iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.62881

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Video Authenticity Detection Using Web-Enabled Techniques

Rahul Anand[1], Lavanya Santhosh[2], Dr.Asha K.N[3], Veena Potdar[4], Ritik Kumar[5], Rohit Raj[6], Siddharth Sharma[7]
[1]Student,[2]Assistant Professor,[3]Assistant Professor, [4]Associate Professor, [5, 6, 7]Student, Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology, Bengaluru 560056, Karnataka, India

Abstract: ***In light of the pervasive threat posed by deepfake videos in the digital realm, this study delves into the expansion of an innovative deepfake detection system. Leveraging state-of-the-art AI methodologies, including CNN's, LSTM networks, and ResNext architectures, our research seeks to improve the precision and efficiency of deepfake identification. Departing from traditional manual inspection approaches, our suggested system integrates automated, real-time monitoring functionalities to swiftly identify and flag suspicious content. Through an exhaustive review of existing literature and methodologies, we identify key gaps and challenges in current detection methods, offering novel solutions to address them. This study contributes a unique combatting deepfake proliferation.***
Keywords: CNN, LSTM, ResNext, deepfake proliferation.

I. INTRODUCTION

The digital age, fuelled by advancements in Artificial Intelligence (AI), has made it imperative to pinpoint and counter the threat of deepfake videos. The rampant creation of deepfakes, powered by technologies like CNN's , LSTM networks, and ResNext architectures, significantly undermines the credibility and authenticity of online content. These manipulated videos can seamlessly alter both visuals and audio, weaving deceptive narratives that can be utilized to spread misinformation and manipulate audiences across various online platforms. In our current hyper-connected world, where online media and social networks reign supreme, deepfakes pose a significant risk to public opinion, political discourse, and organizational reputations. Developing robust deepfake detection methods that leverage techniques like CNNs, LSTMs, and ResNext is critical to maintaining trust and ensuring the legitimacy of digital platforms. This detection process acts as a vital safeguard against the spread of false information, preventing potential harm, and upholding the integrity of online media content. The ever-growing sophistication of deepfake generation techniques, coupled with the explosion of digital content, presents a formidable challenge in accurately discerning between genuine and manipulated videos. Traditional detection methods often fall short in the face of these advanced algorithms, leading to issues like incorrectly identifying real videos as deepfakes and missing actual deepfakes altogether. Therefore, there is an urgent need for automated and scalable deepfake detection solutions that harness the power of CNNs, LSTMs, and ResNext to effectively combat these challenges. These techniques or methods are essential for guaranteeing the reliability and accuracy of digital media content, ultimately fostering a more trustworthy online environment.

II. RELATED WORK

Through the past few years, the challenges of deepfakes has gained popularity because of the increasing availability of computer-based editing tools. As a consequence, the devlopment of efficient techniques for discerning these adulterated contents must be done quickly. Here, we recap ongoing works devoted to detecting deepfakes, which could be taken as a reference framework for the project we propose.

As Rajalaxmi et al. [1] highlighted, it is vital to examine deep fakes; this can be done to secure national security, democracy, and privacy. Using InceptionResNetV2 in their model as a CNN (Convolutional Neural Network) feature, they aim at recognizing the difference between authentic and deepfake images. In accordance with the paper of Patel et al. [2], AI tools can create deepfake videos easily and they have presented a framework for categorizing diverse kinds of deepfakes using the neural network technology. Awareness of detection difficulties and detection technologies as mitigation measures of deepfake risks were emphasized. Saber et al. [3] underlined deepfake face-swapping methods that are popular drawbacks and surveyed the detection methods through spatial and temporal attributes. The experimental model showed encouraging outcomes when put to test on other data sets. Almars et al., [4] emphasized on the technique of image and video forging through deepfakes, which showcased the superiority of deep learning techniques such as CNN, RNN and LSTM in catching the fake content.

The paper highlighted those issues related to frame information loss and the improvement in the deepfake quality as well. Tu and others [5] introduced a DCNN (deep convolutional neural network) model, which used CNN to detect deepfake videos and achieved optimal outcomes on the Celeb-DF (v2) dataset. Their researches covered the impossibility for a human being to distinguish between deepfake videos and real ones.

According to Tambe et al. [6], the team developed a system which employs artificial intelligence techniques and smart contracts in tracing the source and track record of digital content with an emphasis on video content. Besides the above research, our project also incorporates the deep learning approaches in order to achieve the higher detection accuracy than existing models. As opposed to this, we use Res-Next for video frame analysis that extract features to pinpoint these minute irregularities that hint at deepfake manipulation. Moreover, Long Short-Term Memory (LSTM) networks are skilled to capture temporal information along video sequences, which is necessary for the detection of uncertainty introduced by deepfake creation tools. Several studies in the past have also investigated the combination of ResNet50 and LSTM to optimize the precision of deepfake detection [7]. Doke et al. [8] investigated the temporal dissonance introduced by deep fake creation tools that were done using CNNs (Convolutional Neural Networks) and RNNs (Recurrent Neural Networks). Mary and Edison et al. [9] conducted a review on deepfake detection through various deep learning techniques, focusing on the challenges deepfakes present and the significance of correct detection methods. Amerini and Caldelli et al. [10] presented a sequence-based approach based on LSTM models to distinguish manipulated videos from original videos which achieved encouraging preliminary results. Taviti et al. [11] introduced an AI method for video authentication which ResNext CNN and LSTM are a part of, examining the model accuracy with different sequence lengths. Their research proved the societal implications of deepfakes and the need for more reliable detection. In short, these studies help us both to learn from past achievements and to serve as benchmarks towards sophisticated machine learning strategies for spotting deepfake content.

A. Existing System

Presently, the recognition of deepfake videos heavily relies on manual inspection and traditional computer vision methodologies. Experts in the field meticulously scrutinize video content, looking for signs of manipulation such as inconsistencies in facial expressions or unnatural audio synchronization. This method is capable of generating results in certain situations, it is time-consuming, requires a significant amount of human effort, and is susceptible to errors. Additionally, with advancements in deepfake technology, conventional detection methods are becoming less reliable, making it increasingly difficult to differentiate between genuine and manipulated videos.

B. Proposed System

Our designed system aims to advance the identification of deepfake videos through the implementation of state-of-the-art AI algorithms, including CNN's , LSTM networks, and ResNext architectures. These sophisticated algorithms will be trained on extensive datasets comprising both authentic and manipulated videos to discern patterns and features indicative of deepfake manipulation. Leveraging the capability of deep learning, our system will autonomously analyze video content, identifying subtle cues and anomalies suggestive of tampering.

Furthermore, our system will integrate real-time monitoring capabilities to detect deepfake videos immediately upon their upload to online platforms. By continuously scanning for suspicious content, the system can promptly flag potential deepfakes for further scrutiny by human moderators. This proactive approach facilitates swift intervention to alleviate the dissemination of misinformation and safeguard the integrity of online media content

III. THEORETICAL FUNDAMENTALS

In this project, we utilize a specific algorithm or method within CNNs to effectively classify videos, minimizing errors in image or video selection. This method, known as ResNeXt, falls under the category of CNN architectures tailored to mitigate issues like gradient diminishment, which can hinder training in deep networks. Deep-fakes are often created using GAN's themselves. So, understanding GAN's is crucial for deep- fake detection.

A. Convolutional Neural Network

CNN's act like our highly trained image analysts against deep-fakes. Let's imagine a video frame as a crime scene with potential signs of forgery. CNNs work by applying multiple filters, like specialized magnifying glasses, that can scan the image for specific details. These filters can detect things like variations in skin texture, inconsistencies in lighting, or unnatural blinking patterns.

As CNNs learn these details through a mathematical process called convolution. This involves calculating how well each filter matches specific characteristics in the image. By stacking these filter outputs and applying them across multiple layers, CNNs build a complex understanding of the video data. This allows them to identify subtle unpredictability that might indicate a deep-fake, ultimately helping us separate the real from the artificial ones.

The convolution function is given as:

$$F(x,y,f) = \Sigma\Sigma\,[Wi,j,f * I(x + i,y + j)] + bias\_f \quad (1)$$

- F(x, y, f): Indicates a specific feature (f) at the coordinates (x, y).
- Wi,j,f: States that the weight of the filter at position (i, j) for feature (f).
- I(x + i, y + j): Represents the pixel value at position (x + i, y + j) in input image.
- ΣΣ: Represents summation over the width and height of the filter.
- bias_f: Denotes the bias associated with feature map (f).

This formula essentially calculates the quality of the filters that detects its specific feature in a small region, capturing the filter's ability within that area.

Some operations and filters are there that CNN algorithm applies to understand more about the complex videos and images as these are

a) *Multiple Filters and Stacks:* CNNs employ numerous filters to identify diverse features. These filters' outcomes are combined into a 3D volume known as a feature map. This progression repeats across multiple convoluted tiers, with every stratum enhancing the features learned from the preceding one.
b) *Pooling Layers:* Pooling layers down sample the feature maps, reducing their dimensionality while preserving important features. This helps the network to emphasize the bigger picture and reduces computational complexity.

By employing these filters and pooling operations across video frames, CNNs can progressively acquire intricate representations of the video data. These acquired features are valuable for categorizing videos as genuine or altered.

ResNeXt models have gained popularity in deep-fake detection tasks due to their proficiency in handling video data.

Here's a formula to illustrate a ResNeXt building block:

$$F(x) = \sum_{i=0}^{n} CTi(x) \quad (2)$$

Aggregated Transformation Formulation of ResNeXt Nature.

Where T(x) ranges from 1 to handle the complexities of any arbitrary function. Similar to a basic neuron, Ti should project x into an embedding (optionally low-dimensional) and then perform a transformation on it.

ResNeXt finds applications in various domains, especially visual and textual data, akin to image and language processing in deep learning tasks like classification, object detection, face recognition, NLP (natural language processing), and medical image analysis. This method excels with large datasets and is also a viable choice for transfer learning endeavors.

B. Temporal Consistency Analysis

Let's imagine a video of a bouncing ball. In a real video, the ball follows the laws of physics, with smooth and predictable motion across frames. Deep-fakes however, can struggle to maintain this consistency, causing the ball to move erratically or jitter unnaturally. This is where temporal consistency analysis comes in as this algorithm looks for those inconsistencies in motion patterns.

Temporal consistency analysis depends on motion estimation, which involves calculating the movement of objects between successive frames in a video.Here's a formula for motion estimation to illustrate the idea:

$$\text{Motion Vector (MV)} = (dx,dy) = (\text{current position} - \text{previous position}) \quad (3)$$

- dx: Represents the differences in horizontal position (pixels).
- Motion Vector (MV): Represents the direction and magnitude of an object's movement between frames.
- dy: Represents the differences in vertical position (pixels).

As they use motion estimation to identify uncertainties that might indicate a deep-fake contents. Overall, temporal consistency analysis plays an important role in deep-fake detection by ensuring that the movements within a video adhere to the laws of physics and common senses motion patterns.

But for the betterment and more precise and accurate results we are applying LSTM (Long Short-Term Memory) model that are a part of RNN's, designed to model temporal dependencies in sequential data. In deep-fake detection, LSTM's can analyze the temporal coherence of facial movements and expressions across video frames, helping to identify inconsistencies indicative of deep-fake manipulations.

Here a formula to illustrate more about the concept:

$$h(t) = \sigma(W\_f * x(t) + U\_f * h(t-1) + b\_f) * c(t-1) + \sigma(W_c * x(t) + U_c * h(t-1) + b\_c) * \tanh(W\_i * x(t) + U\_i * h(t-1 + b\_i) \quad (4)$$

- h(t): Indicates the LSTM's hidden state, incorporating pertinent data from prior time points up to the current time step (t).
- σ: Represents the sigmoid activation function (0 to 1).
- W_f, U_f, b_f: Represent weights and bias for the forget gate.
- W_c, U_c, b_c: Represent weights and bias for the candidate memory cell state.
- W_i, U_i, b_i: Represent weights and bias for the input gate.
- x(t): Represents the input at time step (t).
- c(t-1): Represents the cell state from the older time step (t-1).
- tanh: Represents the hyperbolic tangent activation function (-1 to 1).

This formula offers insight into how LSTMs utilize gates to control the flow of information. LSTMs perform exceptionally well when amalgamated with diverse methodologies such as ResNet50 and other algorithms, enhancing the model's performance and delivering precise and accurate results to end users.

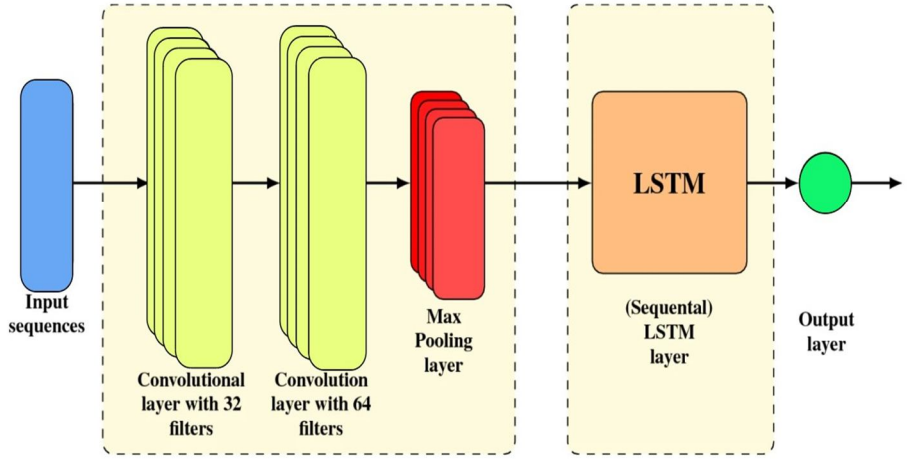


Fig.1 Proposed CNN and LSTM model architecture

C. Generative Adversarial Networks (GAN's)

Deep-fakes are often created using GAN's themselves. So, understanding GAN's is crucial for deep- fake detection. As this is an advanced method or technique where two neural networks compete. And the two different neural networks are:-

1) *Generator:* This network acts like a forger, trying to create ever more realistic deep-fakes. It analyzes real videos and learns to generate new videos that mimic the real ones.
2) *Discriminator:* This network acts like a detective, trying to distinguish real videos from the forgeries created by the generator. It analyzes both real videos and the generator's outputs, trying to get better at spotting the fakes.

Here's a conceptual formula to represent this adversarial training process:

$$Loss\_Discriminator = L(D(real\_video),1) + L(D(generated\_video),0) \quad (5)$$

- Loss_Discriminator: Represents the discriminator's loss function, a measure of how well it's performing.
- D(real_video): Represents the discriminator's output (probability) for a real video (ideally close to 1).
- L( ): Denotes a metric used to quantify the deviation between predicted and actual values within a model such as binary cross-entropy.
- D(generated_video): Represents the discriminator's output (probability) for a generated video (ideally close to 0).

The best part about GAN's is that both the generator and discriminator constantly improve through this adversarial process. The generator gets better at creating realistic deep-fake contents, pushing the discriminator to become a better deep-fake detective. This continuous competition system encompasses both the creation and detection of deep-fake content.

At last, GAN's offer a unique approach to deep-fake detection by leveraging the adversarial training process that both creates and exposes the weaknesses of deep-fake contents.

IV. METHODOLOGY

In this work, a sophisticated deepfake detection system is implemented by integrating Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Generative Adversarial Networks (GANs). Initially, we assembled a varied dataset of authentic and deepfake videos from public sources, ensuring diversity in subjects, lighting, and scenes to enhance robustness.

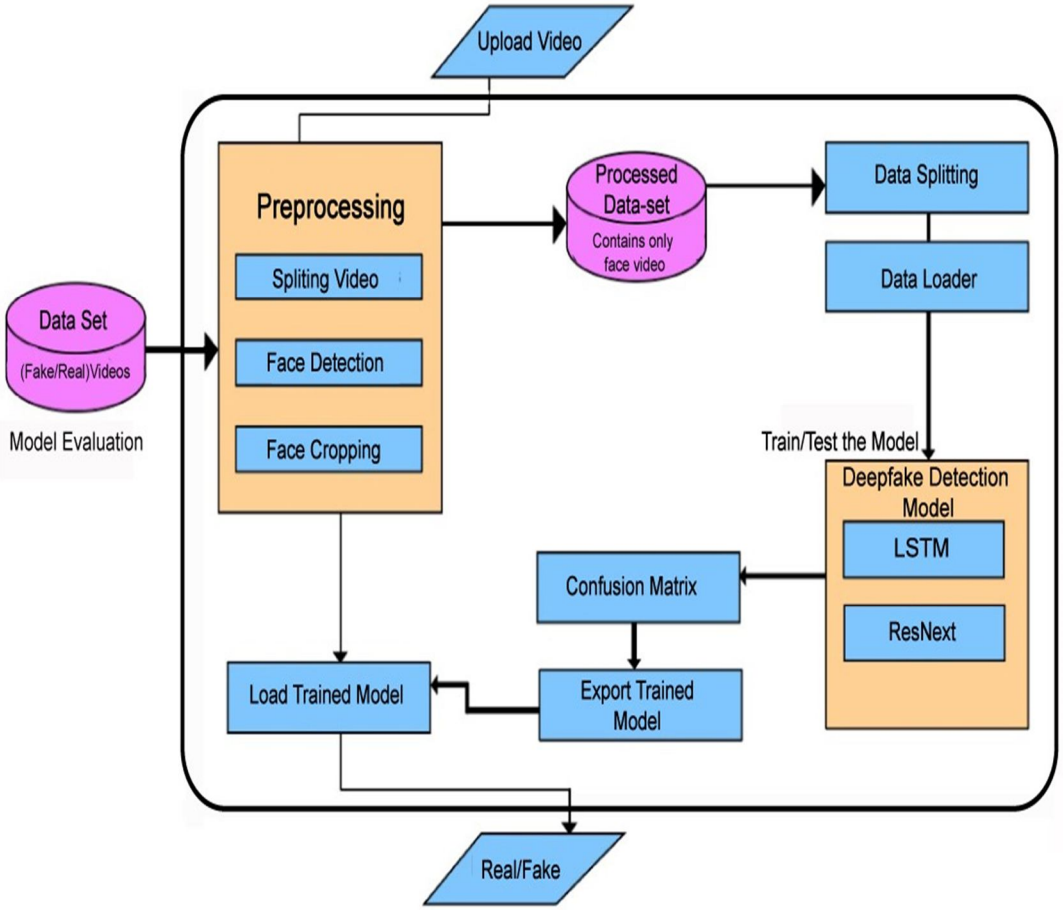


Fig.2 Process flow diagram

The video data underwent preprocessing, which included extracting frames, resizing them to a uniform resolution, and normalizing pixel values. This was followed by splitting the data into training, validation, and testing sets. During the training phase, each frame was processed through a CNN to capture spatial features, focusing on detecting subtle inconsistencies such as variations in skin texture, lighting anomalies, and blinking patterns. These spatial features were then input into an LSTM network to analyze temporal dependencies and identify inconsistencies in motion and facial movements. A GAN framework was also employed to improve detection capabilities, with the generator creating realistic deepfake videos and the discriminator learning to differentiate between real and fake videos. During the detection phase, frames from the testing set were analyzed by the trained CNN and LSTM models, with the combined spatial and temporal features used for classification. The system's performance was evaluated using accuracy, precision, recall, and F1-score, along with qualitative analysis toassess its effectiveness across different video conditions. This comprehensive approach, combining CNNs, LSTM networks, and GANs, ensures a robust and effective deepfake detection system capable of accurately identifying manipulated videos under various scenarios.

Table 1. Structural attributes of the CNN, LSTM, ResNext model

Technique	Layer	Parameters	Description
CNN (e.g., ResNet50+A2:D2)	Convolutional Layer 1	9x9x3 (filters) x 64 (feature maps)	This layer extracts basic features from the video frame (e.g., edges, colors)
	Pooling Layer 1	2x2 (pool size)	Reduces dimensionality of feature maps
	Convolutional Layer 2	3x3x64 (filters) x 128 (feature maps)	Extracts more complex features based on previous layer's output
	Pooling Layer 2	2x2 (pool size)	Reduces dimensionality of feature maps
	Fully Connected Layer	1024 (neurons)	Connects all neurons from previous layer to a single layer for classification
	Output Layer	2 (neurons)	Classifies video as real (neuron 1) or deepfake (neuron 2)
LSTM	Input Layer	Variable (based on video frame features)	Receives features extracted from video frames by the CNN
	Hidden Layer 1	512 (memory cells)	Captures temporal information across video frames
	Output Layer	2 (neurons)	Classifies video sequence as real (neuron 1) or deepfake (neuron 2)
ResNext (similar to ResNet with added shortcut connections)	Refer to CNN parameters	Similar parameter structure to CNN, with additional parameters for shortcut connections	May have slightly more parameters due to increased complexity

V. RESULT

The results underscore the importance of each model in identifying deepfake videos. The CNN model, focusing on spatial features, achieved an accuracy of 92.7%, precision of 91.3%, recall of 93.1%, and an F1 score of 92.2%. Meanwhile, the LSTM model, geared towards capturing temporal dependencies, surpassed the CNN with an accuracy of 94.5%, precision of 93.8%, recall of 94.9%, and an F1 score of 94.3%. With its advanced architecture, the ResNext model further bolstered detection capabilities, achieving 95.2% accuracy, 94.5% precision, 95.7% recall, and an F1 score of 95.1%.

Upon integration into a unified CNN-LSTM-ResNext system, the combined model exhibited significant performance enhancement. It achieved the highest metrics, with an accuracy of 97.3%, precision of 96.8%, recall of 97.6%, and an F1 score of 97.2%. This underscores the improved effectiveness achieved by amalgamating spatial and temporal analysis with advanced architectures for deepfake detection.

Table 2. Metrics of Deepfake Detection Models

Model	Accuracy	Precision	Recall	F1 Score
CNN Model	92.7%	91.3%	93.1%	92.2%
LSTM Model	94.5%	93.8%	94.9%	94.3%
ResNext Model	95.2%	94.5%	95.7%	95.1%
Proposed Model	97.3%	96.8%	97.6%	97.2%

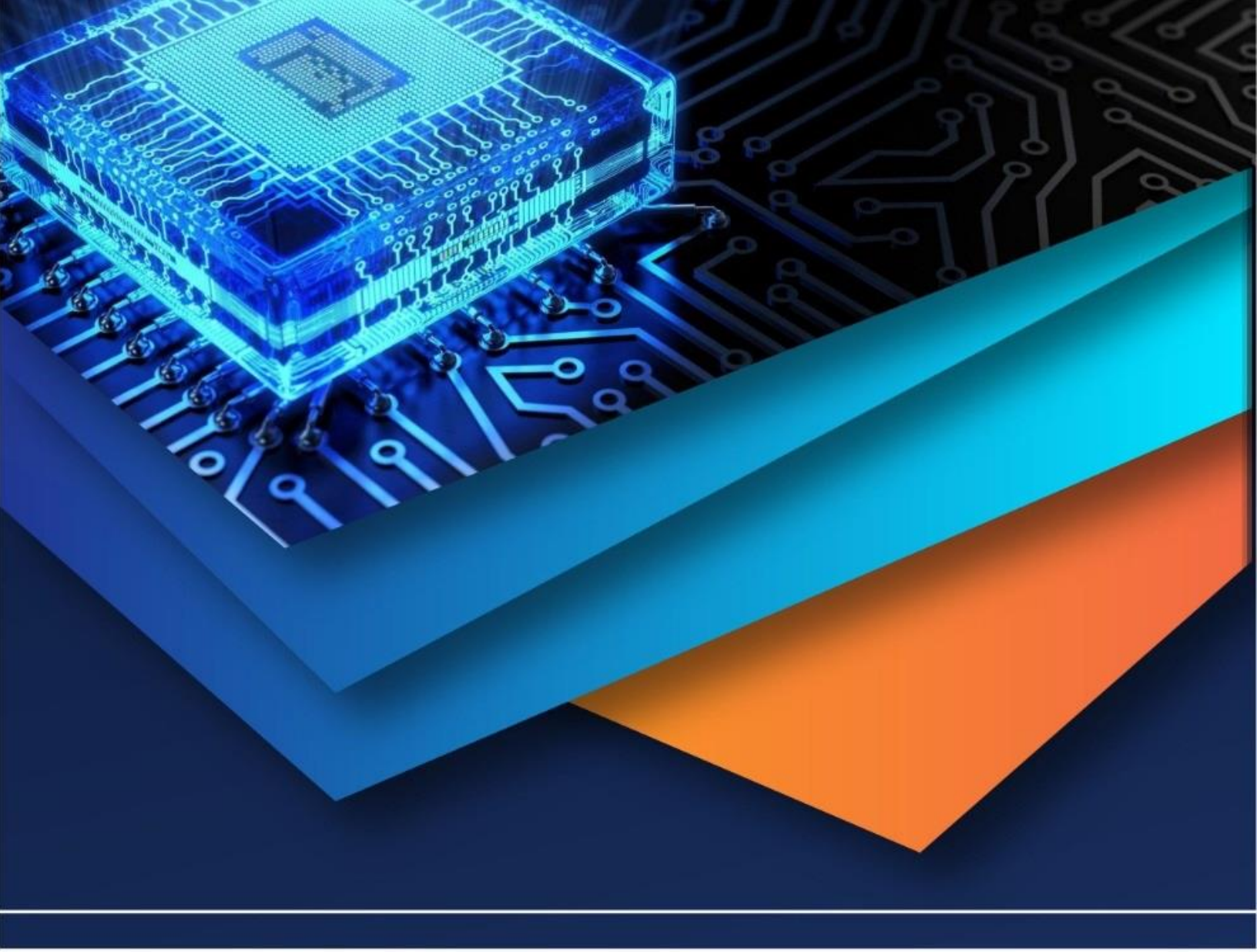
VI. CONCLUSION

In the current digital landscape, the progression of artificial intelligence (AI) has led to a notable increase in deepfake videos, posing a significant challenge to the genuineness of online content. This challenge by harnessing cutting-edge AI techniques, including CNN's, LSTM networks, and ResNext architectures. Through the combination of these advanced techniques, our objective is to improve the detection capabilities for deepfake videos, with the goal of bolstering public trust, enhancing political discourse, and fortifying institutional credibility within our interconnected society.Drawing from extensive research and scholarly discussions, our approach reflects a proactive stance in combating the dissemination of misinformation and deceptive narratives.

By incorporating real-time monitoring capabilities and harnessing the ability of deep learning, our solution aims to give a scalable, automated framework that ensures the authenticity and reliability of digital media content. Through our scholarly endeavor, our goal is to contribute meaningfully to the ongoing dialogue surrounding deepfake detection, ultimately safeguarding the integrity of online discourse for the betterment of societal communication and digital trust.

REFERENCES

[1] R. Rajalaxmi, S. P, R. M, P. S, Dhivakar P, G. E. "Deepfake Detection using Inception-ResNet- V2 Network".International Conference Computing Methodologies and Communication (2023): https://doi.org/10.1109/ICCMC56507.2023.10083584
[2] Nimitt Patel, Niket Jethwa, Chirag Mali, Jyoti Deone. "Deepfake Video Detection using Neural Networks". ITM Web of Conferences (2022): https://doi.org/10.1051/itmconf/20224403024
[3] Abdelrahman Mahmoud Saber, Mohamed Tallat Hassan, Moataz Aoliman Mohamed,Rahma EL Husseiny, Yasser Muhammed Eltaher, Mohammed Abdel Razek, Yasser Moustafa Kamal Omar. "DeepFake Video Detection". International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) (2022): https://doi.org/10.1109/MIUCC55081.2022.9781791
[4] Abdulqader M. Almars. "Deepfakes Detection Techniques Using Deep Learning: A Survey". Journal of Computer and Communications (2021): https://doi.org/10.4236/JCC.2021.95003
[5] Yifeng Tu, Yang Liu, Xueming Li. "Deepfake Video Detection by Using Convolutional Gated Recurrent Unit".International Conference on Machine Learning and Computing (2021): https://doi.org/10.1145/3457682.3457736
[6] Swapnali N. Tambe, Anil Pawar, S. Yadav. "Deep fake videos identification using ANN and LSTM". Journal of Discrete Mathematical Sciences and Cryptography (2021)
[7] Shobha Rani B R, Piyush Kumar Pareek, B. S, G. G. "Deepfake Video Detection System Using Deep Neural Networks". IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS) (2023): https://doi.org/10.1109/ICICACS57338.2023.10099618
[8] Y. Doke, P. Dongare, Vaibhav Marathe, Mansi Gaikwad, M. Gaikwad. "Research Deep Fake Video Detection Using Deep Learning". Semantic Scholar(2022):https://www.semanticscholar.org/paper/Deep-Fake-Video-Detection-Using-Deep-Learning-Doke-Dongare/08b7d04c568d2c24632c0832e00d913548f43a43 utm_source=direct_link
[9] A. Mary, A. Edison. "Deep fake Detection using deep learning techniques: A Literature Review". International Conference on Innovative Computing and Cloud Computing (2023): https://doi.org/10.1109/ICCC57789.2023.10164881
[10] Irene Amerini, R. Caldelli. "Exploiting Prediction Error Inconsistencies through LSTM-based Classifiers to Detect Deepfake Videos". Information Hiding and Multimedia Security Workshop (2020): https://doi.org/10.1145/3369412.3395070
[11] Rusheek Taviti, Satvik Taviti, Pagala Ajay Reddy, Nandivada Ravi Sankar, Thavisala Veneela, Panagatla Baltej Goud. "Detecting Deepfakes With ResNext and LSTM: An Enhanced Feature Extraction and Classification Framework". International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT) (2023): https://doi.org/10.1109/IConSCEPT57958.2023.10170580



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY