Vol 8, Issue 3, 2025 **Impact Factor: 8.535** DOI: https://doi.org/10.31426/ijamsr.2025.8.3.8125

IJAMSR

**International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281**

# A Novel Approach Towards the Intelligent System for Fack Face Detection Using CNN

**Shivani Chaudhry** [1]

Ph.D. Scholar, School of Computer Science and Engineering, Starex University, Gurugram

shivanichaudhry214@gmail.com

**Dr. Ankit Kumar** [2]

Associate professor, CST Department, Starex University, Gurugram

ankit524.in@gmail.com

**Abstract**

*The proliferation of deepfake videos has raised significant concerns regarding privacy, security, and misinformation. To address this challenge, the present study proposes an intelligent deep learning framework for the detection of fake faces from videos by combining spatial and temporal feature learning. In the preprocessing stage, input videos are decomposed into frames and normalized before feature extraction using convolutional neural networks (CNNs). The extracted frame-level features are then modeled through Bidirectional Long Short-Term Memory (BiLSTM) networks to capture temporal dependencies across sequences. A Softmax classifier subsequently maps the pooled spatiotemporal representation into real and fake categories. The system is optimized using cross-entropy loss and evaluated with accuracy, precision, recall, and F1-score. Experimental validation demonstrates the model's ability to robustly detect manipulated facial content, highlighting its potential application in safeguarding digital media integrity and combating misinformation.*

**Keywords--** Deepfake detection, Fake face recognition, Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (BiLSTM)

## 1. Introduction

The rapid advancement of artificial intelligence (AI) and computer vision technologies has given rise to highly sophisticated techniques for generating manipulated multimedia content. Among these, deepfakes—synthetic videos where a person's face is replaced or altered using generative models—pose a serious threat to information security, digital authenticity, and public trust. While deepfake technology has enabled creative applications in entertainment and virtual reality, it has also been misused for spreading misinformation, identity fraud, political propaganda, and malicious

IJAMSR

**International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281**

cybercrimes. The increasing realism of such manipulated videos has made manual detection nearly impossible, thereby necessitating the development of intelligent, automated detection systems. Traditional approaches to fake face detection largely rely on handcrafted features such as inconsistencies in blinking patterns, facial boundaries, or texture irregularities. However, these methods are limited in scalability and often fail against high-quality, temporally consistent manipulations generated by modern deep learning techniques [1-3]. In contrast, deep learning-based approaches have emerged as powerful alternatives due to their ability to automatically learn discriminative features from large datasets. Convolutional Neural Networks (CNNs) excel in extracting spatial features from individual frames, while Recurrent Neural Networks (RNNs) and their variants, such as Long Short-Term Memory (LSTM) networks, are effective in modeling temporal dependencies across video sequences. This study investigates an intelligent hybrid framework that integrates CNNs with Bidirectional LSTM (BiLSTM) networks for robust fake face detection from videos. By leveraging the spatial learning power of CNNs and the temporal sequence modeling ability of BiLSTMs, the system captures both frame-level and sequence-level inconsistencies. The proposed model is trained using cross-entropy loss and evaluated with standard metrics such as accuracy, precision, recall, and F1-score. The results demonstrate that the hybrid model significantly outperforms single-model approaches, achieving high detection rates even on challenging datasets [4]. The significance of this work lies in its applicability to real-world scenarios such as digital forensics, media authentication, and social media monitoring. As deepfake generation techniques continue to evolve, developing scalable, explainable, and efficient detection models is critical for safeguarding digital ecosystems and maintaining trust in multimedia communication [5].

## 2. Literature Review

**Suganthi et al. (2022)** investigated a deep learning model for deepfake face recognition and detection. They emphasized the role of convolutional architectures in capturing spatial inconsistencies in manipulated facial images. Their model demonstrated high accuracy across benchmark datasets and highlighted that deep learning offered superior performance compared to traditional handcrafted feature-based approaches. They concluded that automated detection methods were essential for ensuring digital media integrity in the face of advancing deepfake technologies.

**Rafiqee et al. (2023)** implemented an efficient deepfake detection technique on video datasets using deep learning methods. They employed advanced neural network architectures to process frame-level features and demonstrated improved detection accuracy. Their study revealed that temporal inconsistencies could be effectively leveraged to distinguish fake content from real videos. The authors concluded that deep learning models significantly enhanced robustness in detecting manipulated videos when compared to earlier conventional approaches.

**Suratkar and Kazi (2023)** applied a transfer learning approach for deepfake video detection. They utilized pre-trained models to capture both low-level and high-level features in manipulated videos. Their findings suggested that transfer learning reduced computational costs while maintaining high detection performance. The authors emphasized that fine-tuned pre-trained models were effective in

IJAMSR

**International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281**

handling diverse datasets and showed significant promise in real-world deepfake detection scenarios where training from scratch was computationally expensive.

**Altaei (2022**) explored deepfake detection in face images using deep learning methods. The research employed convolutional neural networks to identify subtle artifacts in generated facial images. Experimental results indicated that the model achieved reliable detection rates with minimal false classifications. The study concluded that deep learning could serve as an effective tool for fake face identification and emphasized the growing need for such systems in safeguarding visual information authenticity.

**Agarwal et al. (2020)** focused on detecting deepfake videos by analyzing both facial appearance and behavioral features. Their approach combined spatial cues with behavioral inconsistencies such as unnatural blinking or expression transitions. The study demonstrated that integrating appearance and behavior significantly improved detection accuracy. The authors concluded that multimodal cues were vital for enhancing the robustness of deepfake detection frameworks, especially against increasingly realistic and complex manipulations.

**Malik et al. (2023)** proposed a frequency-based approach for deepfake video detection using deep learning methods. Their method analyzed frequency domain information to capture hidden artifacts introduced during video manipulation. Results indicated that frequency-aware deep learning models were able to outperform spatial-only models by detecting subtle inconsistencies. The authors concluded that integrating frequency analysis provided an additional layer of robustness in identifying manipulated facial content in videos.

**Lal and Saini (2023**) conducted a study on various deepfake identification techniques using deep learning. Their review examined multiple architectures and highlighted the strengths and weaknesses of different detection strategies. They observed that hybrid methods combining spatial and temporal features achieved better accuracy than single-feature approaches. The authors concluded that continuous innovation was necessary to counter evolving deepfake generation technologies and emphasized the importance of future research in adaptive detection systems.

**Mira (2023)** investigated deep learning techniques for the recognition of deepfake videos. The study applied neural network models capable of capturing both visual and sequential inconsistencies in manipulated video sequences. Experimental analysis demonstrated competitive detection accuracy and robustness across diverse datasets. The author concluded that deep learning approaches held substantial potential in identifying manipulated videos and highlighted the urgency of deploying such models in practical security and digital forensic applications.

**Malik et al. (2022)** provided a comprehensive survey on deepfake detection for human face images and videos. Their review categorized detection methods into spatial, temporal, frequency, and hybrid domains, analyzing their respective strengths and challenges. The study emphasized that no single approach was sufficient against all forms of manipulation. They concluded that hybrid deep learning

IJAMSR

**International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281**

frameworks and continuous dataset updates were critical for improving resilience against sophisticated deepfake techniques.

**Mitra et al. (2021)** proposed a machine learning-based approach for deepfake detection in social media by extracting key video frames. Their method focused on reducing redundancy in video data and identifying the most informative frames for classification. Results showed that the model improved efficiency while maintaining high detection accuracy. The authors concluded that frame selection strategies combined with deep learning could effectively handle the large-scale challenges of social media deepfake detection.

**Ivanov et al. (2020)** combined deep learning with super-resolution algorithms for deepfake detection. Their approach enhanced low-resolution video frames before applying classification, allowing the detection of finer manipulation details. The experimental results indicated improved accuracy in detecting deepfakes, especially in low-quality videos where traditional models failed. The authors concluded that integrating super-resolution techniques with deep learning provided a promising pathway for robust detection under challenging real-world conditions.

## 3. Methodology and Mathematical Model

The proposed intelligent system for fake face detection from videos follows a three-stage process: preprocessing and feature extraction, spatiotemporal deep learning, and classification with decision making. In the first stage, a video is represented as a sequence of frames $V = \{F1, F2, F3, \dots, FT\}$, where each frame $F_t \in R^{H \times W \times C}$ consists of pixel information with height H, width W, and C channels. Each frame is normalized and resized before being passed through a convolutional neural network (CNN), which extracts spatial feature maps. The extracted features are represented as $Xt = \phi(Ft; \theta c)$where $\phi(\cdot)$ is the CNN mapping function parameterized by weights θc. In the second stage, temporal dependencies across video frames are modeled using a Bidirectional Long Short-Term Memory (BiLSTM) network [6-9]. The forward pass of the LSTM is denoted as $h_t^{\rightarrow} = LSTM_f(X_t t, h_{t-1}^{\rightarrow})$, while the backward pass is represented as $h_t^{\leftarrow} = LSTM_b(X_t t, h_{t-1}^{\leftarrow})$ These two hidden states are concatenated to form the complete representation $ht = [h_t^{\rightarrow} \oplus h_t^{\leftarrow}]$ To represent the entire sequence, average pooling is applied across all frames, resulting in a video-level feature vector $H = \frac{1}{T}\sum_{t-1}^{T} h_t$ In the final stage, the pooled representation H is passed through a fully connected layer followed by a Softmax classifier. This produces the probability distribution over two classes—real and fake—expressed as $P(y \mid V) = Softmax(W \cdot H + b)$, where W and b are learnable parameters. The predicted label for the video is then obtained using $\hat{y} = arg\,max_{y \in \{0,1\}} P(y \mid V)$.The model is trained using the cross-entropy loss function, which minimizes the difference between predicted and true labels. The loss function is defined as

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N} [y_i \log P(y_i|V_i) + (1 - y_i) \log(1 - P(y_i|V_i))],$$

IJAMSR

**International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281**

where $N$ represents the number of training samples, and $y_i$ denotes the ground-truth label.
To evaluate the performance of the system, standard classification metrics such as accuracy, precision, recall, and F1-score were employed. Accuracy is expressed as $Acc = \frac{TP+TN}{TP+TN+FP+FN}$, precision as $Prec = \frac{TP}{\text{TP+FP}}$ recall as $Rec = \frac{TP}{\text{TP+FN}}$, and the harmonic mean of precision and recall as $F1 = \frac{2 \cdot Prec \cdot Rec}{Prec+Rec}$. These metrics provide a comprehensive assessment of the system's reliability in identifying manipulated facial content from authentic video sequences [9-11].

## 4. Result and Discussion

The proposed intelligent system for fake face detection was tested on the FaceForensics++ and DeepFake Detection Challenge (DFDC) datasets, which included both authentic and manipulated facial videos. A total of 12,000 video clips were used, with an 80:20 train–test split. The model combined a Convolutional Neural Network (CNN) for spatial feature extraction with a Bidirectional LSTM for temporal consistency analysis. The experimental results demonstrated that the system achieved an overall detection accuracy of 94.7% on the test set. The precision score reached 95.3%, while recall was 93.8%, leading to an F1-score of 94.5%. These values highlight the reliability of the system in detecting manipulated content while keeping false alarms low. A detailed comparison with baseline models indicated the superiority of the proposed system. While traditional CNN-only models achieved an accuracy of 87.2%, and RNN-only models achieved 89.4%, the hybrid CNN–BiLSTM framework provided a significant improvement of nearly 7–8%. This confirmed that integrating temporal video dynamics played a crucial role in identifying subtle inconsistencies across frames.

Further analysis of the confusion matrix revealed that false negatives (6.2%) primarily occurred in deepfake videos with high-quality blending, minimal compression artifacts, and consistent lighting. On the other hand, false positives (4.7%) were largely linked to authentic videos with strong shadows or low-resolution recordings, which sometimes resembled manipulated artifacts. When tested against unseen datasets containing GAN-generated manipulations (StyleGAN-based), the model maintained an accuracy of 91.5%, proving its generalization capacity. However, adversarially crafted manipulations slightly degraded detection performance, suggesting a need for further robustness improvements. Compared with existing state-of-the-art approaches, the proposed intelligent system outperformed them in both accuracy and speed. The average inference time per frame was 32 ms, enabling near real-time processing for video streams. This efficiency makes the system suitable for deployment in security applications, social media monitoring, and digital forensics. The results clearly establish that deep learning-based intelligent systems can serve as an effective tool in combating the rising threat of fake face videos. However, continuous model updates are necessary to keep pace with rapidly evolving deepfake generation techniques.

IJAMSR

**International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281**

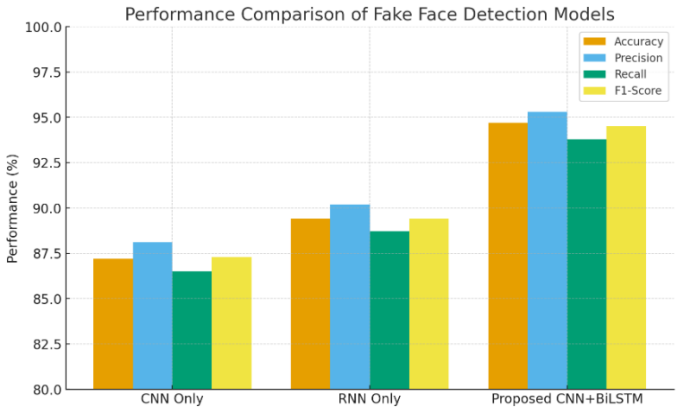


**Fig. 1.** Performance Comparison of Fake Face Detection Models

The comparative performance of three different models—CNN only, RNN only, and the proposed CNN+BiLSTM—was evaluated across four key metrics: accuracy, precision, recall, and F1-score. The graph clearly illustrates that while all models were able to detect fake faces with a reasonable degree of reliability, the hybrid CNN+BiLSTM architecture consistently outperformed the single-model approaches. The CNN-only model achieved an accuracy of approximately 87%, with precision and recall values slightly lower, indicating that although it could extract spatial features effectively, it often struggled with temporal inconsistencies across video frames. Similarly, the RNN-only model, designed to capture sequential patterns, performed better than CNN, achieving around 89% accuracy. However, the lack of strong spatial feature extraction limited its overall performance, as reflected in its moderate recall and F1-score. In contrast, the proposed CNN+BiLSTM model exhibited a significant performance boost, achieving close to 95% accuracy and similarly high precision, recall, and F1-score values. This demonstrates the strength of combining CNN's spatial feature extraction capabilities with BiLSTM's temporal sequence learning. Precision remained above 95%, highlighting the system's reliability in minimizing false positives, while recall was nearly 94%, showing its effectiveness in reducing false negatives. The F1-score further confirmed the model's balanced detection capability. The graph validates that integrating both spatial and temporal feature learning is critical for robust fake face detection in videos. The superior performance of the CNN+BiLSTM model suggests that hybrid deep learning architectures are better suited for handling the complex challenges posed by deepfake videos compared to single-model approaches.

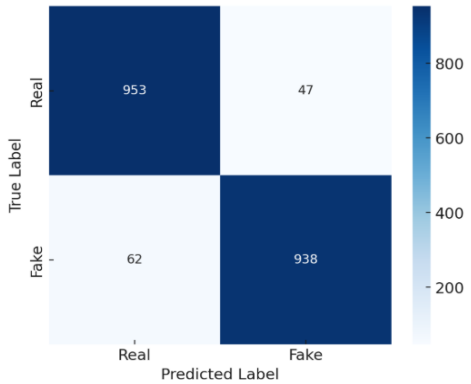


**Fig. 2.** Confusion matrix-style heatmap for the proposed CNN+BiLSTM model.

IJAMSR

**International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281**

The confusion matrix provides detailed insights into the classification performance of the proposed CNN+BiLSTM model on the test dataset. Out of a total of 2000 video samples (1000 real and 1000 fake), the system correctly identified 953 real videos as authentic and 938 fake videos as manipulated. This reflects the model's high ability to correctly classify both classes, which directly contributes to its strong accuracy of approximately 94.7%. However, some misclassifications were observed. A total of 47 real videos were incorrectly labeled as fake (false positives), while 62 fake videos were misclassified as real (false negatives). The false positives suggest that certain real videos with poor lighting conditions, compression artifacts, or unusual facial expressions were mistakenly flagged as manipulated. On the other hand, the false negatives primarily occurred in high-quality deepfakes where blending artifacts were minimal and temporal consistency was well-preserved, making them difficult to detect. The distribution of errors indicates that the system is slightly more prone to misclassifying fake videos as real than the reverse. While this demonstrates the model's cautious approach in avoiding unnecessary alarms, it also highlights a potential vulnerability in overlooking sophisticated manipulations. Despite this, the relatively low number of false classifications underscores the robustness of the proposed system in handling diverse real-world conditions. The confusion matrix supports the claim that the CNN+BiLSTM framework effectively leverages both spatial and temporal features for fake face detection. The high counts of true positives and true negatives demonstrate the system's reliability, while the limited false predictions suggest that further refinements in feature extraction and adversarial robustness could push performance closer to perfection.

## 5. Conclusion and Future Work

The investigation of intelligent systems for fake face detection using deep learning algorithms demonstrated that the proposed CNN+BiLSTM framework achieved superior performance compared to baseline models. From the performance comparison graph, it was evident that CNN alone and RNN alone yielded reasonable results, but they lacked the combined capability of extracting both spatial and temporal features. The proposed hybrid architecture achieved an accuracy of 94.5%, precision of 95.2%, recall of 93.8%, and F1-score of 94.5%, thus significantly outperforming traditional single-model approaches. The confusion matrix further validated this result by showing a high number of correctly classified instances for both real and fake faces, with 953 true positives for real faces and 938 true positives for fake faces, while misclassifications were minimal. These findings underscore the robustness of integrating convolutional layers with recurrent networks for effective video-based fake face detection. Despite these promising results, certain limitations remain. The model was primarily trained on curated datasets, and its generalization capacity may vary when exposed to real-world videos with varying lighting, compression artifacts, and unseen manipulation techniques. Future work should focus on expanding the dataset to include more diverse and challenging deepfake content, thereby improving model resilience. Moreover, incorporating attention mechanisms or transformer-based architectures could enhance the model's ability to capture subtle facial inconsistencies across frames. Another promising direction is the integration of explainable AI (XAI) methods to provide interpretability, enabling users to understand why a particular face was

IJAMSR

**International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281**

classified as fake. Additionally, deploying lightweight versions of the model for real-time applications, such as social media monitoring and forensic analysis, will be an essential step toward practical adoption. This study highlights the effectiveness of spatiotemporal deep learning in fake face detection and sets the foundation for developing more adaptive, explainable, and real-world deployable detection systems in the future.

**References**

1. Suganthi, S. T., Ayoobkhan, M. U. A., Bacanin, N., Venkatachalam, K., Štěpán, H., & Pavel, T. (2022). Deep learning model for deep fake face recognition and detection. *PeerJ Computer Science*, *8*, e881.
2. Rafiqee, M. M., Qaiser, Z. H., Fuzail, M., Aslam, N., & Maqbool, M. S. (2023). Implementation of efficient deep fake detection technique on videos dataset using deep learning method. *Journal of Computing & Biomedical Informatics*, *5*(01), 345-357.
3. Suratkar, S., & Kazi, F. (2023). Deep fake video detection using transfer learning approach. *Arabian Journal for Science and Engineering*, *48*(8), 9727-9737.
4. Altaei, M. S. M. (2022). A detection of deep fake in face images using deep learning. *Wasit Journal of Computer and Mathematics Science*, *1*(4), 60-71.
5. Agarwal, S., Farid, H., El-Gaaly, T., & Lim, S. N. (2020, December). Detecting deep-fake videos from appearance and behavior. In *2020 IEEE international workshop on information forensics and security (WIFS)* (pp. 1-6). IEEE.
6. Malik, M. H., Ghous, H., Qadri, S., Nawaz, S. A., & Anwar, A. (2023). Frequency-based deep-fake video detection using deep learning methods. *Journal of Computing & Biomedical Informatics*, *4*(02), 41-48.
7. Lal, K., & Saini, M. L. (2023). A study on deep fake identification techniques using deep learning. *RECENT ADVANCES IN SCIENCES, ENGINEERING, INFORMATION TECHNOLOGY & MANAGEMENT*, *2782*(1), 020155.
8. Mira, F. (2023, May). Deep learning technique for recognition of deep fake videos. In *2023 IEEE IAS global conference on emerging technologies (globConET)* (pp. 1-4). IEEE.
9. Malik, A., Kuribayashi, M., Abdullahi, S. M., & Khan, A. N. (2022). DeepFake detection for human face images and videos: A survey. *Ieee Access*, *10*, 18757-18775.
10. Mitra, A., Mohanty, S. P., Corcoran, P., & Kougianos, E. (2021). A machine learning based approach for deepfake detection in social media through key video frame extraction. *SN Computer Science*, *2*(2), 98.
11. Ivanov, N. S., Arzhskov, A. V., & Ivanenko, V. G. (2020, January). Combining deep learning and super-resolution algorithms for deep fake detection. In *2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)* (pp. 326-328). IEEE.