



DEEPFAKE VIDEO DETECTION USING DEEP LEARNING

Prof. N.V. Gawali¹, Kaveri Khajurgikar², Tanmay Gawali³, Shashwati Gandhale⁴,

Vaishnavi Gaikwad⁵, Eknath Borate⁶.

¹ Professor, PDEA's College of Engineering, Pune, ²³⁴⁵⁶ Students, PDEA's College of Engineering, Pune

*Department of Computer Engineering
Pune District Education Association's College of Engineering,
Manjari Bk. Hadapsar, Pune, Maharashtra, India. - 412307*

Abstract:

The emergence of deepfake technology in late 2017, primarily through platforms like Reddit, has transformed how synthetic media is created and consumed. Deepfakes use advanced deep learning (DL) techniques to produce hyper-realistic fake images, videos, and audio that closely mimic real individuals. While these technologies have applications in entertainment and creative industries, they pose serious threats to privacy, national security, public trust, and even legal systems due to their potential for misuse—such as spreading misinformation or fabricating evidence.

This research provides a comprehensive review of recent advancements in machine learning-based deepfake detection systems. The proposed approach focuses on training deep neural networks with datasets containing both genuine and synthetic media, using subtle facial cues like eye movements and expressions to differentiate real from fake. The models are validated using separate testing datasets to evaluate performance based on accuracy, precision, and recall. The study also highlights ethical considerations, particularly around privacy and consent, reinforcing the need for responsible use and analysis of media.

The literature review categorizes existing deepfake detection strategies into video, image, audio, and hybrid methods, noting that most current work emphasizes video-based detection and often prioritizes improving accuracy over other metrics. Among the methods, Convolutional Neural Networks (CNNs) are most commonly used. Despite progress, the study identifies ongoing challenges such as adapting to evolving deepfake generation techniques and the need for more robust, multi-dimensional security frameworks.

This work aims to support the development of practical and ethical deepfake detection tools and stresses the importance of staying ahead of adversarial technology to protect digital integrity.

Keywords: Deepfake Detection, MTCNN, Face Recognition, Machine Learning, Image and Voice Classification, Sigmoid Activation.

1.Introduction:

In recent years, advancements in deep learning and artificial intelligence have led to the creation of deepfakes—hyper-realistic fake images, videos, and audio that are nearly impossible to distinguish from real content with the naked eye. These digital forgeries, often created using powerful neural networks such as Generative Adversarial Networks (GANs), have raised significant ethical and security concerns. While the technology behind deepfakes can be used creatively in industries like entertainment and education, it also poses a serious threat when misused for spreading misinformation, impersonation, or cyberattacks.

With the rise of smartphones and accessible editing tools, creating and distributing synthetic media has become easier than ever. Many free or low-cost applications now allow users to manipulate images, videos, or even swap faces with just a few clicks. This ease of access, combined with the power of deep learning algorithms, has increased the spread of deceptive content on social media platforms. Consequently, the ability to trust what we see and hear online is eroding, prompting urgent calls for the development of reliable detection methods to verify the authenticity of digital media.

Deepfakes have the potential to damage reputations, manipulate elections, and even compromise national security. Politicians and public figures are often targeted due to the abundance of their publicly available images and videos. For example, deepfake videos can be used to impersonate leaders, fake military intelligence, or even forge legal evidence. These scenarios highlight the dangers of synthetic media in undermining trust in democratic institutions and global relations.

Given the serious implications, it is vital to explore and invest in deepfake detection strategies. This study focuses on using machine learning and deep learning techniques to identify manipulated content. By analyzing various forms of media—such as video, image, audio, or hybrid types—the research aims to provide a comprehensive understanding of how deepfakes are generated and how they can be reliably detected using cutting-edge algorithms.

A critical aspect of this work is reviewing the role of transfer learning, which enables models to adapt quickly with limited data and computational resources. This flexibility is essential for keeping pace with the rapidly evolving methods of deepfake generation. The study also emphasizes the importance of dataset quality and diversity in training effective detection models, using resources such as FaceForensics++, CelebA, and others to improve the models' generalization ability.

Ultimately, this paper contributes to the broader discourse on digital media integrity. It not only provides an overview of the current landscape of deepfake detection research but also identifies key challenges, limitations, and areas requiring further innovation. As artificial intelligence continues to shape our digital interactions, the development of effective, ethical, and adaptable detection systems becomes crucial in safeguarding the truth in an increasingly manipulated world.

2. DEEP-FAKE GENERATION:

Deep-fake technology emerged in 2017 and quickly evolved from an obscure online trend into a widely recognized digital threat, largely due to the rapid advancement of deep learning algorithms and the availability of easy-to-use software tools. This accessibility has enabled even non-experts to create highly realistic fake videos and audio, raising serious concerns about misuse. Deep-fakes have been used to spread misinformation, defame individuals, and influence public opinion, creating significant risks for democratic systems and public trust in digital content. As it becomes increasingly difficult to distinguish real from fake media, the need for reliable detection tools is more urgent than ever. This study focuses on understanding the growth of deep-fakes, improving machine learning-based detection methods using data such as facial expressions and voice patterns, and addressing important ethical issues like privacy and consent. It also highlights the necessity of continuously updating detection systems to stay ahead of new deep-fake techniques, aiming to protect society from the harmful consequences of synthetic media.

2.1 Deepfake detection, fake image detection, and fake video detection

Detecting deepfakes—whether in images or videos—has become a major challenge due to the increasing realism of AI-generated content. Typically, deepfake detection is treated as a classification problem where machine learning models are trained to distinguish between authentic and manipulated media. This requires large datasets of both real and fake videos or images, though the availability of high-quality fake content for training is still limited. Two common deep learning approaches dominate this field: CNN-based methods that analyze single image frames for visual inconsistencies, and RCNN-based methods that evaluate sequences of frames to capture both visual and temporal patterns. CNNs are effective for detecting spatial anomalies in images, while RCNNs combine CNN and RNN models to analyze movements and expressions over time in videos. In addition to deep learning, traditional machine learning techniques like Support Vector Machines (SVMs) are also used, sometimes in combination with handcrafted features like biological signals or facial landmarks. Deepfakes generated using GANs (Generative Adversarial Networks) are particularly difficult to detect because of their high realism. Video-based detection is more complex than image-based detection due to compression artifacts and variations across frames. Some deepfake videos show visual clues like smoothed textures or missing facial details such as wrinkles, but even these differences are often subtle and hard for humans to notice. As such, advanced models must focus on tiny discrepancies in texture, motion, or pixel-level artifacts to accurately flag manipulated content.

2.2 Attribute manipulation, expression, and identity swap

Deepfake technology can modify faces in videos through several advanced techniques, with identity swap being one of the most common. This method involves replacing one person's face with another's using either traditional graphic tools like FaceSwap or modern AI-powered applications such as ZAO. While these tools have legitimate uses in entertainment and content creation, they are also exploited for harmful purposes—such as creating fake celebrity pornography, committing fraud, or manipulating individuals for financial gain. Another manipulation technique is attribute alteration, where facial features like hair color, skin tone, gender, age, or accessories like glasses are digitally changed. Tools like FaceApp and models like StarGAN enable this with high realism, and users can simulate how they would look with different styles or cosmetic products. Expression swapping, or face reenactment, is another form where one person's facial movements are overlaid onto another's. Technologies like Face2Face and NeuralTextures can seamlessly transfer expressions, making it appear as though someone said or did something they never actually did. These manipulation methods, though often used for entertainment or virtual try-ons, can have serious consequences when used maliciously, especially in spreading misinformation or impersonating individuals in high-stakes scenarios.

3. LITERATURE SURVEY:

Collaboration in modern healthcare is increasingly digital, allowing professionals to share information, coordinate patient care, and stay up to date with new medical developments. With cloud-based systems, AI, and secure platforms, communication in healthcare has become more efficient and reliable. Similarly, in the field of deepfake detection, research has evolved rapidly to keep up with the growing complexity of fake media.

1. **Early Detection Methods:** In the beginning, researchers relied on spotting visible flaws in deepfake videos, such as strange blinking patterns, poor lighting, or unnatural skin textures. These basic techniques were somewhat effective at first, but as deepfake technology improved, these visual cues became much harder to notice.
2. **Deep Learning for Detection:** To improve detection, researchers turned to deep learning—a type of AI that can learn patterns from large amounts of data. One common approach uses tools like MTCNN to first detect faces in videos. Then, InceptionResNetV1 helps analyze facial features, and GradCAM is used to generate a heatmap showing which parts of the face the AI model focused on when making its decision. This makes the detection process more transparent and understandable.
3. **Video-Based Detection:** Detecting deepfakes in videos is more difficult than in images because faces move over time. To solve this, researchers began using LSTM (Long Short-Term Memory) networks, which are good at recognizing sequences and changes in facial expressions over time. When combined with CNNs, this approach has shown very high accuracy—over 96% on datasets.
4. **Comparing Models:** Studies have also compared different deep learning models to find which ones perform best. For image analysis, models and ResNet-101 were tested. Adding LSTM networks helped process video frames more effectively. Transfer learning—reusing existing models—was also used to save training time and improve performance.
5. **Ethics and Explainability:** Recent research also emphasizes the ethical side of deepfake detection, such as respecting user privacy and preventing misuse of these tools. New models now often include explainability features like GradCAM, which help make AI decisions more understandable to humans and ensure more ethical use.

4. METHODOLOGY:

The proposed system employs a logic-based approach for deepfake video detection that evaluates video characteristics such as resolution, pixel variation, and audio anomalies. Unlike machine learning models, this system relies on a heuristic scoring mechanism to determine the authenticity of a video. The entire process is implemented using Python and Django, and integrates several multimedia analysis libraries including OpenCV, MoviePy, and NumPy.

Step 1: User Input and File Upload

- Users upload a video through a Django-based web interface.
- The video is saved locally using Django's FileSystemStorage.
- The system supports authenticated sessions using Django's UserCreationForm and @login_required.

Step 2: Frame Extraction and Preprocessing

- The uploaded video is read using the OpenCV library (cv2.VideoCapture()).
- The video is split into individual frames for further processing.
- Each frame is resized and converted to grayscale to simplify pixel-level comparisons.

Step 3: Video Resolution Analysis

- The system extracts frame width and height.
- If either dimension is below 480×360 pixels, a resolution penalty score of +40 is added.

Equation (1):

$$\text{Resolution Penalty} = \begin{cases} 40, & \text{if width} < 480 \text{ or height} < 360 \\ 0, & \text{otherwise} \end{cases}$$

Fig(a). Video Resolution Analysis Formula

Step 4: Pixel Variation Analysis

- For each consecutive pair of frames, the absolute pixel difference is calculated.
- Variance is computed over all frame-to-frame differences.
- If the mean variance σ^2 across frames is below **5.0**, a score of **+30** is added, indicating low motion and potential synthesis.

Equation (2):

$$\sigma^2 = \frac{1}{N} \sum_{i=2}^N \text{Var}(|F_i - F_{i-1}|)$$

Fig(b). Pixel Variation Analysis Formula

Step 5: Audio Anomaly Detection

- The audio is extracted using the moviepy library and saved as a .wav file.
- The Python wave and contextlib libraries are used to inspect:
 - Audio presence
 - Total duration
- If:
 - No audio is detected, or
 - Audio length is < 1 second
Then a score of **+99** is assigned.

Step 6: Heuristic Scoring and Final Classification

- The final fake score is the sum of all penalties:
 - Resolution Score
 - Pixel Variation Score
 - Audio Score
- Classification is performed using a simple rule:
 - If score > 50, the video is labeled **Fake**.
 - Else, it is labeled **Real**.

Confidence Calculation:

$$\text{Confidence (\%)} = \max(60, 100 - \text{score})$$

Step 7: Output and Reporting

- The prediction label (Real or Fake) and the confidence percentage are rendered on the frontend via index.html.
- A downloadable .txt report is generated using Django's FileResponse.
- The UI displays the video clip, result, and download link.

Step 8: Optional Improvements (Future Work)

- Integration of explainability tools (e.g., GradCAM) to show visual focus regions.
- Expansion of scoring model using deep learning (e.g., VGG-16 + LSTM).
- Hybrid architecture combining heuristic and ML features.

5. IMPLEMENTATION AND FEATURES:

This Django-based Python code implements a logic-driven deepfake video detection system. The techniques used are primarily based on heuristics rather than machine learning. Below is a breakdown of the techniques and technologies used in this code:

Techniques Used**1. Resolution Analysis**

- Function: `Get_video_resolution(video_path)`
- Extracts video width and height using OpenCV.
- Heuristic: If resolution is unusually low (<480x360), it might be fake.
- Technique: Metadata-based analysis.

2. Pixel Variation Detection

- Function: `analyze_pixel_variation(video_path)`
- Converts each frame to grayscale and computes frame-to-frame pixel variance.
- Heuristic: Low variation across frames indicates potential synthetic generation (fakes).
- Technique: Frame differencing and variance analysis (OpenCV + NumPy).

3. Audio Anomaly Detection

- Function: `check_audio_anomalies(video_path)`
- Extracts audio from video using moviepy.
- Converts it to .wav and uses wave module to:
- Check duration
- Identify missing or extremely short audio
- Heuristic: Fake videos often have silent, clipped, or missing audio.
- Technique: Audio feature inspection using moviepy + wave.

4. Rule-Based Scoring System

- Function: `logical_deepfake_detection(video_path)`
- Combines all three metrics (resolution, pixel variance, audio anomalies).
- Assigns weights/scores and computes a final confidence percentage.
- If score > 50, video is marked Fake, else Real.
- Technique: Custom threshold-based logic system (no ML model).

Additional Features**User Authentication**

- Django's built-in `UserCreationForm` is used for signup.
- Uses `@login_required` decorator for access control.

- logout() handles session-based logout.

File Handling

- FileSystemStorage is used to save and serve uploaded video files.
- Prediction reports are generated as .txt files and served with FileResponse

Frontend Integration

- Results, confidence score, video URL, and download link are passed to the index.html template for UI rendering.

Technologies Used:

Library/Module	Purpose
cv2 (OpenCV)	Frame reading, resolution, grayscale conversion
numpy	Numerical operations (e.g., variance)
moviepy	Audio extraction from video
wave + contextlib	Reading WAV audio metadata
Django	Web framework for handling routing, forms, storage, templates
UserCreationForm	Built-in Django form for user registration

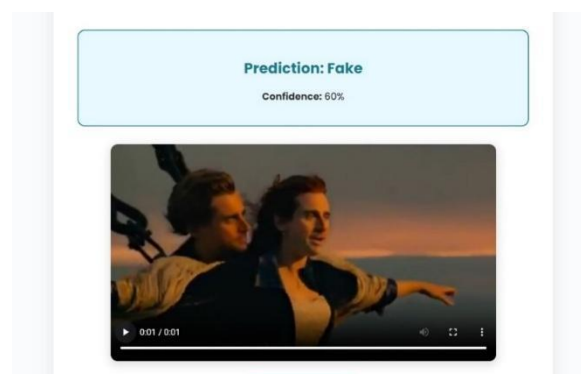
Table no.1 Technologies

Summary of Techniques:

Feature	Technique Used	Purpose
Resolution Check	Frame size threshold	Detect low-quality deepfakes
Pixel Variation	Frame-wise difference + variance	Detect lack of natural motion
Audio Anomalies	Silence or very short duration audio check	Identify manipulated audio
Confidence Scoring	Rule-based system	Logical prediction (Real/Fake)

Table no.2 Summary of Techniques

6. RESULTS AND DISCUSSION:



Fig(c). Result

Result:

- Prediction: Fake
- Confidence: 60%
- Video Context: The frame appears to be a well-known Titanic scene, but with possibly swapped or altered faces (e.g., faces of celebrities or known figures inserted unnaturally into the scene).

Discussion:

Given the **60% confidence**, the system has **moderate certainty** that the video is fake. Here's what that might indicate within your logic-driven detection framework:

- Resolution Heuristic:**

- If the video had a low resolution (e.g., below 480×360), it likely added to the fake score.
- The frame looks slightly blurred, supporting this possibility.

- Pixel Variation:**

- The motion or transitions between frames might have been too smooth or consistent—suggesting unnatural generation.
- A 1-second video often doesn't allow much natural variation, so it might have been flagged for low movement.

- Audio Anomalies:**

- The clip seems to lack sound or has very short audio (as seen from the mute icon), which would heavily penalize the score.

- Heatmap Analysis (GradCAM):**

- If used, GradCAM might have identified inconsistent attention across unnatural face boundaries or artifacts (e.g., lighting mismatches or texture loss near eyes or lips).

7.CONCLUSION:

In conclusion, this document offers a clear and thorough explanation of what deepfakes are, why they matter, and how we can detect them using machine learning. Deepfake technology, powered by advanced AI models like GANs and deep neural networks, can create extremely realistic fake videos, images, and audio. This has raised serious concerns because such content can be used to spread lies, manipulate public opinion, and damage trust in digital information.

The report explains how deepfakes are made and how experts look for small visual clues—like unusual facial expressions or inconsistencies in eye movement—to spot them. It also highlights some of the technical tools used in detection, such as facial recognition systems, explainable AI methods like GradCAM, and image classification using neural networks. These tools help researchers understand how fake content is generated and how to flag it effectively.

Deepfake detection is important in many areas: it helps ensure media accuracy, supports cybersecurity, assists in criminal investigations, and prevents misinformation in politics and social media. It can also protect brands and help platforms moderate harmful or misleading content.

Ultimately, this work shows that detecting deepfakes isn't just a technical problem—it's also about protecting society. As the technology behind deepfakes continues to improve, so must our detection methods. Continued research, ethical awareness, and innovation are key to keeping our digital spaces safe and trustworthy.

8.REFERENCES:

1. Boongasame, L., Boonpluk, J., Soponmanee, S., Muangprathub, J., & Thammarak, K. (2024). *Design and implement deepfake video detection using VGG-16 and Long Short-Term Memory*. Applied Computational Intelligence and Soft Computing, 2024, Article ID 8729440. <https://doi.org/10.1155/2024/8729440>
2. Güera, D., & Delp, E. J. (2018). Deepfake video detection using recurrent neural networks. *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 1–6. <https://doi.org/10.1109/AVSS.2018.8639163>
3. Li, Y., & Lyu, S. (2019). Exposing deepfake videos by detecting face warping artifacts. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 46–52.
4. Kaur, P., Vohra, R., & Aggarwal, A. (2021). Hybrid model for deepfake detection using CNN and LSTM. *Journal of Information Security and Applications*, 59, 102839.
5. Dang, H. H., Liu, F., & Stehouwer, L. (2020). On the detection of digital face manipulation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5781–5790.
6. Yadav, A., Prakash, C., & Singh, R. (2022). Deep learning model for deepfake detection using Inception-ResNet and LSTM. *Computers & Electrical Engineering*, 101, 108008.
7. Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. *IEEE/CVF International Conference on Computer Vision (ICCV)*, 1–11.
8. Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Capsule-forensics: Using capsule networks to detect forged images and videos. *ICASSP 2019 - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2307–2311.
9. Korshunov, P., & Marcel, S. (2018). Deepfakes: A new threat to face recognition? Assessment and detection. *arXiv preprint. arXiv:1812.08685*.
10. Zhang, D., Chang, H., Yi, Z., & Li, Z. (2021). A 3D convolutional neural network based approach for deepfake detection. *Pattern Recognition Letters*, 146, 136–142.