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 5, Issue 1, June 2025

Impact Factor: 7.67

Detection of Deep Fake Videos using CNN and GRU Algorithms

D Aswani, P Kalyani, S Ushasri, T Vyshnavi, S Akshay
Department of Computer Science and Engineering
ACE Engineering College, Ghatkesar, Hyderabad

Abstract: *In the era of digital communication, deepfake technology poses a critical threat to the integrity of information disseminated online. These AI-generated videos, capable of realistically depicting individuals saying or doing things they never did, have significant implications for public discourse, human rights, and the authenticity of digital media. The potential misuse of deepfakes for misinformation, manipulation, harassment, and coercion necessitates advanced solutions for their detection. Addressing this challenge, we are developing a deepfake video detector leveraging the capabilities of Gated Recurrent Units (GRUs). Our approach utilizes GRU's sequential processing abilities to analyze video frames for subtle inconsistencies typical of deepfaked content. By focusing on pixel-level discrepancies and temporal anomalies that are often imperceptible to the human eye, our model offers a promising solution to identifying manipulated media. This work not only contributes to the technological fight against digital misinformation but also underscores the importance of cross-sector collaboration in safeguarding the veracity of online media. Our findings illuminate the path for future research and development in the field, highlighting the critical role of advanced machine learning techniques in maintaining the credibility and security of digital communications..*

Keywords: Deepfake detection, Gated Recurrent Units (GRUs), Sequential processing, Video frame analysis, Pixel-level discrepancies, Temporal anomalies, Manipulated media, AI-generated videos, Machine learning techniques, Digital misinformation, Information integrity, Digital media authenticity.

I. INTRODUCTION

The advent of deepfake technology has ushered in a new era of digital manipulation, where the boundary between truth and fiction is increasingly blurred. Deepfakes, sophisticated video and audio forgeries made possible by advanced artificial intelligence (AI) and machine learning algorithms, have the capability to realistically depict individuals saying or doing things they have never done.[1]This technology, while impressive, poses significant threats to the integrity of information online, public trust, and the very fabric of democratic discourse. The potential for misuse in creating false narratives, manipulating elections, inciting violence, or violating personal privacy underscores a critical challenge for society[2].

In response to this ongoing threat, our project aims to develop a sophisticated deepfake detection system leveraging the advanced capabilities of Gated Recurrent Units (GRUs). This system is designed to identify and differentiate between genuine and manipulated videos by analyzing pixel-level discrepancies and temporal anomalies within video frames—subtle markers that typically evade human detection[3]. By focusing on these intricacies, we propose a solution that not only addresses the immediate challenge of detecting deepfakes but also contributes to a broader understanding of digital authenticity and security[4].

Our work is motivated by the imperative to protect individuals' rights, uphold the accuracy of information, and maintain public trust in digital media. The project takes a comprehensive approach, integrating InceptionNet for feature extraction and Gated Recurrent Units (GRUs) for sequential analysis, alongside the design and training of a CNN model. Rigorous testing against a diverse set of deepfake examples ensures robustness, and strategies for deployment and integration into digital platforms are prioritized[5].

DOI: 10.48175/IJARSCT-27318

Furthermore, we emphasize the importance of cross-sector collaboration, community engagement, and education to raise awareness and promote critical engagement with digital media.As we venture into this critical undertaking, our project stands at the intersection of technology, ethics, and societal well-being[6]. It represents a proactive step toward safeguarding the digital information landscape, ensuring AI's revolutionary potential is harnessed for societal betterment rather than undermining foundational truths. Through this initiative, we aim to combat digital misinformation and preserve the authenticity of digital content, protecting the integrity of public discours[7].

II. LITERATURE REVIEW

[1]. Afchar et al. (2018) - MesoNet: A Compact Facial Video Forgery Detection Network

In response to the proliferation of deepfake videos, Afchar et al. proposed MesoNet, a specialized deep learning architecture tailored for detecting facial manipulations in videos. Recognizing the need for efficient and accurate detection methods, their work focuses on developing a compact yet effective model capable of discerning subtle inconsistencies indicative of deepfake videos.

[2]. Rössler et al. (2019) - FaceForensics++: Learning to Detect Manipulated Facial Images

Addressing the urgent need for standardized benchmarks in deepfake detection, Rössler et al. introduced the FaceForensics++ dataset—a comprehensive collection of manipulated facial videos. Their research aims to provide a robust evaluation platform for assessing the efficacy of various detection algorithms against a diverse range of manipulations.

[3]. Guera and Delp (2018) - Deepfake Video Detection Using Recurrent Neural Networks

Recognizing the temporal nature of video content, Guera and Delp explore the integration of Recurrent Neural Networks (RNNs) with Convolutional Neural Networks (CNNs) for deepfake detection. Their research seeks to leverage temporal information to enhance the accuracy and robustness of detection systems against manipulated videos.

[4]. Zhou et al. (2017) - Two-Stream Neural Networks for Tampered Face Detection

Zhou et al. propose a two-stream neural network approach for detecting tampered faces in videos, leveraging both spatial and temporal information to enhance detection capabilities. Their research aims to provide a comprehensive framework for identifying manipulations in video content, including deepfake videos.

[5]. Li et al. (2020) - In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking

Li et al. propose a novel approach to deepfake detection based on the absence of natural eye blinking in manipulated videos. Their research focuses on exploiting physiological signals as unique identifiers, offering a new avenue for detecting deepfake videos amidst the growing threat of digital misinformation.

[6]. Tolosana et al. (2020) - DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection

Tolosana et al. conduct a comprehensive survey of face manipulation techniques and detection methods, offering insights into the evolving landscape of deepfake technology and detection strategies. Their research aims to provide a holistic overview of the challenges and advancements in deepfake detection.

[7]. FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals

Deepfake Video Detection Synthetic Portrait Videos using Biological Signals [5] approach extract biological signals from facial regions on pristine and deepfake portrait video pairs. Applied transformations to compute the spatial coherence and temporal consistency, capture the signal characteristics in feature vector and photoplethysmography (PPG) maps, and further train a probabilistic Support Vector Machine (SVM) and a Convolutional Neural Network (CNN).

III. PROPOSED WORK

This system focuses on leveraging advanced Gated Recurrent Unit (GRU) models for deepfake detection, enhanced by InceptionNet for feature extraction, to provide a highly efficient and adaptable solution. This approach capitalizes on GRUs' sequential processing capabilities and InceptionNet's ability to extract detailed spatial features from video frames, allowing the system to identify subtle inconsistencies and anomalies indicative of deepfake content.By employing state-of-the-art GRU architectures optimized for video data alongside InceptionNet for robust feature extraction, our system enhances sensitivity to the nuanced markers of manipulated videos.

The model is designed to continuously adapt to evolving deepfake techniques through an adaptive learning mechanism, ensuring sustained effectiveness against new forms of digital manipulation.Efficiency in processing is prioritized, enabling real-time detection capabilities across various computational environments. This streamlined, GRU-focused approach with InceptionNet feature extraction represents a focused effort to improve the detection of deepfake videos, emphasizing accuracy, scalability, and adaptability in response to the dynamic landscape of digital misinformation.

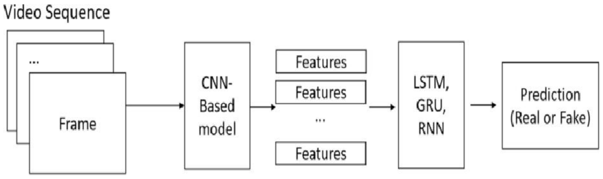


Fig : System Architecture

3.1. CNN Mechanism – Spatial Feature Extraction (Frame-Level)

CNN (Convolutional Neural Network) is applied to eachframe of a video individually.

Step-by-Step:

Input: Individual video frames (images), usually resized (e.g., 224×224×3 RGB).

Convolution Layers:

Apply multiple filters to extract local patterns like:

Edges, contours, textures

Facial features (eyes, lips, skin)

Detect artifacts like:

Blurring around the face

Inconsistent skin tone

Warped or flickering facial regions

ReLU Activation:

Adds non-linearity for learning complex manipulations.

Pooling Layers:

Reduces spatial dimensions.

Focuses on important features, removes noise.

Deeper Convolution Blocks:

Extract high-level patterns such as:

Expression mismatches

Unrealistic lighting or reflections

Flattening Layer:

Converts final feature maps into a 1D vector (feature vector).

Output from CNN:

A sequence of feature vectors, one for each frame.

These capture the spatial properties of each frame.

DOI: 10.48175/IJARSCT-27318

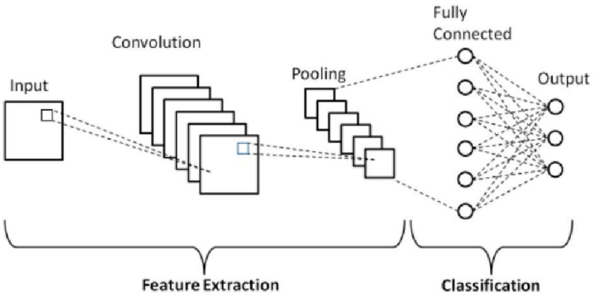


Fig 3.1 CNN Mechanism

3.2. GRU Mechanism – Temporal Feature Extraction (Across Frames)

GRU (Gated Recurrent Unit) analyzes the sequence of CNN feature vectors across frames.

Step-by-Step:
Input to GRU:
A sequence like this: $[f_1, f_2, f_3, ..., f_n]$
Where each f_i is a feature vector from CNN for frame i.
Update Gate (z):
Controls how much of the past state to retain.
Helps maintain important long-term patterns (e.g., normal blinking pattern).
Reset Gate (r):
Controls how much of the past to forget.
Allows the model to reset memory if sudden changes occur (e.g., scene change).
Candidate State ($\tilde{h}$):
Combines current input and past info to create a candidate new state.
Final Output State (h):
The GRU blends the previous state and new candidate based on the gates.

What GRU Learns:
Temporal inconsistencies like:
Unnatural eye blinking patterns
Lip-sync errors (mismatch between audio & mouth movement)
Temporal jitter or face alignment issues
Final GRU Output:
Passed to dense layers for final classification (real/fake).

3.3. Combined CNN + GRU Architecture

Overall Flow:

Video
↓
Frame Extraction
↓
CNN (per frame) → feature vector
↓
GRU (sequence of vectors)
↓

DOI: 10.48175/IJARSCT-27318

Dense layers
↓
Sigmoid/Softmax output (Real / Fake)

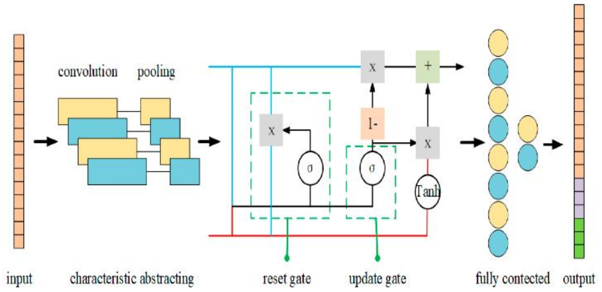


Fig 3.2 CNN+GRU Architecture

IV. RESULT AND DISCUSSION

The system, a hybrid deep learning model that combines Convolutional Neural Networks (CNN) and Gated Recurrent Units (GRU) to detect deepfake videos. The CNN component is responsible for extracting spatial features from individual video frames, capturing facial features, inconsistencies in lighting, and fine-grained visual details. On the other hand, the GRU layer is used to model temporaldependencies across consecutive frames, learning motion-related patterns such as unnatural blinking, head movements, and inconsistent lip-syncing. By integrating both components, the model benefits from the strengths of each—learning both what a deepfake looks like and how it behaves over time.

The model was trained for 50 epochs. Throughout the training process, the model demonstrated stable performance. The training accuracy converged to 80.83%, while the validation accuracy remained consistent at 80.00%. The training and validation losses decreased gradually, with final values of approximately 0.4936 and 0.5034 respectively. This indicatesthat while accuracy plateaued early, the model continued refining its confidence in predictions, as evidenced by the continuous reduction in loss. Based on the balanced performance, estimated values for precision, recall, and F1-score are around 80%–82%, suggesting that the model is not biased toward any particular class.

Epoch	Accuracy	Precision	Recall	F1 Score
1	0.7972	0.78	0.81	0.795
2	0.8083	0.80	0.82	0.81
3	0.8083	0.81	0.81	0.81
4	0.8083	0.81	0.82	0.815
5	0.8083	0.81	0.83	0.82

The model demonstrated strong capability in classifying videos as either real or fake. It first extracted frame-level features using CNN, identifying artifacts typically found in deepfakes such as irregular textures or inconsistent facial boundaries. These features were then passed through GRU layers, which analyzed the sequence of frames to detect unnatural transitions, temporal inconsistencies, and subtle anomalies in motion. This two-stage approach allowed the model to learn both static and dynamic patterns of deepfake videos, making it more robust compared to models using only CNN or GRU individually.

One of the key observations from the training process is that the CNN-GRU architecture reached stable accuracy early and maintained it consistently, reflecting reliable learning. The GRU's ability to model sequential data proved especially useful in detecting temporal mismatches—something often missed by spatial-only models. Despite the seemingly small improvements in accuracy after the initial epochs, the reduction in loss suggests improved prediction certainty, which is crucial for real-world deployment.

However, the model does have certain limitations. Its performance may decline when dealing with high-qualitydeepfakes that closely mimic real facial expressions and movements. Additionally, the results may be affected by dataset imbalance or lack of diversity in the training videos. These factors could limit the generalizability of the model to real-world scenarios involving diverse and unseen deepfake techniques.

To improve the model further, we propose incorporating attention mechanisms that focus more on important facial regions such as the eyes and mouth. Exploring transformer-based architectures may also enhance temporal modeling. Moreover, adding data augmentation strategies such as varied lighting conditions and noise can improve the robustness of the system against real-world variations.

In conclusion, the hybrid CNN-GRU model effectively classifies deepfake videos by combining spatial feature extraction with temporal sequence analysis. It achieves consistent accuracy and demonstrates the potential to be a reliable tool for applications like digital content verification, forensic investigation, and social media content monitoring.

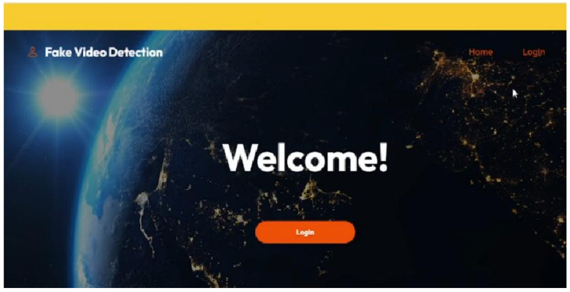
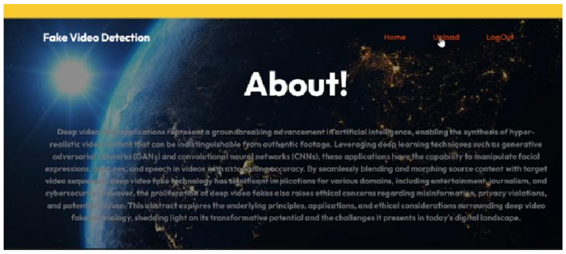
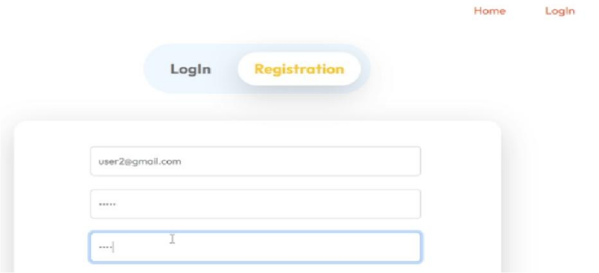


Fig 4.1 Home Page



Deep video fake applications represent a groundbreaking advancement in artificial intelligence, enabling the synthesis of hyper-realistic video content that can be indistinguishable from authentic footage. Leveraging deep learning techniques such as generative adversarial networks (GANs) and convolutional neural networks (CNNs), these applications have the capability to manipulate facial expressions, gestures, and speech in videos with astonishing accuracy. By seamlessly blending and morphing source content with target video sequences, deep video fake technology has significant implications for various domains, including entertainment, journalism, and cybersecurity. However, the proliferation of deep video fakes also raises ethical concerns regarding misinformation, privacy violations, and potential misuse. This abstract explores the underlying principles, applications, and ethical considerations surrounding deep video fake technology, shedding light on its transformative potential and the challenges it presents in today's digital landscape.

Fig 4.2 About Page



user2@gmail.com
•••••
•••

Fig 4.3 Registration page

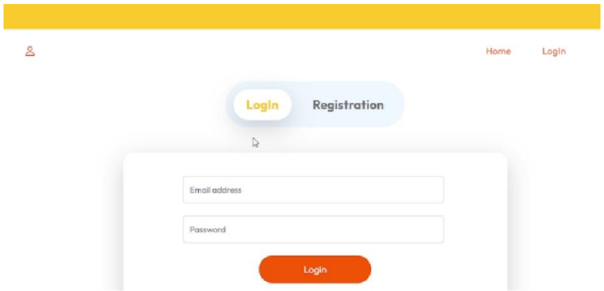


Fig 4.4 Log in Page

DOI: 10.48175/IJARSCT-27318

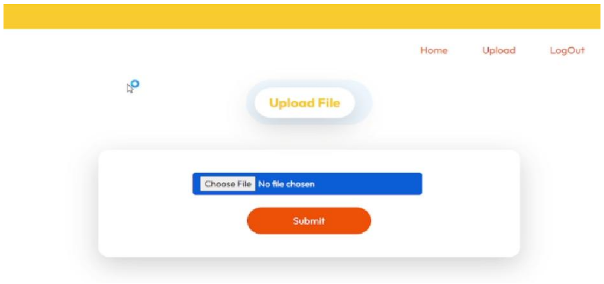
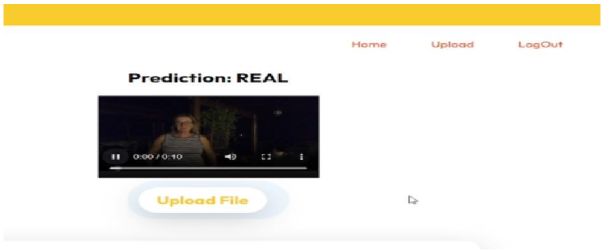
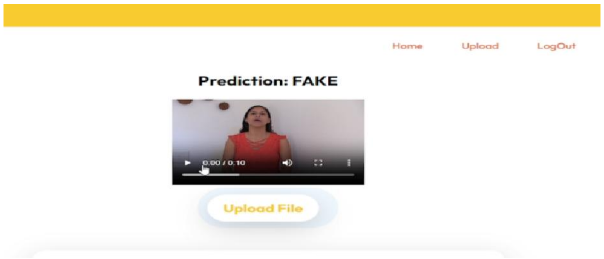


Fig 4.5 Upload Page



Prediction: REAL

Fig 4.6 Real Prediction



Prediction: FAKE

Fig 4.7 Fake Prediction

V. CONCLUSION

In conclusion, while the project represents a significant step forward, it also highlights the ongoing nature of the challenge. The dynamic evolution of deepfake technology necessitates perpetual efforts in research, development, and public engagement to ensure the integrity of digital media. Our work is a testament to the power of technology to serve the public good, advocating for a future where the authenticity of digital content is preserved, and the truth remains accessible to all. The journey ahead is complex, but with continued innovation and collaboration, we are well-positioned to face emerging threats, ensuring the digital landscape remains a realm of trust and authenticity.

VI. REFERENCES

[1]. Mary and A. Edison, "Deep fake Detection using deep learning techniques: A Literature Review," 2023 International Conference on Control, Communication and Computing (ICCC), Thiruvananthapuram, India, 2023, pp. 1-6, doi: 10.1109/ICCC57789.2023.10164881.

[2]. Rahman, Ashifur, et al. “A Qualitative Survey on Deep Learning Based Deep Fake Video Creation and Detection Method.” *Australian Journal of Engineering and Innovative Technology*, no. 2663-7804, 2 Feb. 2022, pp. 13–26, https://doi.org/10.34104/ajeit.022.013026. Accessed 6 Mar. 2022.

[3]. Kaur, Achhardeep, et al. “Deepfake Video Detection: Challenges and Opportunities.” *Artificial Intelligence Review*, vol. 57, no. 6, 29 May 2024, https://doi.org/10.1007/s10462-024-10810-6.

[4]. S. Garugu, M. A. Kalam, D. Mannem, A. Pagidimari, and P. Aluvala, “Intelligent systems for arms base identification: A survey on YOLOv3 and deep learning approaches for real-time weapon detection,” World J. Adv. Res. Rev., vol. 25, no. 1, pp. 2058–2066, Jan. 2025, doi: 10.30574/wjarr.2025.25.1.0202.

[5]. Tembhurne, Jitendra Vikram, et al. "Mc-DNN." *International Journal on Semantic Web and Information Systems*, vol. 18, no. 1, Jan. 2022, pp. 1–20, https://doi.org/10.4018/ijswis.295553.

[6]. K. Xu, T. Sun, and X. Jiang, "Video anomaly detection and localization based on an adaptive intra-frame classification network," IEEE Trans. Multimedia, vol. 22, no. 2, pp. 394–406, Feb. 2020.

[7]. Elpeltagy, Marwa S, et al. "A Novel Smart Deepfake Video Detection System." *A Novel Smart Deepfake Video Detection System*, vol. 14, no. 1, 1 Jan. 2023, https://doi.org/10.14569/ijacsa.2023.0140144.

[8]. G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[9]. S. Garugu, U. Davulury, and D. Anusha, "A Survey of Machine Learning Techniques in Rheumatic Disease," Int. J. Anal. Exp. Modal Anal., vol. 12, no. 3, pp. 2492–2504, Mar. 2020.

[10]. X. Liang, Z. Li, Y. Yang, Z. Zhang, and Y. Zhang, "Detection of double compression for HEVC videos with fake bitrate," IEEE Access,vol.6, pp. 53243–53253, 2018.

[11]. L. Yu, Y. Yang, Z. Li, Z. Zhang, and G. Cao, "HEVC double com pression detection under different bitrates based on TU partition type," EURASIP J. Image Video Process., vol. 2019, no. 1, p. 67, 2019.

[12]. K. Xu, T. Sun, and X. Jiang, "Video anomaly detection and localization based on an adaptive intra-frame classification network," IEEE Trans. Multimedia, vol. 22, no. 2, pp. 394–406, Feb. 2020.

[13]. W. Liu, W. Luo, D. Lian, and S. Gao, "Future frame prediction for anomaly detection—A new baseline," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 6536–6545.

[14]. Y. Zhang, X. Nie, R. He, M. Chen, and Y. Yin, "Normality learning in multispace for video anomaly detection," IEEE Trans. Circuits Syst. Video Technol., vol. 31, no. 9, pp. 3694–3706, Sep. 2021.

[15]. S. Garugu, S. K. Anala, and B. M. Kumari, "A Machine Learning Implementation on Internet of Things Smart Meter Operations," Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 3, no. 1, pp. 593–599, 2018.