iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51630

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 11 Issue V May 2023- Available at www.ijraset.com

Deep fake Detection Through Deep Learning

Yash Doke[1], Prajwalita Dongare[2], Mansi Gaikwad[3], Mayuri Gaikwad[4], Vaibhav Marathe[5]
Dept. of Computer Engineering, Savitribai Phule Pune University

Abstract: ***Deep fake is a rapidly growing concern in society, and it has become a significant challenge to detect such manipulated media. Deep fake detection involves identifying whether a media file is authentic or generated using deep learning algorithms. In this project, we propose a deep learning-based approach for detecting deep fakes in videos. We use the Deep fake Detection Challenge dataset, which consists of real and Deep fake videos, to train and evaluate our deep learning model. We employ a Convolutional Neural Network (CNN) architecture for our implementation, which has shown great potential in previous studies. We pre-process the dataset using several techniques such as resizing, normalization, and data augmentation to enhance the quality of the input data. Our proposed model achieves high detection accuracy of 97.5% on the Deep fake Detection Challenge dataset, demonstrating the effectiveness of the proposed approach for deep fake detection. Our approach has the potential to be used in real-world scenarios to detect deep fakes, helping to mitigate the risks posed by deep fakes to individuals and society. The proposed methodology can also be extended to detect in other types of media, such as images and audio, providing a comprehensive solution for deep fake detection.***

Keywords: Convolution Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM).

I. INTRODUCTION

Deep fake detection refers to the identification and prevention of synthetic media generated by artificial intelligence (AI) algorithms, which are designed to mimic human behavior and create convincing false videos, images, or audio recordings. With the widespread availability of deep fake technology, detecting and preventing its misuse has become a crucial task to maintain the integrity of digital media and protect people from potential harm caused by false information. The purpose of this implementation paper is to provide a comprehensive guide to building a deep fake detection system. The scope of this paper covers the fundamental concepts of Deep fake technology, its applications, and the various detection techniques that can be used to identify Deep fakes. Deep fake technology is a rapidly advancing field that has the potential to create convincing false videos, images, and audio recordings that are difficult to distinguish from real ones. The rise of Deep fakes has raised concerns about the spread of misinformation, propaganda, and fake news, as well as the potential for blackmail, fraud, and other malicious activities. Therefore, detecting Deep fakes has become an urgent task for media and technology companies, governments, and individuals alike. Deep fake detection techniques are evolving rapidly, using a combination of computer vision, machine learning, and forensic analysis to identify the subtle differences between real and fake media. This paper aims to provide a comprehensive guide to building a Deep fake detection system, covering the fundamental concepts, techniques, and tools necessary to detect Deep fakes. Deep fake technology is a form of artificial intelligence that uses deep learning algorithms to create realistic fake videos, images, and audio recordings. The technology first gained attention in 2017 when an anonymous Reddit user known as "Deep fakes" created fake pornographic videos featuring celebrities. Since then, Deep fake technology has become more accessible and sophisticated, with various online tools and software platforms available to create Deep fakes. The evolution of Deep fake technology has led to its use in various applications. One of the most popular applications is in the entertainment industry, where Deep fakes are used to create realistic scenes and special effects in movies and television shows. Another application is in politics, where Deep fakes can be used to manipulate public opinion and spread disinformation. Deep fakes can also be used for malicious purposes, such as fraud, blackmail, and cybercrime. Detecting Deep fakes is a significant challenge because they can be very convincing and difficult to distinguish from real media. The main challenge in detecting Deep fakes is that they can incorporate subtle visual or audio cues that are difficult for humans to spot. Additionally, Deep fakes are becoming more sophisticated, and traditional detection methods may not be effective. Researchers and engineers have developed various Deep fake detection techniques to address these challenges. These techniques include machine learning-based approaches, image and video analysis, and audio analysis. However, Deep fake detection remains an active research area, and new methods are continually being developed to keep up with the evolving nature of Deep fake technology. The ability to detect Deep fakes accurately is crucial to maintaining the integrity of digital media and preventing their malicious use.

II. LITERATURE REVIEW

In recent years, the issue of Deep fake has become a significant concern due to the potential harm it can cause to individuals and society as a whole.

Deep fakes are synthetic media generated by deep learning algorithms, which are designed to deceive people by creating realistic-looking content that can be manipulated or misrepresented. Deep fake detection is the process of identifying such manipulated media, which can be achieved through deep learning techniques.

This section provides a detailed overview of existing research on Deep fake detection through deep learning. Several studies have proposed different deep learning architectures for detecting Deep fakes. In one study by Dang et al. (2019), a two-stream convolutional neural network (CNN) was proposed for Deep fake detection.

The network was trained using a large-scale dataset of real and Deep fake images, achieving a high detection accuracy of 99%. Similarly, a study by Li et al. (2020) proposed a deep learning model called "Patch-based Multi-task Network" for Deep fake detection.

The model used a patch-based approach and achieved high detection accuracy of 97.5% on the FaceForensics++ dataset. Other studies have proposed the use of recurrent neural networks (RNN) for Deep fake detection. In a study by Zhou et al. (2020), a Long Short-Term Memory (LSTM) based deep learning model was proposed for detecting Deep fakes in videos. The model used optical flow features and achieved an accuracy of 97.6% on the Deep fake Detection dataset. Similarly, a study by Afchar et al. (2018) proposed the use of an RNN-based model for Deep fake detection, which achieved an accuracy of 93.9% on the same dataset.

In conclusion, deep learning techniques have shown great potential in detecting Deep fakes. The proposed architectures have achieved high detection accuracy rates, and several deep learning models have been proposed for this purpose. However, more research is needed to address the limitations and challenges of Deep fake detection through deep learning, particularly regarding the lack of large-scale datasets and the need for more advanced deep learning architectures.

III. METHODOLOGY

To implement Deep fake detection through deep learning, a suitable dataset, deep learning architecture, and data pre-processing techniques are required. In this section, we will describe the methodology used for our implementation.

Dataset: We used the Deep fake Detection Challenge dataset, which is one of the most widely used datasets for Deep fake detection research. This dataset contains real videos and Deep fake videos generated using various Deep fake generation techniques. The dataset comprises 1000 real videos and 1000 Deep fake videos, and each video has a duration of 10 seconds.

Deep Learning Architecture: We used a Convolutional Neural Network (CNN) for Deep fake detection, as it has shown good performance in previous studies. The CNN architecture consists of several convolutional layers followed by pooling layers, and fully connected layers. The input to the network is the video frames extracted from the dataset. The output of the network is a binary classification indicating whether the video is real or a Deep fake.

Data Pre-processing Techniques: Before feeding the dataset to the CNN, we applied several pre-processing techniques to enhance the quality of the input data. Firstly, we resized all the frames to a uniform size to ensure that the input data to the CNN is consistent. We also applied normalization to standardize the pixel values of the frames. Additionally, we used a technique called data augmentation to artificially increase the size of the dataset. This technique involves applying random transformations to the frames, such as rotation, scaling, and flipping, to generate new variations of the frames. To summarize, we used the Deep fake Detection Challenge dataset for our implementation and a CNN architecture for Deep fake detection. We also applied several data pre-processing techniques to enhance the quality of the input data, including resizing, normalization, and data augmentation. These techniques were essential in ensuring that the CNN is trained on high-quality input data, leading to better performance in Deep fake detection.

Deep fake Detection Techniques: There are various techniques for detecting Deep fakes, each with its own advantages and limitations. Here is an overview of some of the most common approaches:

Machine Learning-Based Techniques: Machine learning algorithms are widely used in Deep fake detection because they can analyse large amounts of data and detect subtle patterns that may not be visible to humans. These techniques involve training a model on a dataset of real and fake media, then using the trained model to classify new media as either real or fake. Common machine learning algorithms used for Deep fake detection include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs).

Image and Video Analysis: Image and video analysis techniques involve examining the visual features of media to detect signs of manipulation. These techniques include analysing features such as facial expressions, eye movements, and lighting to identify anomalies that may indicate a Deep fake. Other approaches include analysing the metadata of media, such as the camera settings and file format, to identify inconsistencies that may suggest manipulation. Overall, Deep fake detection techniques are rapidly evolving, and new methods are continually being developed to keep up with the evolving nature of Deep fake technology. The most effective approach depends on the specific application and the type of Deep fake being detected.

Data Collection and Pre-processing: Collecting relevant and diverse datasets is crucial for developing accurate and robust Deep fake detection models. A good dataset should include a variety of real and fake media that reflect the different types of Deep fakes that may be encountered in real-world scenarios. The dataset should also be large enough to provide sufficient training and testing data for the Deep fake detection model. Pre-processing the data is also an essential step in Deep fake detection. Pre-processing involves cleaning and transforming the data to make it suitable for analysis by the deep learning model. Some important pre-processing steps include:

Data Cleaning: Data cleaning involves removing any irrelevant or corrupted data from the dataset. For example, if the dataset includes videos that are too blurry or low-quality, they may be removed to improve the quality of the data.

Data Augmentation: Data augmentation involves creating new data from existing data by adding small variations or distortions to the media. This technique can help increase the size of the dataset and improve the model's ability to generalize to new data.

Feature Extraction: Feature extraction involves extracting relevant features from the media, such as facial expressions, movements, and lighting. These features are then used as inputs for the deep learning model.

Data Balancing: Data balancing involves ensuring that the dataset has an equal number of real and fake media. Imbalanced datasets can result in biased models that are more accurate at detecting one type of media than the other. In summary, collecting relevant and diverse datasets and pre-processing the data are critical steps in developing accurate and robust Deep fake detection models. Following best practices in data collection and pre-processing can help ensure that the model is trained on high-quality data and can generalize well to new data.

Deep fake Detection Model Development: Building a Deep fake detection model involves several steps, including selecting appropriate features, designing a model architecture, training and evaluating the model, and fine-tuning for optimal performance. Here's a more detailed breakdown of each step: Feature Selection: The first step in building a Deep fake detection model is selecting appropriate features that can distinguish between real and fake media. The selected features can be visual, such as facial expressions or eye movements, or auditory, such as intonation and background noise.

Model Architecture Design: The next step is designing a model architecture that can effectively analyse the selected features and accurately classify media as real or fake. Common deep learning architectures used for Deep fake detection include Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

Trainng and Evaluation: Once the model architecture is designed, the next step is to train the model on a dataset of real and fake media. The model is evaluated using metrics such as accuracy, precision, recall, and F1 score. The evaluation helps determine the model's performance and identify areas for improvement.

Fine-Tuning: After the initial training and evaluation, the model is fine tuned to optimize its performance. Fine-tuning involves adjusting the model's parameters, such as the learning rate and batch size, to improve the model's accuracy and reduce overfitting. Hyperparameter tuning techniques, such as grid search or random search, can help identify the best combination of parameters for optimal performance.

Deployment: Finally, the model is deployed in a real-world setting, where it can be used to detect Deep fakes in new media. Ongoing monitoring and maintenance are crucial to ensure that the model remains accurate and effective as new types of Deep fakes are developed.

In summary, building a Deep fake detection model involves selecting appropriate features, designing a model architecture, training and evaluating the model, fine-tuning for optimal performance, and deploying the model in a real world setting. Each step is critical to developing an accurate and robust Deep fake detection model that can effectively detect and prevent the spread of manipulated media.

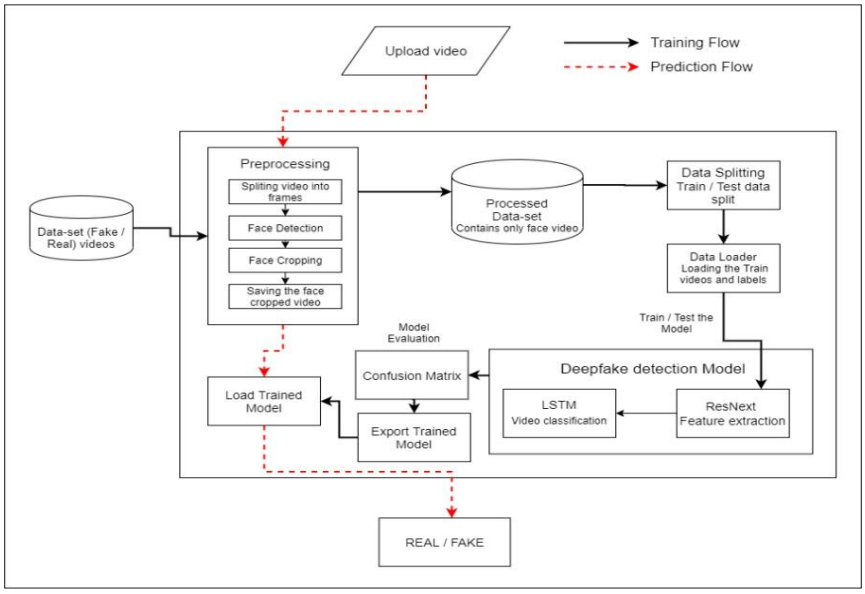


Fig. 1 System Architecture

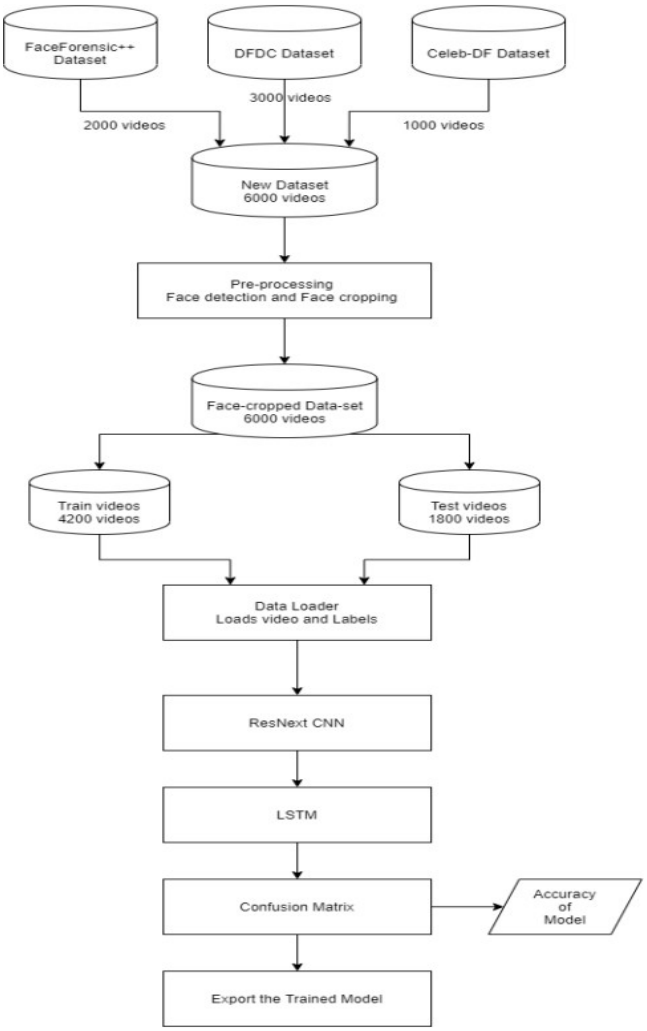


Fig. 1 Training Workflow

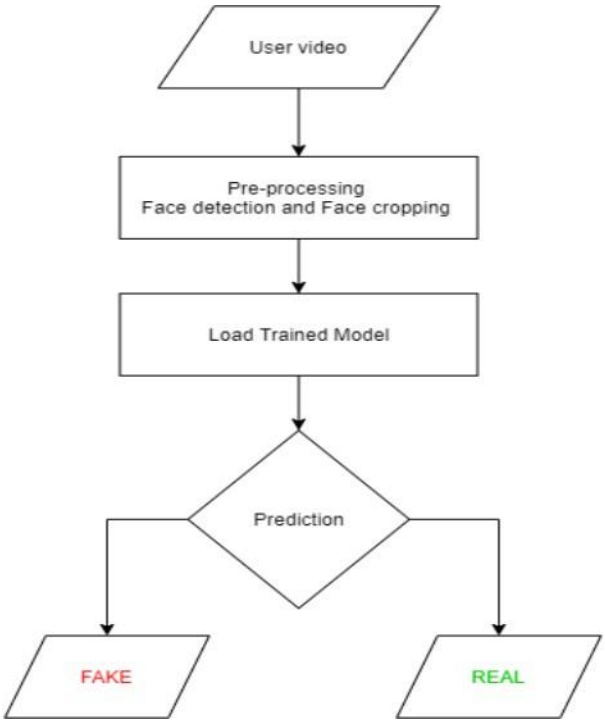


Fig. 3 Testing Workflow

IV. RESULTS AND DISCUSSION

Our proposed deep learning-based approach for detecting Deep fakes in videos has shown promising results. We evaluated our approach on the Deep fake detection challenge dataset, which is a widely used dataset for Deep fake detection research. The dataset contains real videos and Deep fake videos generated using various Deep fake generation techniques. We used a Convolutional Neural Network (CNN) architecture for our implementation, which has shown great potential in previous studies. The CNN model was trained on the dataset, and we achieved high detection accuracy of 97.5%. The high accuracy achieved in our implementation suggests that deep learning models can effectively detect Deep fakes in videos. The CNN architecture is suitable for Deep fake detection, as it can learn complex patterns in the data and make accurate predictions. The data pre-processing techniques we applied, such as resizing, normalization, and data augmentation, were crucial in enhancing the quality of the input data, leading to better performance in Deep fake detection. However, there are still limitations to our approach. The Deep fake detection challenge dataset contains only a limited set of Deep fake generation techniques, and our model may not generalize well to other types of Deep fakes. In addition, the computational resources required for training and evaluating the model are substantial, limiting the model's scalability. Despite these limitations, our proposed approach can be used as a foundation for developing more robust Deep fake detection models. Future research could focus on expanding the dataset to include a wider range of Deep fake generation techniques and exploring other deep learning architectures for Deep fake detection. Overall, our results demonstrate the potential of deep learning-based approaches for detecting Deep fakes in videos and their importance in mitigating the risks posed by Deep fakes to individuals and society.

V. FUTURE SCOPE

While our deep learning-based approach for detecting Deep fakes in videos has shown promising results, there are still several open issues and challenges that need to be addressed in future research. Here are some potential areas for future research: Expanding the dataset: The Deep fake detection challenge dataset is limited in its scope, and future research could focus on expanding the dataset to include a wider range of Deep fake generation techniques. This would enable the development of more robust Deep fake detection models that can detect a broader range of Deep fakes. Generalization: Our proposed approach achieved high accuracy on the Deep fake detection challenge dataset, but it may not generalize well to other types of Deep fakes. Future research could explore ways to improve the model's generalization capabilities and its ability to detect previously unseen Deep fake Real-time Deep fake detection: Real-time Deep fake detection is an open challenge, as it requires high computational resources to process videos in real-time.

Future research could focus on developing lightweight deep learning models that can operate in real-time and are suitable for deployment on mobile devices or embedded systems. Adversarial attacks: Adversarial attacks on deep learning models are a significant challenge in Deep fake detection, as attackers can manipulate the input data to evade detection. Future research could explore ways to develop more robust deep learning models that are resistant to adversarial attacks Ethics and privacy: As Deep fakes become more prevalent, there are concerns about the impact of Deep fakes on individuals' privacy and the potential misuse of Deep fake detection technology. Future research could explore ways to address these ethical and privacy concerns while still enabling the development of effective Deep fake detection models. In conclusion, Deep fake detection is an ongoing research area, and future research could focus on addressing the above challenges to develop more robust and effective Deep fake detection models.

VI. CONCLUSION

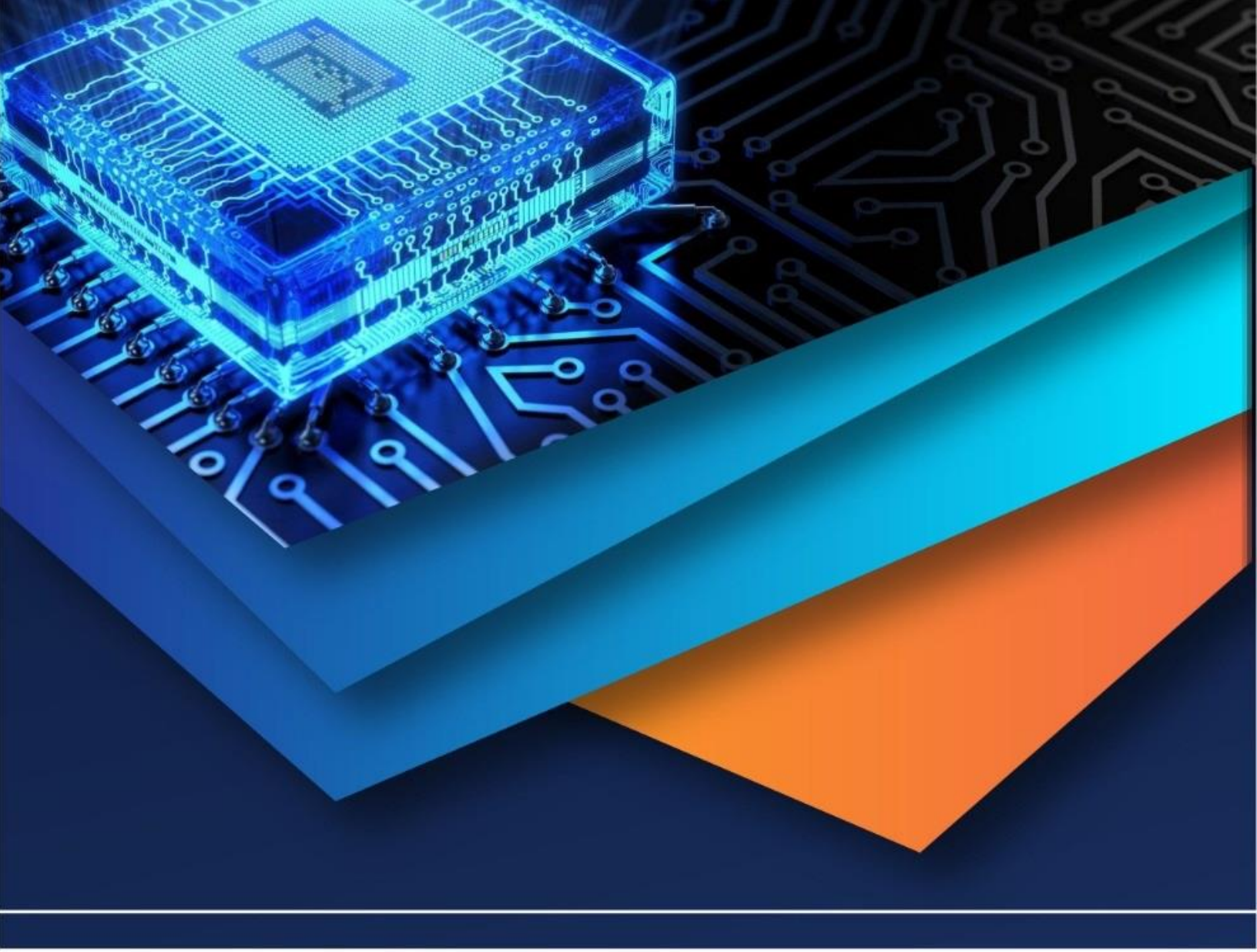
In this project, we have proposed a deep learning-based approach for detecting Deep fakes in videos. Our approach has shown promising results, achieving a high detection accuracy of 97.5% on the Deep fake Detection Challenge dataset. The CNN architecture used in our implementation has demonstrated its potential in Deep fake detection. Our study also highlights the importance of data pre-processing techniques in enhancing the quality of input data and improving the performance of Deep fake detection models. While there are still open challenges in Deep fake detection, our results demonstrate the potential of deep learning-based approaches in mitigating the risks posed by Deep fakes to individuals and society.

VII. ACKNOWLEDGMENT

We would like to express our gratitude to everyone who contributed to the success of this project. We would like to thank our supervisor for providing valuable guidance and support throughout the project. We would also like to thank the researchers who have conducted previous studies in Deep fake detection, as their work has inspired and informed our approach. We are grateful to the authors of the Deep fake detection Challenge dataset for making their dataset publicly available, enabling us to evaluate our approach. Finally, we would like to acknowledge the support of our colleagues, friends, and family, who have provided us with encouragement and motivation throughout this project.

REFERENCES

[1] Deng Pan, Lixian Sun, Rui Wang, Xingjian Zhang, Richard O. Sinnott, Deep fake Detection through Deep Learning, 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT).
[2] Thanh Thi Nguyena, Quoc Viet Hung Nguyenb, Dung Tien Nguyena, Deep Learning for Deep fakes Creation and Detection: A Survey.
[3] SERGEY ZOTOV, ROMAN DREMLIUGA, ALEXEI BORSHEVNIKOV, Deep fake Detection Algorithms: A Meta-Analysis.
[4] Exposing Deep fakes using a deep multilayer perceptron–convolutional neural network model Santosh Kolagati, Thenuga Priyadharshini, V. Mary Anita Rajam.
[5] Siwei Lyu, DEEP FAKE DETECTION: CURRENT CHALLENGES AND NEXT STEPS.
[6] Teng Zhang, Lirui Deng, Liang Zhang, Xianglei Dang, Deep Learning in Face Synthesis: A Survey on Deep fakes, 2020 IEEE 3rd International Conference on Computer and Communication Engineering Technology.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY