International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 1772-1777
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0231

Deep Fake Detection System

Ms.Ch.Jeevana Priya[1], Komarvalli Pravallika[2], Nadakuditi Bhavana[3], Kona Jyoshnavi[4], Chinthakindi Kaveri[5]
[1]Assistant Professor, Department CSE-Artifical Intelligence and Machine Learning, SRK Institute of Technology, Vijayawada,India
[2,3,4,5]Students, Department CSE-Artifical Intelligence and Machine Learning, SRK Institute of Technology,Vijayawada,India
***Emails**: jeevanapriya710@gmail.com[1], aksapravallika@gmail.com[2], bhavananadakuditi25@gmail.com[3], 4konajyoshnavi7@gmail.com[4], kaverich01@gmai.com[5]*

Abstract

Deepfake technology uses advanced artificial intelligence and deep learning techniques to generate highly realistic synthetic images. While this technology has useful applications in areas such as entertainment and media production, it can also be misused to manipulate digital content and spread misleading information. Because of this, identifying deepfake images has become an important challenge in digital media security. This project presents an AI-based Deepfake Detection System designed to distinguish between real and manipulated images. The proposed approach utilizes a Convolutional Neural Network (CNN) to analyze facial features and extract spatial patterns from images. The model learns to recognize visual inconsistencies such as texture distortions, blending artifacts, and unnatural facial characteristics that often appear in manipulated images. By analyzing these patterns, the system classifies images as real or fake and provides a confidence score for the prediction. The experimental results show that the CNN-based approach can effectively detect deepfake images and improve the reliability of digital media verification.

Keywords: *Deepfake Detection, Convolutional Neural Network (CNN), Deep Learning, Image Forensics, Artificial Intelligence, Media Authentication, Digital Image Manipulation Detection.*

1. Introduction

Deepfake technology uses artificial intelligence and deep learning techniques to generate very realistic fake images. Even though it is useful in areas such as entertainment and digital media, it can also be misused for activities like spreading false information, manipulating identities, and committing online fraud. As the quality of deepfake images improves, identifying them manually becomes very challenging. Therefore, automated detection systems based on deep learning are necessary. In this project, a Deepfake Detection System is developed using a Convolutional Neural Network (CNN). The CNN model analyzes image features such as facial patterns and possible digital irregularities. By examining these patterns, the system determines whether an image is real or fake with improved accuracy.

2. Literature Survey

2.1. Early Deepfake Image Detection Techniques

Early deepfake detection methods focused on identifying visible artifacts and inconsistencies present in manipulated images. One of the initial approaches used convolutional neural networks to detect manipulation traces in facial regions. The MesoNet architecture was introduced as a lightweight CNN model designed specifically for detecting facial forgery patterns in images [1]. Similarly, researchers observed that deepfake generation often produces face warping artifacts due to imperfect image synthesis, which can be used to identify manipulated images [2]. These early approaches demonstrated that spatial artifacts present in fake images could be effectively detected using deep learning techniques.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 1722-1777
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0231

2.2. Deep Learning–Based Image Forgery Detection

With the advancement of deep learning, several CNN-based architectures were proposed to improve deepfake image detection performance. The FaceForensics++ dataset provided a large benchmark dataset containing manipulated facial images for training and evaluating deep learning models [3]. Capsule Networks were later introduced to preserve spatial relationships between facial features, improving the ability to detect subtle manipulation patterns [4]. The Face X-Ray technique further enhanced detection by analyzing blending boundaries between real and manipulated regions in images [5]. In addition, statistical feature-based methods were developed to identify inconsistencies between real and synthetic images [6].

2.3. Large-Scale Datasets for Deepfake Detection

The availability of large datasets has played a crucial role in advancing deepfake detection research. The Celeb-DF dataset introduced more realistic manipulated facial images that improved model training and evaluation [7]. Similarly, the DeepFake Detection Challenge (DFDC) dataset provided a large collection of real and fake images and videos for benchmarking deepfake detection models [8]. These datasets helped researchers develop more robust models capable of detecting deepfake images in real-world scenarios.

2.4. Frequency-Based and Hybrid Detection Approaches

Recent studies have explored frequency-domain features to detect deepfake images. Researchers observed that synthetic images often lack natural frequency distributions found in real images. Frequency-aware detection techniques were introduced to analyze these inconsistencies in manipulated images [9]. Adversarial training methods were also proposed to improve the generalization capability of detection models [10]. Additionally, two-stream neural networks were introduced to combine spatial features and frequency features, resulting in improved detection accuracy [11]. Other approaches used frequency analysis and spectral features to detect subtle artifacts left by generative models [12].Advanced Deepfake Detection and Security Approaches

Recent research focuses on improving detection robustness and understanding how deepfakes are generated. Some studies proposed methods to detect deepfakes by analyzing head pose inconsistencies and abnormal facial geometry [15]. Other approaches investigated image manipulation detection through self-consistency learning [16]. Generative models such as StyleGAN demonstrated the ability to produce highly realistic synthetic images, highlighting the need for stronger detection techniques [17]. Researchers also discovered that generative models leave unique fingerprints that can be used to attribute fake images to specific GAN architectures [18]. Further studies showed that CNN-generated images contain identifiable artifacts that can be detected using deep learning models [19]. Recently, spatial attention mechanisms have been applied to improve deepfake image detection performance by focusing on important facial regions[20].

3. Related Work

Deepfake detection has become an important research area due to the rapid growth of generative models that create highly realistic fake images. Early detection methods focused on identifying visual artifacts and facial inconsistencies such as face warping, unnatural textures, and blending errors. With the advancement of deep learning, Convolutional Neural Network (CNN) based models became widely used for extracting spatial features from manipulated images. The availability of large-scale datasets also helped improve the training and evaluation of these models. Modern approaches analyze facial patterns, digital artifacts, and texture inconsistencies to identify manipulated images. Although significant progress has been achieved, detecting highly realistic deepfake images remains a challenging task. Therefore, deep learning-based CNN architectures are widely used to improve accuracy and provide reliable detection of Real and Fake images in real-world scenarios.

4. Existing System

Existing deepfake detection systems mainly use deep learning models to classify images as real or fake. Most approaches rely on Convolutional Neural

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 1722-1777
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0231

Networks (CNNs) to analyze facial images and detect spatial artifacts such as texture distortions, face warping, and blending inconsistencies. Some advanced methods also use capsule networks or frequency-based analysis to identify manipulated image patterns. However, these systems still have several limitations. Many CNN-based models struggle to detect highly realistic deepfake images, especially when the images are compressed, low resolution, or captured under poor lighting conditions. Because of these challenges, accurately detecting advanced deepfake images remains a difficult task.

5. Methodology

The proposed Deepfake Detection System follows a systematic approach to classify images as real or fake using deep learning techniques. Initially, a dataset containing labeled real and fake images is collected from publicly available sources. This dataset is carefully organized to ensure a balanced distribution of both classes, which helps the model learn effectively. Before feeding the data into the model, preprocessing steps are applied to improve input quality. These steps include face detection, cropping the relevant facial region, resizing images to a fixed size, and normalizing pixel values. This process ensures that all images have a consistent format and enhances the model's learning capability. After preprocessing, the images are provided as input to a Convolutional Neural Network (CNN). The CNN automatically learns important visual patterns such as edges, textures, and subtle inconsistencies present in manipulated images. These learned patterns are then transformed into meaningful representations that can be used for classification. The model is trained using labeled data with a suitable loss function and optimization algorithm to minimize errors during learning. The dataset is divided into training and validation sets to monitor performance and avoid overfitting. Evaluation metrics such as accuracy, precision, recall, and F^1-score are used to measure the effectiveness of the model. Finally, the trained model is integrated into a web-based system where users can upload images for analysis. The system processes the input image and produces a prediction result along with a confidence score, indicating whether the image is real or fake. This approach ensures reliable, efficient, and scalable deepfake image detection.

6. System Architecture

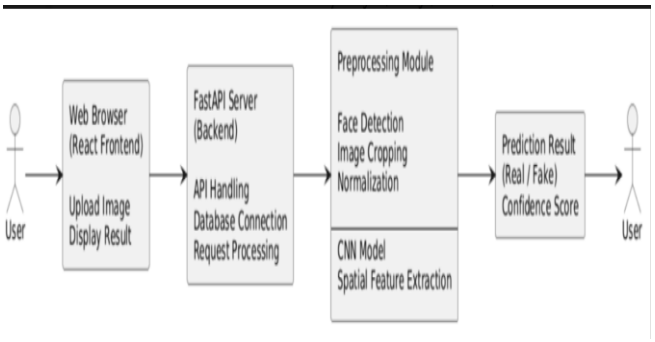


Figure1 System Architecture

In Figure 1 the user uploads an image through the web browser (React frontend). The image is then sent to the FastAPI backend server, which handles the request and processes the data. After that, the preprocessing module detects the face in the image, crops it, and normalizes it for better analysis. The processed image is then given to the CNN model, which extracts important visual features from the face. Finally, the system analyzes these features and gives the prediction result as Real or Fake with a confidence score, which is displayed to the user.

7. System Implementation

User Module

Interacts through a web interface Uploads image Receives prediction result (Real / Fake)

System Modules

- Input: Frontend (React.js): Accepts user image upload Backend (FastAPI): Receives image file and sends it to preprocessing module
- Processing: Preprocessing Module: Detects face, crops the face region, resizes image to fixed resolution, and normalizes pixel values. CNN Model: Extracts spatial features such as facial textures, artifacts, and inconsistencies Classification Layer: Produces final prediction based on extracted features
- Output Prediction Result: Classifies image as Real or Fake Frontend Display: Shows result to user in a simple and easy-to-understand format The User Module is responsible for

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 1722-1777
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0231

interaction between the user and the Deepfake Detection System. It provides a simple and user-friendly web interface where users can upload images for analysis. The user selects an image file and submits it to the system for processing. After the image is analyzed by the model, the system displays the prediction result indicating whether the image is Real or Fake. It also shows a confidence score to help the user understand the accuracy of the result. This module ensures easy access and smooth communication between the user and the system.

7.1. Home Page Of The Deepfake Detection System

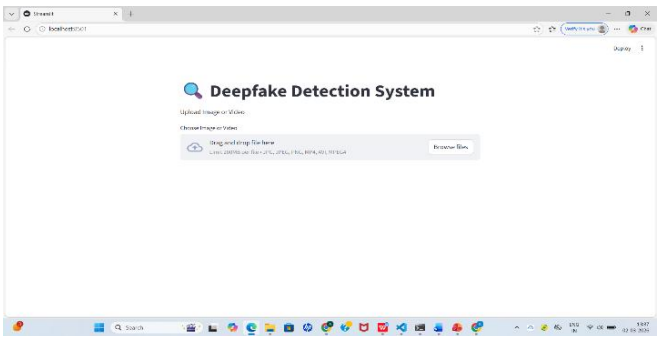
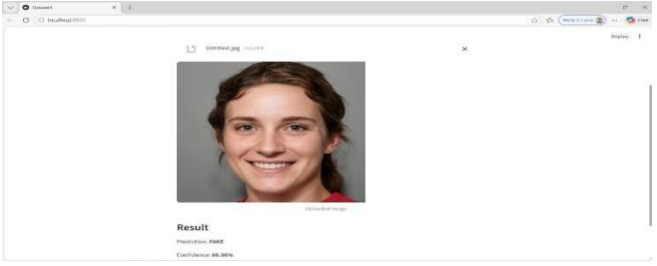


Figure 2 Home Page of the Deepfake Detection System

In Figure 2 shows the user interface of the Deepfake Detection System. The webpage allows users to upload an image or video either by dragging and dropping the file or by selecting it using the browse option. After the file is uploaded, the system processes the media and displays the prediction result indicating whether it is Real or Fake.

7.2. Deep Fake Detection – Result Interface



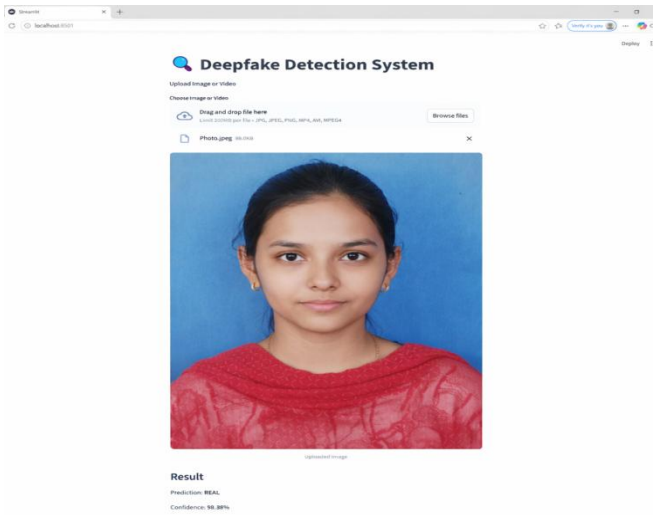


Figure 3 Deep Fake Detection – Result Interface

Figure 3 shows the result interface of the Deep Fake Detection system. In this interface, the user uploads a facial image, and the system analyzes it using a deep learning model. After processing the image, the system displays the prediction result along with the confidence score. This helps users easily understand whether the uploaded image is real or fake.

8. Experimental Results And Observations

The proposed Deepfake Detection System was evaluated using a dataset containing both real and fake images. The dataset was divided into training and testing sets to measure the model's performance accurately. A Convolutional Neural Network (CNN) model was trained to extract spatial features from images such as facial textures, artifacts, and inconsistencies. The performance of the system was evaluated using standard metrics such as Accuracy, Precision, Recall, and F^1-Score. The results show that the CNN model effectively distinguishes between real and fake images. The system successfully identified visual artifacts present in manipulated images and achieved high detection accuracy. It also demonstrated efficient processing time, making it suitable for practical and near real-time applications. These results confirm that the proposed system provides reliable and accurate deepfake image detection. Key Observations:

- The CNN model improves deepfake image detection accuracy by learning important visual features.
- CNN effectively captures spatial artifacts in

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 1722-1777
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0231

images such as texture distortions and facial inconsistencies.

- The system provides reliable Real/Fake classification with confidence scores.
- The model shows efficient processing time, making it suitable for near real-time image detection.

Conclusion and Future Work

- Conclusion: A Deepfake Detection System based on a CNN architecture was proposed and implemented. The system analyzes spatial features in images to accurately classify them as Real or Fake. Experimental results show that the CNN model effectively detects facial artifacts and visual inconsistencies present in manipulated images. The system also provides confidence scores and processes images efficiently, making it suitable for practical and near real-time applications. Overall, the proposed method provides a reliable solution for detecting deepfake images.
- Future Enhancements: In the future, the Deepfake Detection System can be improved in several ways. A larger and more diverse image dataset can be used to increase accuracy and make the model more robust. Advanced deep learning models can be integrated to improve detection performance for high-quality deepfake images. The system can also be optimized to reduce processing time and improve efficiency. Additionally, the system can be deployed as a mobile application or browser extension so that users can easily check fake images on social media platforms. Continuous model updates and cloud integration can further improve scalability and reliability, making the system more suitable for real-world applications.

References

[1]. H. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A Compact Facial Video Forgery Detection Network," in Proc. IEEE Int. Workshop on Information Forensics and Security (WIFS), 2018, pp. 1–7.

[2]. Y. Li and S. Lyu, "Exposing DeepFake Videos by Detecting Face Warping Artifacts," in Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 46–52.

[3]. A. Rössler et al., "FaceForensics++: Learning to Detect Manipulated Facial Images," in Proc. IEEE Int. Conf. Computer Vision (ICCV), 2019, pp. 1–11.

[4]. H. Nguyen, J. Yamagishi, and I. Echizen, "Capsule-Forensics: Using Capsule Networks to Detect Forged Images," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 2307–2311.

[5]. L. Li, J. Bao, T. Zhang, and B. Yang, "Face X-Ray for More General Face Forgery Detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2020, pp. 5001–5010.

[6]. R. Durall, M. Keuper, F. Pfreundt, and J. Keuper, "Unmasking DeepFakes with Simple Features," in Proc. IEEE Int. Conf. Image Processing (ICIP), 2020, pp. 6–10.

[7]. Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2020, pp. 3207–3216.

[8]. B. Dolhansky et al., "The DeepFake Detection Challenge (DFDC) Dataset," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 3068–3077.

[9]. Y. Qian, G. Yin, L. Sheng, Z. Chen, and J. Shao, "Thinking in Frequency: Face Forgery Detection by Mining Frequency-Aware Clues," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 43, no. 12, pp. 1–15, 2021.

[10]. Z. Yan et al., "Asymmetric Deepfake Detection with Adversarial Training," in Proc. IEEE Int. Conf. Multimedia and Expo (ICME), 2021, pp. 1–6.

[11]. Y. Zhou and S. Lim, "Two-Stream Neural Networks for Tampered Face Detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW),

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 1722-1777
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0231

2019, pp. 1831–1839.

[12]. J. Frank et al., "Leveraging Frequency Analysis for Deep Fake Image Recognition," in Proc. Int. Conf. Machine Learning (ICML), 2020, pp. 3247–3258.

[13]. S. Agarwal et al., "Protecting World Leaders Against Deep Fakes," in Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 38–45.

[14]. P. Korshunov and S. Marcel, "DeepFakes: A New Threat to Face Recognition? Assessment and Detection," IEEE Access, vol. 7, pp. 164208–164220, 2019.

[15]. X. Yang, Y. Li, and S. Lyu, "Exposing Deep Fakes Using Inconsistent Head Poses," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 8261–8265.

[16]. M. Huh et al., "Fighting Fake News: Image Splice Detection via Learned Self-Consistency," in Proc. European Conf. Computer Vision (ECCV), 2018, pp. 101–117.

[17]. T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4401–4410.

[18]. N. Yu et al., "Attributing Fake Images to GANs: Learning and Analyzing GAN Fingerprints," in Proc. IEEE Int. Conf. Computer Vision (ICCV), 2019, pp. 7556–7566.

[19]. X. Wang et al., "CNN-Generated Images Are Surprisingly Easy to Spot… for Now," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2020, pp. 8695–8704.

[20]. K. Zhang, J. Xu, and H. Li, "Detecting Deepfake Images with Spatial Attention Mechanisms," IEEE Trans. Circuits and Systems for Video Technology, vol. 32, no. 5, pp. 1–12, 2022.

IEEE conference templates contain