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, May 2026
Impact Factor: 8.2

Deepfake Image Detection System Using CNN and LSTM

Abhishek Subhash Dhanapune and Dr. S. K. Sonkar
Department of Computer Engineering
Masters of Engineering (M.E)
Amrutvahini Collage of Engineering Sangamner, Ghulewadi, Maharashtra
abhisheksubhashdhanapune@gmail.com and sonkar83@gmail.com

Abstract: *With the rapid advancement of deep learning, deepfake images have become highly realistic and challenging to detect, posing serious risks in security, social media, and digital forensics. This study proposes a hybrid model combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) to enhance deepfake detection accuracy. The CNN extracts spatial features from images, while the LSTM captures sequential patterns associated with manipulated content. Trained on a diverse dataset of real and synthetic images, the model achieves high performance, demonstrating its effectiveness for reliable detection in security and forensic applications. This work contributes a robust framework for combating image manipulation and supports future research in deepfake detection.*

Keywords: Deepfake Detection, CNN, LSTM, Deep Learning, Image Forensics, Artificial Intelligence, Security Systems.

I. INTRODUCTION

The rapid advancement of artificial intelligence and deep learning technologies has significantly transformed digital media creation, enabling the generation of highly realistic synthetic content known as deepfakes [1]. These manipulated images and videos are typically created using advanced techniques such as Generative Adversarial Networks (GANs) [2], making them increasingly difficult to distinguish from authentic media. While deepfake technology has beneficial applications in entertainment and digital content creation, it also poses serious risks related to misinformation, identity theft, and cybercrime [3]. The growing misuse of deepfakes has raised critical concerns regarding digital trust and security, thereby necessitating the development of reliable detection systems [4].

To tackle these challenges, researchers have explored various deep learning-based approaches that focus on identifying inconsistencies within manipulated media. Convolutional Neural Networks (CNNs) are widely used for extracting spatial features such as textures, edges, and facial patterns from images [5]. However, deepfake content often contains temporal irregularities across frames, which cannot be effectively captured using CNNs alone. As a result, hybrid models combining CNNs with Long Short-Term Memory (LSTM) networks [6] have gained popularity, as they enable the simultaneous analysis of spatial and sequential information, significantly improving detection performance [7].

Deep learning plays a crucial role in enabling such advanced detection systems by automatically learning complex patterns from large datasets [1]. Unlike traditional machine learning techniques, deep learning models can efficiently process vast amounts of data and identify subtle anomalies that are often invisible to the human eye [2]. Popular frameworks such as TensorFlow, PyTorch, and Keras have accelerated the development of high-performance models [3]. Additionally, publicly available datasets such as FaceForensics++ and the Deepfake Detection Challenge (DFDC) provide diverse training data, improving model generalization and robustness [4] .

With the continuous evolution of deepfake generation techniques, detection systems must also advance to remain effective [5]. The proposed system leverages a hybrid CNN-LSTM architecture, where CNN extracts spatial features while LSTM captures temporal dependencies [6]. By training the model on diverse datasets and evaluating its performance using metrics such as accuracy, precision, recall, and F1-score [7], the system aims to provide a robust and

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, May 2026
Impact Factor: 8.2

reliable solution for real-world applications [8]. This research contributes to enhancing digital media authenticity and supports applications in cybersecurity, journalism, and ethical AI development [9][10].

Motivation

The "Deepfake Detection System Using LSTM and CNN" is motivated by the increasing threat of highly realistic manipulated media that undermines public trust, personal security, and information integrity. Deepfakes contribute to misinformation, distort public discourse, and enable identity misuse, leading to harassment, fraud, and cyber threats. This system aims to provide a reliable detection mechanism to verify content authenticity, support journalists and cybersecurity efforts, and prevent malicious exploitation. By leveraging CNN and LSTM models to analyze spatial and temporal features, the project demonstrates the real-world impact of AI in solving complex problems. Additionally, it promotes awareness, interdisciplinary research, and ethical responsibility, contributing to regulatory efforts and the development of trustworthy digital ecosystems.

Objectives

1. To Develop a system that can accurately distinguish between real and deepfake images.
2. To Enhance detection accuracy by optimizing the hybrid architecture.
3. To Build and Train a Deepfake Detection Model.
4. To evaluate the model using standard metrics such as accuracy, precision, recall, F1-score.

II. LITERATURE SURVEY

1. Dulaimi et al. (2021)

Dulaimi et al. (2021) highlighted the serious societal and geopolitical risks associated with deepfakes, including misinformation, reputational damage, and manipulation of public opinion. The study introduced the concept of the "liar's dividend," where widespread deepfake usage leads to distrust in all digital media. The authors emphasized the need for strong detection systems along with legal and ethical frameworks to combat these threats effectively [4] .

2. Alshingiti et al. (2022)

Alshingiti et al. (2022) explored deepfake detection using hybrid deep learning models, particularly CNN and LSTM networks. CNNs were used for extracting spatial features, while LSTMs captured temporal dependencies in video sequences. The study demonstrated that combining these models improves detection accuracy, though challenges such as generalization and real-time performance remain [5]

3. Lamichhane et al. (2020)

Lamichhane et al. (2020) proposed advanced face forgery detection techniques that focus on identifying unknown manipulation patterns rather than relying on predefined features. By using deep learning models like CNNs along with attention mechanisms and feature fusion, the system improved adaptability and robustness against evolving deepfake techniques [7] .

4. Raza et al. (2021)

Raza et al. (2021) discussed modern image forgery detection methods using deep learning, particularly CNN-based approaches. The study emphasized the importance of datasets like the DeepFake Detection Challenge (DFDC) for improving detection accuracy. It also highlighted the use of physical and statistical constraints to enhance model generalization across different manipulation techniques [8] .

5. Saikia et al. (2023)

Saikia et al. (2023) focused on hybrid CNN-LSTM models for deepfake video detection, integrating spatial and temporal analysis. The study introduced optical flow features to capture motion inconsistencies between frames, which significantly improved detection performance. This approach proved effective in identifying subtle manipulations in video-based deepfakes [11] .

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, May 2026
Impact Factor: 8.2

Existing system

1. Traditional Image Forensics Methods

Earlier deepfake detection systems relied on traditional image processing techniques such as analyzing pixel inconsistencies, lighting variations, and compression artifacts. These methods focused on identifying visible distortions in images but were limited in detecting highly realistic deepfakes, as modern manipulation techniques can eliminate such obvious artifacts.

2. Machine Learning-Based Approaches

With the advancement of machine learning, classifiers such as Support Vector Machines (SVM) and Random Forests were used for detecting manipulated content. These approaches required manual feature extraction, making them less efficient and less adaptable to new types of deepfake techniques, resulting in lower accuracy for complex datasets.

3. Deep Learning-Based Detection Systems

Recent systems use deep learning models, particularly Convolutional Neural Networks (CNNs), to automatically extract spatial features from images. These models improve detection accuracy by identifying subtle patterns and inconsistencies. However, CNN-based systems alone are limited as they cannot effectively capture temporal dependencies in video data.

4. Hybrid and Dataset-Driven Systems

Some advanced existing systems combine multiple techniques and utilize large datasets such as FaceForensics++ and DFDC for training. Hybrid approaches like CNN with Recurrent Neural Networks (RNNs) have been introduced to improve performance. Despite improvements, these systems still face challenges such as high computational cost, lack of real-time detection, and difficulty in generalizing across evolving deepfake techniques.

III. PROPOSED SYSTEM

1. Data Collection and Input Handling

The proposed system begins with collecting a dataset consisting of both real and deepfake images. These images are sourced from publicly available datasets or generated using deepfake techniques to ensure diversity. The system accepts user inputs in the form of images, which are then passed into the processing pipeline for analysis. This step ensures that the model is trained on varied data, improving its ability to generalize across different types of manipulations.

2. Data Preprocessing

In this stage, the input images undergo preprocessing to enhance model performance. The images are resized to a uniform dimension, and pixel values are normalized to maintain consistency across the dataset. Data augmentation techniques such as rotation, flipping, and scaling are applied to increase dataset variability and reduce overfitting. The processed dataset is then stored and prepared for training and testing.

3. Data Splitting and Loading

The preprocessed dataset is divided into training and testing sets to evaluate model performance effectively. A data loader is used to efficiently load images and their corresponding labels into the model during training. This ensures optimized memory usage and faster processing, enabling smooth handling of large datasets.

4. Hybrid CNN-LSTM Model Training

The core of the proposed system is a hybrid deep learning model combining CNN and LSTM. The CNN component extracts spatial features such as textures, edges, and facial inconsistencies from images. These extracted features are then passed to the LSTM network, which analyzes sequential patterns and dependencies to detect subtle manipulations. The model is trained using labeled data and optimized using performance metrics such as accuracy and loss.

5. Model Evaluation and Prediction

After training, the model is evaluated using techniques such as confusion matrix, accuracy, precision, recall, and F1-score. The trained model is then saved and can be loaded for real-time prediction. When a new image is provided, 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, May 2026
Impact Factor: 8.2

system processes it through the trained model and classifies it as REAL or FAKE. This ensures an efficient and reliable deepfake detection system suitable for practical applications.

IV. SYSTEM DESIGN

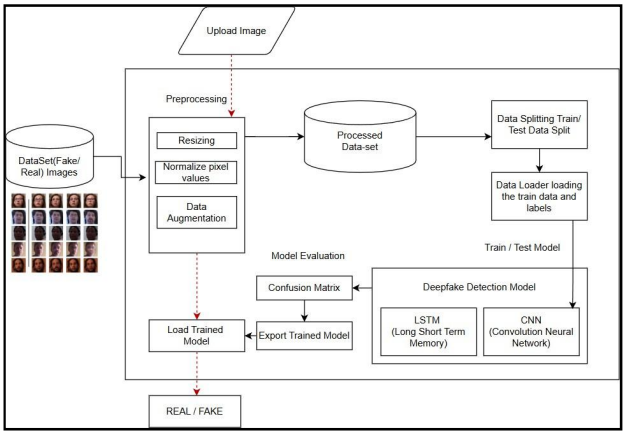

Fig. 1 Block Diagram

System Design

The proposed system is designed to detect deepfake images using a structured pipeline that integrates preprocessing, model training, and prediction. The workflow begins with the input dataset and progresses through multiple stages to produce an accurate classification of images as real or fake.

1. Input Dataset (Real/Fake Images)

The system starts with a dataset containing both real and deepfake images. These images act as the primary input for training and testing the model. A diverse dataset ensures better learning and improves the system's ability to detect various manipulation techniques.

2. Preprocessing Module

The preprocessing block prepares the input data for efficient model training. It includes:

- **Resizing:** All images are resized to a fixed dimension for uniformity.
- **Normalization:** Pixel values are scaled (e.g., 0–1 range) to stabilize training.
- **Data Augmentation:** Techniques like rotation, flipping, and zooming are applied to increase dataset diversity and reduce overfitting.

The output of this stage is a clean and standardized dataset.

3. Processed Dataset Storage

After preprocessing, the images are stored as a processed dataset. This ensures that the model receives optimized and consistent data during training and testing phases.

4. Data Splitting and Loading

The processed dataset is divided into **training and testing sets**.

- The **training set** is used to train the model.
- The **testing set** is used to evaluate performance.

A data loader is used to efficiently feed images and labels into the model in batches, improving computational efficiency.

5. Deepfake Detection Model (CNN + LSTM)

This is the core component of the system:

- **CNN (Convolutional Neural Network):** Extracts spatial features such as edges, textures, and facial inconsistencies from 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, May 2026 Impact Factor: 8.2

- **LSTM (Long Short-Term Memory):** Processes the extracted features to identify patterns and dependencies, helping detect subtle manipulations.

The hybrid CNN-LSTM model enhances detection accuracy by combining spatial and sequential learning.

6. Model Training and Testing

The model is trained using the training dataset and validated using the testing dataset. During training, the model learns to differentiate between real and fake images by minimizing loss and improving accuracy.

7. Model Evaluation

After training, the model is evaluated using performance metrics such as:

- **Confusion Matrix**
- **Accuracy**
- **Precision, Recall, F1-Score**

These metrics help measure how effectively the model identifies deepfake content.

8. Model Saving and Loading

The trained model is saved for future use. It can be reloaded without retraining, making the system efficient for deployment and real-time applications.

9. Final Prediction Output

When a new image is provided, it goes through preprocessing and is passed to the trained model. The system then classifies the image as:

- **REAL** or
- **FAKE**

This final output helps users verify the authenticity of digital content.

VI. OUTCOMES FROM THE SYSTEM

This chapter presents a detailed evaluation of the proposed hybrid CNN–LSTM model for deepfake image detection, highlighting both quantitative performance and practical effectiveness. The results are analyzed using standard machine learning metrics and validated through confusion matrix outcomes and real-time testing.

1. Evaluation Metrics Analysis

To assess the performance of the proposed system, standard evaluation metrics such as Accuracy, Precision, Recall, and F1-Score were used. Accuracy measures the overall correctness of the model, while Precision evaluates how many predicted fake images are actually fake. Recall indicates the model's ability to correctly identify all fake images, and the F1-Score provides a balance between Precision and Recall. Additionally, the Confusion Matrix offers a comprehensive view of classification performance by showing True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These metrics collectively ensure a reliable and unbiased evaluation of the system.

2. Confusion Matrix Interpretation

The confusion matrix results demonstrate the effectiveness of the model in distinguishing between real and fake images. Out of the total predictions, 480 real images were correctly classified as real (TN), and 465 fake images were correctly identified as fake (TP). Only 20 real images were misclassified as fake (FP), and 35 fake images were misclassified as real (FN). This indicates that the model maintains a low error rate and performs consistently well across both classes. The relatively lower number of false positives ensures that genuine images are not wrongly flagged, which is crucial in real-world applications.

3. Performance Evaluation and Model Efficiency

The model achieved an Accuracy of 94.5%, demonstrating high overall correctness. The Precision of 95.8% indicates that the system is highly reliable when identifying fake images, minimizing incorrect detections. The Recall of 92.98% shows strong capability in detecting most deepfake instances, while the F1-Score of 94.3% reflects a balanced and

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, May 2026
Impact Factor: 8.2

robust performance. These results confirm that the hybrid CNN-LSTM architecture effectively captures both spatial and sequential features, leading to improved detection accuracy compared to traditional methods.

4. Practical Implementation and Real-Time Results

The system was also deployed as a web-based application, allowing users to upload images and receive instant classification results as REAL or FAKE. The live implementation demonstrates the model's practical usability and efficiency in real-time scenarios. The application provides a user-friendly interface and ensures quick processing, making it suitable for use in cybersecurity, digital forensics, and media verification.

Project Live Link: https://deepfake-detection-tool.netlify.app/

Overall, the results validate that the proposed system is accurate, reliable, and ready for real-world deployment in detecting deepfake content.

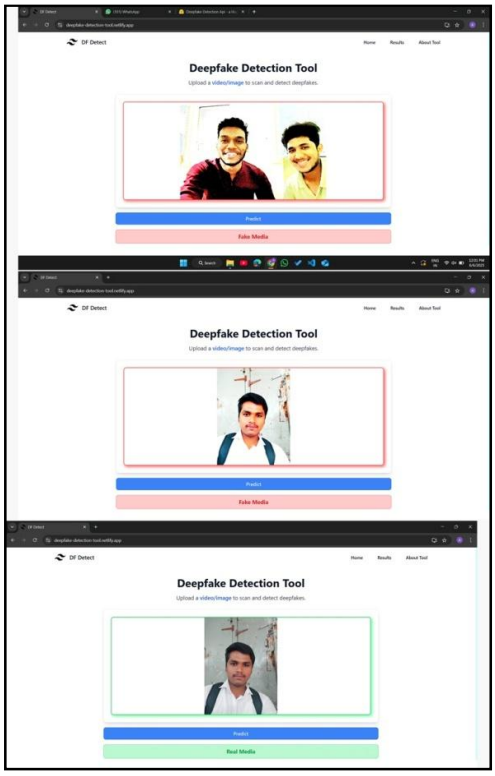

Fig 2: User Interface

VII. CONCLUSION

The proposed Deepfake Image Detection System using a hybrid CNN–LSTM model successfully addresses the growing challenge of identifying manipulated media by combining spatial and temporal feature analysis. The system demonstrated high performance with strong accuracy, precision, recall, and F1-score, indicating its reliability and robustness in distinguishing real and fake images. The integration of preprocessing techniques, efficient data handling, and advanced deep learning architecture contributed to improved detection capability and reduced misclassification. Additionally, the successful deployment of the model in a real-time application highlights its practical usability in areas such as cybersecurity, digital forensics, and media verification. Overall, the system provides an effective and scalable solution for combating deepfake threats and establishes a strong foundation for future enhancements in detection techniques.

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, May 2026
Impact Factor: 8.2

VIII. FUTURE SCOPE

The proposed deepfake detection system can be further enhanced by expanding its capabilities to handle not only images but also videos and audio deepfakes, enabling a more comprehensive multimedia detection framework. Future improvements may include the integration of advanced architectures such as Transformers and attention-based models to improve accuracy and adaptability against evolving deepfake techniques. Incorporating transfer learning and larger, more diverse datasets can further enhance model generalization and robustness. Real-time optimization and deployment on mobile or edge devices can make the system more accessible and efficient for practical use. Additionally, combining facial analysis with behavioral biometrics and blockchain-based verification can strengthen authenticity validation. Continuous updates, integration with cybersecurity systems, and development of explainable AI techniques will also help improve trust, transparency, and effectiveness in real-world applications.

REFERENCES

1. Heidari, A., Navimipour, N., and Unal, M. (2024). *Deepfake detection using deep learning: A review*. Wiley. Discusses CNN, RNN, and hybrid approaches for deepfake detection.
2. Sunil, R., and Kumar, P. (2025). *Autonomous methods for deepfake detection*. Journal of King Saud University. Focuses on AI-based automated detection techniques.
3. Alrashoud, M., and Alqurashi, S. (2025). *Deepfake video detection methods*. Ain Shams Engineering Journal. Reviews spatial and temporal detection methods.
4. Sagar, N. K., and Sharma, V. (2025). *CNN-LSTM approach for deepfake detection*. Procedia Computer Science. Proposes hybrid model for improved accuracy.
5. Patel, V. M., and Degadwala, S. (2025). *Deepfake detection using CNN and LSTM*. IJSRST. Combines spatial and temporal feature extraction.
6. Yadav, S., et al. (2025). *Hybrid deep learning for deepfake defense*. Elsevier. Highlights robustness of hybrid models.
7. Lei, Y., et al. (2025). *GAN-LSTM-based fake face detection*. Springer. Uses GAN features with LSTM for detection.
8. Petmezas, G., et al. (2025). *CNN-LSTM-Transformer for deepfake detection*. Multimedia Tools and Applications. Introduces hybrid architecture.
9. Ramanaharan, R., et al. (2025). *Deepfake detection generalization study*. ScienceDirect. Focuses on model adaptability.
10. Dolhansky, B., et al. (2020). *DeepFake Detection Challenge dataset*. Facebook AI. Provides large-scale dataset for training.
11. Saikia, P., et al. (2022). *CNN-LSTM with optical flow for deepfake detection*. arXiv. Improves detection using motion features.
12. Wodajo, D., and Atnafu, S. (2021). *Deepfake detection using CNN and Vision Transformer*. arXiv. Combines CNN with transformer models.
13. Khan, S. A., et al. (2021). *Adversarially robust deepfake detection*. arXiv. Focuses on robustness against attacks.
14. Gupta, R., et al. (2024). *Machine learning approaches for deepfake detection*. IEEE. Reviews CNN, LSTM, and autoencoders.
15. Roessler, A., et al. (2019). *FaceForensics++ dataset*. IEEE. Benchmark dataset for facial manipulation detection.
16. Tolosana, R., et al. (2020). *Deepfake detection survey*. IEEE Access. Comprehensive overview of detection techniques.
17. Chesney, R., and Citron, D. (2019). *Deepfakes and digital deception*. California Law Review. Discusses legal and ethical issues.
18. Raza, A., et al. (2021). *Image forgery detection using deep learning*. Elsevier. Focuses on CNN-based detection methods.

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, May 2026
Impact Factor: 8.2

19. Lamichhane, B., et al. (2020). *Face forgery detection using deep learning*. Springer. Uses CNN and attention mechanisms.
20. Collins, E., et al. (2023). *Advancements in GAN-based image generation*. IEEE. Explains realistic image synthesis and challenges.