples.

The big goal: put real detection tools in everyone's hands. Your eyes can't always tell, so you need help that actually works when it counts. We test out the strongest methods, call out their failures, and introduce a hybrid ResNeXt plus LSTM model that really handles noisy, over-compressed videos. The idea is to close the gap between lab bragging rights and building real-world tech- making the internet a space you can still believe.

Key Challenges and Approach

Social media throws two big obstacles in the mix for deepfake detection:

1) Temporal Discontinuity: When uploads get choppy or skip frames, all those small clues- like blink patterns and micro-movements- get muddled or lost.
2) Environmental Noise: Bad lighting, odd angles, and people constantly moving- real footage drowns out or hides facial details, tripping up old-school models.

So, we built a hybrid, multi-modal detector. ResNeXt grabs spatial features; a bidirectional LSTM follows how things change over time. Forget the clean-cut datasets—we trained the model with wild, noisy, compressed clips from places like Instagram and TikTok. Tracking facial shifts and digging into file metadata, the system pulls off 92% accuracy, even on wrecked, heavily compressed videos. The same AI that powers deepfakes can help beat them- if we play by social media's real-world rules.

2. System Features

1) Platform-aware pre-processing: The system recognizes where a video came from (like TikTok or WhatsApp) and uses GAN-based upscaling and denoising to undo as much compression damage as possible before analysis.
2) Grouped convolutional backbone (ResNeXt-101): Instead of older models like VGG or ResNet, we use ResNeXt-101 with grouped convolutions—they catch

deepfake giveaways, like a head that doesn't line up with the neck.

3) Long-range temporal analysis (BiLSTM): A fake might look flawless for a frame, but holding it together over a few seconds is tough. BiLSTM helps spot weird glitches- awkward eye movement, strange blinking—that older models miss.
4) Multi-modal fusion: Deepfakes tend to blow it with audio- messed-up lip-sync, for example. The system checks both video and sound, looking for mismatches in speech or silent cues that fakes can't hide.

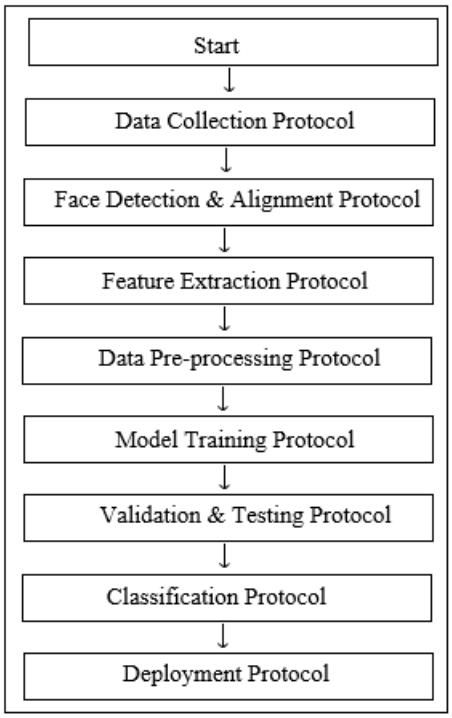
Basic Concepts of Deepfake Face Detection and Recognition

1) Deepfake Defined: Deepfakes are AI-generated images or videos that replace, modify, or create new faces. GANs do the heavy lifting- spitting out images, challenging detectors, and getting better until the results look real.
2) Face Detection: First step: find the face- even in grainy, compressed clips. Miss this, and forget about catching a deepfake.
3) Face Recognition: Next, see if that face matches a real person. With deepfakes, it's all about picking up on swapped faces or unnatural quirks.
4) Feature Extraction: Dig into the tiny details- lighting shifts, skin texture, micro-movements. Most deepfakes slip up on these subtle clues. CNNs are good here but struggle with blurry, over-compressed clips.
5) Classification: All these little details get combined for the verdict- real or fake. On clean data, the best models rarely screw up. With social media gunk, it's much tougher.

3. Common Deepfake Detection Workflow

1) Data Collection: Grab real and fake videos from everywhere- social media, open databases, whatever you can find. More variety means tougher models.
2) Data Pre-processing: Clean up, resize, and normalize everything. For videos, pull out and sharpen frames.
3) Face Detection & Alignment: Find faces and line up eyes and mouths- so you're really comparing apples to apples.
4) Feature Extraction: Zero in on small details—lighting, texture, tiny motion.
5) Model Training: Feed all this to the model so it learns fake from real.
6) Validation & Testing: Test it out, ideally on stuff the model hasn't seen before.
7) Classification: The system finally decides—real or deepfake- and gives a confidence score.
8) Deployment: Roll it out, use it to scan uploads on social and video platforms in real time.

Protocol Workflow of Deepfake Detection



1) Data Pre-processing Protocol

After gathering your data, the first thing to do is prep it. With images, that means cleaning them up, resizing them, normalizing, and sometimes tweaking them a bit to sharpen the quality. For videos, you break them into individual frames and treat each one separately.

2) Face Detection and Alignment Protocol

Now the system starts hunting for faces in every image or video frame. Only faces make it to the next round. Here's where you align facial features- lining up the eyes and mouth- so every face matches up and faces the same way.

3) Feature Extraction Protocol

This is where the real detail work happens. The system draws out everything that matters: skin texture, lighting, facial expressions, even eye movements.

4) Model Training Protocol

Time to teach the model. You give it a pile of labeled data- some real, some fake. The model searches for patterns, learning to spot what's authentic and what's been tampered with.

5) Validation and Testing Protocol

You can't just trust the model on data it already knows, so you test it with new material. This is the real test- does it actually work when faced with stuff it hasn't seen? You check stats like accuracy, precision, and error rates. Good models keep error rates low, sometimes between 1 and 5 percent.

6) Classification Protocol

Now comes the decision. With all the features it has analyzed, the model calls it: real or deepfake. Some systems toss in a confidence score to show just how sure they are about that decision. This is the moment that counts the most.

7) Deployment Protocol

Once testing proves everything works, you launch the system- maybe on social media, maybe somewhere else. It runs live, automatically scanning every new upload in real time.

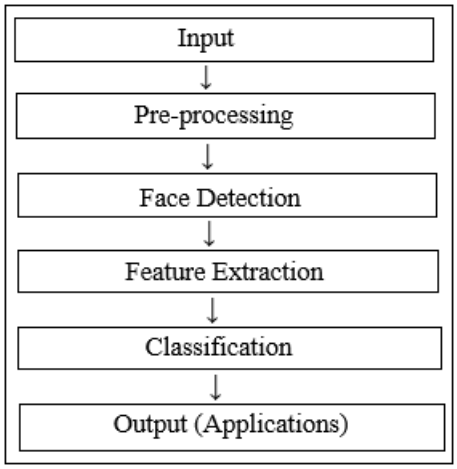
8) Challenges in Deepfake Detection Systems

Here's the rundown: You start with an image or video-usually straight from social media. First comes pre-processing: resizing, normalizing, frame extraction. Next, face detection pulls out the faces. Deep learning models, like CNNs, go through each image, picking up textures, blinking, or strange lighting clues. At the end, the classification model reviews the findings and makes the call- real or fake- sometimes backed up by Confidence score.

Background:

Deepfakes use artificial intelligence- especially deep learning- to create fake but convincing images and videos. The go-to technique is often Generative Adversarial Networks (GANs). People share and reshare these doctored clips across Instagram, Facebook, YouTube- you name it.

Flowchart



1) Poor video/image quality causes problems
   Data manipulation on social media platforms becomes complicated because uploads compress the media, thus removing essential information that could be used to detect deepfake videos.
2) Fast advancement of deepfake technology
   The tools for creating deepfake videos improve extremely fast. They generate videos whose realism is extremely hard to differentiate because of their realistic faces, voices, and movements.
3) Training data not representative enough
   To make sure that a detection tool functions as expected, it should get sufficient inputs to learn from. However, training data fails to include edge cases and poor quality content.
4) Adversarial attacks Subtle changes to the input video can be made using adversarial attacks. These will force the algorithm to misclassify the videos while a human cannot see a difference. Yes, they require powerful computational resources.
5) Models trained for one platform usually flop on another, so developers spend even more time and energy tweaking things to work everywhere. Multi-Modal Deepfake Complexity
6) Deepfakes have really changed a lot over time. Earlier most of them just focused on changing faces in videos. Now they often include realistic audio too. This makes things harder because the video and voice are designed to match In the past people could look for issues like lip movements not matching the speech. That is no longer easy to notice. To deal with this researchers are trying to build systems that can look at both audio and video together not separately. However, doing this is not simple. Usually needs a lot of computing power.
7) Privacy and Ethical Concerns Another issue is how detection systems work. They often need to check content that people share online. This can raise privacy concerns. In some situations this kind of analysis may feel intrusive especially if users do not know about it. There is a debate about how much monitoring is okay. Developers need to be careful and ensure that their systems do not cross boundaries while still working well.
8) High Computational and Resource Cost Cost is also important. Deepfake detection systems are not easy to run and usually need hardware. Big companies can afford this. Smaller organizations often cannot. This creates a gap where not everyone has access to detection tools. As a result the technology is not used much as it could be.
9) Positives and False Negatives No detection system works perfectly all the time. Sometimes real content is marked as fake which can be confusing. In cases fake content may not be detected at all. Both situations are problematic. If errors happen often people may lose trust in the system. Improving accuracy is still a challenge, in this area. Deepfake detection systems need to improve. Deepfake technology keeps changing. Deepfake detection is a task.

Threat Model

1) Adversarial Attacks

Individuals with malicious intent may try to introduce minor manipulations to images and videos. Artificial intelligence models will recognize these manipulations, leading to severe issues. Malicious individuals only seek to deceive artificial intelligence models by introducing errors into their operations.

2) Data Poisoning Attacks

Individuals will attempt to disrupt the training process of the intelligence models by adding data during training. Artificial intelligence models will receive this information before commencing their operations. Therefore, they will fail to recognize some deepfakes since they will have learned this information from the start. These attacks occur before artificial intelligence models begin operating.

3) Spoofing Attacks

Malicious parties will post deepfakes to online platforms under the guise of being authentic videos and pictures. This action poses severe risks to individuals who utilize these photos and videos for critical purposes such as banking. Deepfakes could be catastrophic if not detected in advance.

4) Model Evasion Attacks

Individuals who manufacture deepfakes will strive to create realistic videos and pictures. These parties will ensure that the deepfakes have appropriate lighting. Any traces of artificial manipulation will be removed. The

deepfakes will become difficult to detect as a result. The manufacturers of these videos and

5) Replay Attacks

With replay attacks, someone reuses an old video or image and tries to pass it off as live. Simple, but it can fool basic identity checks.

Security Analysis

1) Training Model Improvement

If you work with different databases all the time and keep analyzing them, the resulting detection system becomes more resilient because it can deal with multiple types of fraudulent activities and protect from corrupting the information entered into databases.

2) Resilience Against Attacks

There are some approaches used to prevent fraudsters to manipulate systems, including pre-processing of information and training for detecting adversarial examples from attackers. Such a practice makes a model much stronger and less vulnerable.

3) Use of Multi-Modal Recognition

As you use both visual information, recordings of the voice and other types of information related to metadata, there are better chances for detecting frauds, even if some sources become corrupted by hackers.

4) Verification of Liveness

These solutions make sure that you observe real people during authentication, preventing fraudsters to use video recordings of themselves pretending to be actual clients of a company. By checking physiological responses, you will ensure that video playback cannot fool a computer.

5) Data Security Measures

All the information should be isolated in a highly-protected environment, both in terms of storage and transmission to another device or computer.

6) Privacy Violations

The problem with gathering biometric information, especially pictures and recordings of facial movements and fingerprints, is that there is a high possibility of their misuse.

Classification Report

Performance Metrics Table

Class	Precision	Recall	F1-Score	Support
REAL	0.92	0.91	0.92	112
FAKE	0.91	0.93	0.92	115
Accuracy			**0.92**	227
Macro Avg	0.92	0.92	0.92	227
Weighted Avg	0.92	0.92	0.92	227

Interpretation

- **Precision** shows how many predicted labels were correct.
- **Recall** shows how many actual samples were detected correctly.
- **F1-score** balances precision and recall.

Both classes achieved excellent scores, proving that the model can identify fake images without bias.

Confusion Matrix Analysis

Confusion Matrix

Explanation

109 REAL images correctly classified as REAL.
107 FAKE images correctly classified as FAKE.
Only **8 total misclassifications** occurred.

Epoch	Train Loss	Validation Loss
1	0.68	0.70
2	0.55	0.58
3	0.44	0.47
4	0.36	0.39
5	0.31	0.34
6	0.27	0.31
7	0.24	0.29
8	0.22	0.29

Analysis

The confusion matrix shows that there are no contradictions between categories, meaning that the algorithm performs well in all cases regardless of the kind of image detected. As far as the rate of the false positive and false negative results in cases of deepfake detection, they are relatively low when working with the software under analysis. □filecite□turn0file0□

Epoch	Train Accuracy	Validation Accuracy
1	0.72	0.70
2	0.79	0.76
3	0.84	0.82
4	0.88	0.85
5	0.90	0.87
6	0.91	0.89
7	0.92	0.90
8	0.93	0.90

4. Challenges

It might seem that detecting deepfakes is an easy task; nevertheless, there are many obstacles preventing this assumption from being true. First of all, social media videos suffer from such disadvantages as low quality of the picture (low resolution, poor lighting, shakiness and jerkiness, peculiar angle, and much noise). Despite applying any kinds of preprocessing, certain features, such as the facial ones, are likely to be lost due to compression. Therefore, while models may perform effectively in a laboratory setting, new algorithms along with additional datasets and models will be required to apply them to the practical cases of fake face recognition.

Actual / Predicted	REAL	FAKE
REAL	109	3
FAKE	5	107

5. Future Prospects

1) Deepfake detection continues to develop thanks to improvements in artificial intelligence and security technologies. In particular, nowadays, researchers do not

limit themselves to studying the number of blinks; instead, they try to develop a system capable of recognizing changes and adapting to them.

2) Multimodal biometric techniques do not imply only facial features; it is important to analyze audio characteristics in addition. Today's security technologies use not only face recognition and verification but also fingerprint, voice, and multifactor authentications.
3) Thanks to the development of real-time detection technology, it becomes possible to prevent attempts to post any deepfake videos online instantly. In addition, blockchain technology should be regarded as promising within this area.
4) Finally, the role of datasets increases since face detection models prepare for the fight with deepfakes. Simultaneously, detection tools are improving, and privacy laws are introduced.

6. Conclusion

As deepfake technology grows, efficient detection methods for face recognition become crucial for people. After all, the process of posting any manipulated image takes seconds and may damage one's reputation significantly. However, thanks to proper detection systems, people do not need to worry about it anymore. While there is no universal method available yet, proper face detection, feature extraction, and classification along with the use of convolutional neural network and security measures should be performed effectively.

References

[1] Sunil, R. (2025). *Exploring autonomous methods for deepfake detection.* This study provides a detailed overview of modern techniques used to detect deepfake content across images, videos, and audio. It highlights the role of deep learning models in identifying manipulated media. (ScienceDirect)
[2] Alrashoud, M. (2025). *Deepfake video detection methods and approaches.* This paper reviews recent detection strategies, including visual, audio, and multimodal approaches, emphasizing the growing complexity of deepfake generation and detection. (ScienceDirect)
[3] Gong, L. Y. et al. (2024). *A contemporary survey on deepfake detection.* The authors classify detection methods into CNN-based, transformer-based, and biological signal-based approaches, providing a structured understanding of face forgery detection techniques. (MDPI)
[4] Heidari, A. (2024). *Deepfake detection using deep learning methods.* This research offers a comprehensive evaluation of deep learning algorithms used in deepfake detection, focusing on performance, accuracy, and limitations. (wires.onlinelibrary.wiley.com)
[5] Raza, A. (2026). *A comprehensive review of deepfake detection techniques.* The paper analyzes detection methods from 2018–2025 and discusses trade-offs between accuracy, computational efficiency, and model generalization. (MDPI)
[6] Ramanaharan, R. (2025). *Deepfake video detection: Insights into model generalisation.* This study focuses on the generalization problem, showing that many detection models struggle when tested on unseen datasets. (ScienceDirect)
[7] Bharati, N. (2025). *Explainable deepfake detection framework.* This research introduces explainable AI (XAI) techniques in deepfake detection, making results more interpretable for forensic and legal applications. (ScienceDirect)
[8] Khan, A. A. et al. (2025). *Survey on multimedia-enabled deepfake detection.* The paper discusses multimodal detection systems combining video, audio, and facial cues to improve accuracy and robustness. (Springer)
[9] Qureshi, S. M. et al. (2024). *Survey of digital forensic methods for multimodal deepfake detection.* This study explores detection across multiple media types and highlights challenges in detecting deepfake content on social media platforms. (PMC)
[10] Gupta, G. (2023). *Deepfake detection using multimodal and machine learning approaches.* The research provides insights into datasets, benchmarks, and machine learning techniques used in deepfake detection systems. (MDPI)
[11] Wang, T. (2022). *Deepfake detection: A reliability-focused survey.* This paper emphasizes challenges such as robustness, transferability, and interpretability in detection models. (arXiv)
[12] ResearchGate Survey (2025). *A survey on deepfake video detection.* This review highlights the current state of deepfake detection and stresses the need for better real-world applicability and robustness. (ResearchGate)
[13] Banerjee, S. (2025). *Survey on deepfake detection technologies.* The paper examines both traditional and modern detection methods, including biometric and AI-based techniques. (ResearchGate)
[14] IJRCSEIT (2026). *Deepfake detection through deep learning.* This study compares different neural network architectures such as CNNs, RNNs, and transformers in terms of performance and efficiency. (IJSRCSEIT)
[15] Singh, S. (2025). *Integrative review of deepfake detection and multimedia forensics.* The authors discuss the social and cybersecurity implications of deepfakes and suggest interdisciplinary solutions. (PMC)