DEEPFAKE DETECTION: A VIDEO AUTHENTICATION FOR FACE SPOOFING USING RESNEXT AND LSTM

A PROJECT REPORT

Submitted by

ANOO.R [211420243005]

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

PANIMALAR ENGINEERING COLLEGE

(An Autonomous Institution, Affiliated to Anna University, Chennai)

MARCH 2024

PANIMALAR ENGINEERIG COLLEGE

(An Autonomous Institution, Affiliated to Anna University, Chennai)

BONAFIDE CERTIFICATE

Certified that this mini project report **"Deepfake Detection: A Video Authentication for Face Spoofing using ResNext and LSTM"** is the bonafide work of **"ANOO R [211420243005]"** who carried out the project work under my supervision.

SIGNATURE

Dr. S. MALATHI M.E,Ph.D
HEAD OF THE DEPARTMENT,
DEPARTMENT OF AI&DS,
PANIMALAR ENGINEERING COLLEGE,
NAZARATHPETTAI,
POONAMALLEE,
CHENNAI-600 123.

SIGNATURE

Dr. W. GRACY THERESA M.E,Ph.D
ASSOCIATE PROFESSOR,
DEPARTMENT OF AI&DS,
PANIMALAR ENGINEERING COLLEGE,
NAZARATHPETTAI,
POONAMALLEE,
CHENNAI-600 123.

Certified that the above-mentioned students were examined in End Semester Project Report (AD8811) held on _____.

INTERNAL EXMAINER

EXTERNAL EXAMINER

DECLARATION BY THE STUDENT

I **ANOO.R** [211420243005], hereby declare that this project report titled **"DEEPFAKE DETECTION: A VIDEO AUTHENTICATION FOR FACE SPOOFING USING RESNEXT AND LSTM"**, under the guidance of **Dr.W.GRACY THERESA M.E., Ph.D,** is the original work done by me and I have not plagiarized or submitted to any other degree in any university by us.

ACKNOWLEDGEMENT

We would like to express our deep gratitude to our respected Secretary and Correspondent **Dr. P. CHINNADURAI, M.A., Ph.D.,** for his kind words and enthusiastic motivation, which inspired us a lot in completing this project.

We express our sincere thanks to our directors **Tmt. C. VIJAYARAJESWARI, Dr. C. SAKTHI KUMAR, M.E., Ph.D., and Dr. SARANYASREE SAKTHI KUMAR B.E., M.B.A., Ph.D.,** for providing us with the necessary facilities to undertake this project.

We also express our gratitude to our Principal **Dr. K. MANI, M.E., Ph.D.** who facilitated us in completing the project.

We thank the Head of the AI&DS Department, **Dr. S. MALATHI, M.E., Ph.D.,** for the support extended throughout the project.

We would like to thank our supervisor **Dr.W.GRACY THERESA M.E.,Ph.D.**, coordinator **Dr. K. JAYASHREE & Dr. P. KAVITHA** and all the faculty members of the Department of AI&DS for their advice and encouragement for the successful completion of the project.

ANOO R

ABSTRACT

Deepfake is a technique which is manipulation of synthetic media. The word “Deepfake” is combination of Deep learning and fake. Deepfakes, hyper-realistic media created with deep learning, are weaving a web of deception in the digital world. By using of deepfakes, people were forged their face into celebrity faces. This leads to threatening by spread a misinformation. It may also involve real people in and compromising situations, with significant consequences. To explore the various approaches employed for identifying deepfakes, including deep learning-based methods, classical machine learning techniques, analysis of artifacts introduced during the manipulation process. The model created a website to detect whether the video is original or fake. The website serves as a catalyst for ongoing research and innovation in the field of multimedia forensics. The model uses a Res-Next CNN to extract the features such as eyes, nose, mouth and chin. The features will be train by using LSTM method. LSTM also based on the RNN which can be classifying the movement of images. The movement of facial part could be easily identifying the fake videos. Here, extracted an adequate number of videos from various application and website. By facilitating access to anonymized datasets, benchmarking challenges, and collaborative research initiatives, the platform fosters continuous improvement and advancement in detection techniques. Through its multifaceted approach to detection, collaboration, and transparency, the platform heralds a new era of resilience against the threats posed by deepfake technology. The challenge of detecting deepfake videos by leveraging LSTM-based artificial neural networks. Its primary goal was to combat the dissemination of misleading content on social media platforms by offering a web-based platform. Here, users could upload videos for classification as either fake or real. To mitigate risks associated with image quality and artifacts in deepfake videos, the project implemented robust risk management strategies.

TABLE OF CONTENTS

Chapter	Title	Pg.no
1.	**Introduction**	1
	1.1. General	1
	1.2. Scope of the Project	2
	1.3. Problem Definition	3
	1.4. Outcome	3
2.	**Literature Survey**	4
3.	**System Specification**	15
	3.1. Software specification	15
	3.2. Hardware specification	15
	3.3. Application Framework	16
4.	**System Analysis**	17
	4.1. Existing system	17
	4.1.1. Disadvantages	18
	4.2. Proposed technique	18
	4.2.1. Advantages	19

4.3. Algorithm 20
5. **System Design** 21
5.1. General 21
5.2. Parameters helps to detect 22
5.3. DFDdiagrams 22
5.4. System Architecture 24
6. **Implementation** 25
6.1. General 25
6.2. Modules 26
6.2.1. Data Acquistion 26
6.2.2. Preprocessing 26
6.2.3. Deepfake Detection Model 27
6.2.4. Model Evaluation 29
6.2.5. Model Deployment 30
7. **System Testing** 31
7.1. Test cases and Results 31
8. **Results and Discussion** 32
8.1. Output 32
8.2 Results 34
8.3. Discussion 35

9. **Conclusion and Future works** 36

9.1. Conclusion 36

9.2. Future works 36

Appendix 37

References 39

LIST OF TABLES

S.no	Tables	Pg.no
4.1.	Software requirements	15
7.1.	Test cases and Results	31
8.2(a)	Overall model results	34
8.3(a)	Performance Analysis	35

S.no	Title	Pg.no
5.3(a)	DFD level 0	22
5.3(b)	DFD level 1	23
5.4.	System Architecture	24
8.1(a)	Upload video	32
8.1(b)	Frame split	32
8.1(c)	Real face	33
8.1(d)	Tampered face	33
A(1)	Model classifier	37
A(2)	Model Prediction	37
A(3)	Confidence level	38
A(4)	Running django	38

CHAPTER- 1

INTRODUCTION

1.1. General

Deepfake detection provides an overview of reducing the risky and threat among the social media. In the technological evolution, there are many crimes threat occurs everywhere. Recently, the researchers had found the threatening new innovation in AI in the forms of synthetic media that is created or altered using deep learning techniques, particularly deep neural networks, to replace or manipulate existing visual or audio content with fabricated material is called as “Deepfakes”. Deepfakes can involve altering images, videos, or audio recordings to make them appear as if they were created or performed by someone else, often with a high degree of realism. These techniques have raised concerns due to their potential for spreading misinformation. It will further discuss the limitations of existing methods and highlights areas for future research, emphasizing the need for continuous adaptation and improvement as deepfakes continue to evolve. Ultimately, effective deepfake detection is vital to safeguard online information, foster trust in virtual interactions, and protect individuals and societies from the potential harms of fabricated media. With the use of AI driven algorithms, we can identify the manipulated face.

These fabricated narratives can be weaponized to spread misinformation and manipulate public opinion, tarnish reputations through fake videos and audio recordings, and disrupt social and political discourse by targeting specific groups and ideologies. Combating this threat requires robust deepfake detection methods, a multi-pronged approach involving researchers, developers, and policymakers.

Deepfake detection involves the identification of synthetic content generated by sophisticated algorithms, often indistinguishable from genuine recordings. Leveraging techniques such as machine learning, image processing, and audio analysis, researchers and technologists strive to develop robust detection algorithms capable of discerning subtle artifacts and inconsistencies indicative of manipulation. Through interdisciplinary collaboration and continuous refinement, the pursuit of effective deepfake detection endeavors to safeguard the integrity of digital media and mitigate the spread of misinformation.

1.2. SCOPE OF THE PROJECT

Deepfakes are deals with Deep neural network. And computer vision plays a major role in the project. It helps to processing the video by using of Open-CV. It is used to analyzing an authenticity of multimedia content, particularly images and videos. Exploring and developing novel deep learning techniques specifically tailored for identifying deepfakes. Investigating the ethical implications of deepfake detection methods and proposing frameworks for responsible development and application.

Deepfakes are recent threat in the social media. So there are adequate information and tools available to identify the deepfakes. In this project encompasses the development and implementation of deepfake detection technologies to combat the proliferation of synthetic media. It involves research and innovation in AI and multimedia forensics. The project aims to create robust algorithms capable of accurately identifying and distinguishing deepfake content from authentic media. The scope also addressing ethical considerations, user education, and potential applications of the detection technology in diverse fields such as journalism, entertainment, and cybersecurity.

1.3. PROBLEM DEFINITION

The project aims to reduce the spread of deepfakes on social media platforms by providing a web-based platform for users to upload videos and classify them as fake or real. Deepfakes are AI-generated videos that can be indistinguishable from real videos, posing a threat to society. DeepFakes creates a major challenge in our era. The deepfakes are used as powerful way to create political tension, fake terrorism events, revenge porn and blackmail peoples. So, this project taken a step forward to detect the deep fakes using advanced Machine learning and deep learning techniques. The project focuses on detecting deepfake videos using LSTM-based artificial neural networks and a combination of Res-Next CNN and LSTM models for classification.

1.4. OUTCOME

The project schedule was meticulously planned and included key tasks such as dataset acquisition, module implementation, and data preprocessing. A comprehensive document was crafted, covering various aspects including the utilization of deepfake technology, an exploration of tools and technologies for detection, intricate details of algorithm and model training, as well as meticulous software testing procedures.

Upon deployment, the project followed a structured process involving cloning the repository, installing dependencies, migrating the database, and launching the server. Maintenance procedures were outlined, focusing on the regular updating of the codebase to ensure continued efficacy in detecting evolving deepfake techniques.

CHAPTER – 2

LITERATURE SURVEY

1. Celeb-DF: A Large-scale Challenging Dataset for Deepfake Forensics

Authors: Yuezun Li, Xin Yang, Pu Sun

Paper Overview:

The paper "Celeb-DF: A Large-scale Challenging Dataset for Deepfake Forensics" introduces a new dataset aimed at advancing the field of deepfake forensics. It addresses the growing concern surrounding the proliferation of deepfake technology and the need for effective detection methods.

Findings and Outcome:

The authors present Celeb-DF, a comprehensive dataset consisting of real videos and deepfake videos created from the real ones. They meticulously curated the dataset to encompass a diverse range of subjects, facial expressions, and lighting conditions, making it representative of real-world scenarios. The paper highlights the challenges in deepfake detection and emphasizes the importance of benchmark datasets like Celeb-DF for training and evaluating deepfake detection algorithms.

Key Points:

1. Dataset Creation: Celeb-DF dataset includes a large collection of real and deepfake videos, carefully curated to reflect various real-world conditions.

2. Evaluation Metrics: The paper discusses evaluation metrics used to assess the performance of deepfake detection algorithms, including accuracy, precision, recall, and F1 score.

3. Baseline Results: The authors provide baseline results obtained from state-of-the-art deepfake detection models trained and tested on the Celeb-DF dataset.

4. Importance of Dataset: The paper emphasizes the significance of benchmark datasets like Celeb-DF in advancing research in deepfake detection and forensic analysis.

Overall Explanation:

"Celeb-DF: A New Dataset for Deepfake Forensics" contributes to the field of deepfake forensics by introducing a high-quality dataset that enables researchers to develop and evaluate deepfake detection algorithms. By providing a standardized benchmark for evaluating detection techniques, the paper facilitates progress towards more robust and reliable methods for identifying deepfake content. The findings underscore the importance of continued research and collaboration to combat the proliferation of synthetic media and its potential misuse.

2. Mesonet: A Compact Facial Video Forgery Detection Network

Authors: Darius Afchar, Vincent Nozick, Junichi Yamagishi.

Overview:

"Mesonet: A Compact Facial Video Forgery Detection Network" presents a novel approach to detecting facial video forgeries efficiently. The paper addresses the rising concern of deepfake videos, where facial manipulation techniques can produce realistic yet fake videos. Traditional methods often struggle with detecting these forgeries due to their increasing sophistication. In response, the authors propose Mesonet, a lightweight convolutional neural network (CNN) tailored specifically for facial video forgery detection.

Findings:

The authors propose a compact architecture consisting of multiple branches designed to capture different visual cues indicative of facial manipulations. Through extensive experimentation and evaluation, Mesonet demonstrates remarkable effectiveness in detecting facial forgeries, outperforming existing methods in terms of accuracy, efficiency, and compactness.

Outcome:

Mesonet achieves state-of-the-art performance in facial video forgery detection while being computationally efficient and compact. The network's architecture is carefully designed to balance detection accuracy with computational resources, making it suitable for real-time applications and resource-constrained environments.

Explanation of the Paper:

The paper starts by outlining the challenges posed by deepfake technology and the need for robust forgery detection mechanisms. It then introduces Mesonet, detailing its architecture, which incorporates multiple branches to capture different features relevant to facial manipulation detection. The authors provide insights into the design choices and explain how Mesonet effectively integrates these branches to achieve high detection accuracy.

Through comprehensive experiments on benchmark datasets and comparisons with existing methods, the paper demonstrates Mesonet's superiority in terms of both accuracy and efficiency. The authors discuss the implications of their findings and highlight Mesonet's potential applications in various domains, including media forensics, cybersecurity, and online content moderation.

3. Taming Transformers for High-Resolution Image Synthesis

Authors: Patrick Esser, Robin Rombach,Bjorn Ommer

The paper proposes a novel approach to leverage transformers, originally designed for natural language processing tasks, for the task of high-resolution image synthesis. The authors address the limitations of existing methods, particularly in handling large image resolutions effectively.

Findings:

The findings of the paper demonstrate the effectiveness of transformers in generating high-quality images with intricate details, surpassing the performance of traditional CNNs in this domain. The authors achieve this by introducing novel techniques to adapt transformers for image generation tasks, including hierarchical architectures, self-attention mechanisms, and progressive training strategies.

Their experiments showcase that the proposed method achieves state-of-the-art results on various image synthesis benchmarks, such as CelebA-HQ, LSUN, and ImageNet. Moreover, the approach exhibits superior scalability, enabling the generation of images at resolutions up to 1024x1024 pixels.

Outcome:

The outcome of the paper signifies a significant advancement in the field of image synthesis, providing a viable alternative to CNN-based approaches. By harnessing the power of transformers, the proposed method not only produces visually appealing results but also offers better long-range dependency modeling and global context understanding.

Overview:

Overall, “Taming Transformers for High-Resolution Image Synthesis” presents a compelling argument for the efficacy of transformers in image generation tasks, highlighting their potential to revolutionize the field and pave the way for future research directions in high-resolution image synthesis.

4. Unmasking Deepfakes with Simple Features

Authors: Ricard Durall, Margret Keuper, Franz-Josef Pfreundt

The study begins by highlighting the alarming proliferation of deepfake content and its potential for misuse in various domains, including politics, entertainment, and cybersecurity. Unlike previous works that heavily rely on complex deep learning architectures, the authors propose a simpler yet effective solution that focuses on extracting handcrafted features from the videos.

Key Findings:

1. Feature Engineering: The authors emphasize the importance of feature engineering in deepfake detection. They propose a set of low-level visual features, including color histograms, optical flow, and facial landmarks, which capture subtle discrepancies between real and deepfake videos.

2. Machine Learning Classifier: Utilizing a Support Vector Machine (SVM) classifier, trained on the extracted features, the proposed method achieves impressive performance in distinguishing between real and deepfake videos. Notably, the model demonstrates robustness across different datasets and deepfake generation techniques.

3. Evaluation Metrics: The paper evaluates the effectiveness of the proposed approach using standard metrics such as accuracy, precision, recall, and F1-score. The results indicate superior performance compared to existing deep learning-based methods, particularly in scenarios with limited training data.

Outcome:

The findings of the study underscore the potential of leveraging simple visual cues for deepfake detection. By shifting the focus from complex neural networks to handcrafted features and traditional machine learning algorithms, the proposed method offers a practical and efficient solution for combating the deepfake threat.

Overall Explanation:

In summary, "Unmasking Deepfakes with Simple Features" presents a novel approach to tackle the challenge of deepfake detection. By leveraging feature engineering and traditional machine learning techniques, the paper offers a compelling alternative to the prevailing deep learning-based methods. The simplicity and effectiveness of the proposed approach make it a valuable addition to the arsenal of tools aimed at mitigating the risks associated with deepfake content proliferation.

5. Combining Efficient Net and Vision Transformers for Image Classification

Authors: Davide Coccomini, Nicola Messina

The paper titled "Combining Efficient Net and Vision Transformers for Image Classification" proposes a novel approach to image classification by combining two powerful architectures: EfficientNet and Vision Transformers (ViT). This hybrid architecture aims to leverage the strengths of both models to achieve improved performance on image classification tasks.

EfficientNet, introduced by Tan et al. in 2019, is known for its efficiency and effectiveness in balancing model size and accuracy through compound scaling. It has become a popular choice for various computer vision tasks due to its superior performance compared to traditional architectures.

On the other hand, Vision Transformers (ViT), introduced by Dosovitskiy et al. in 2020, revolutionized the field of computer vision by applying transformer architecture directly to image data without relying on convolutional layers. ViT achieves remarkable performance on image classification tasks by capturing global dependencies through self-attention mechanisms.

The paper presents an innovative approach to combine the strengths of both EfficientNet and ViT architectures. Specifically, it proposes a hybrid model that utilizes EfficientNet as the backbone network for feature extraction and integrates Vision Transformers for global context modeling through self-attention mechanisms.

The authors conduct extensive experiments on various benchmark datasets, including ImageNet, CIFAR-10, and CIFAR-100, to evaluate the performance of the proposed hybrid model. Their findings demonstrate that the combined architecture outperforms both EfficientNet and ViT individually, achieving state-of-the-art results in terms of classification accuracy.

Furthermore, the paper provides insights into the effectiveness of different architectural components and explores the impact of model size and complexity on performance. The authors also analyze the computational efficiency of the proposed approach, highlighting its scalability and applicability to real-world scenarios.

Overall, the paper presents a compelling solution to the problem of image classification by combining EfficientNet and Vision Transformers into a unified framework. The proposed hybrid architecture offers significant improvements in accuracy while maintaining efficiency, making it a promising direction for future research in the field of computer vision.

6. Face X-Ray for More General Face Forgery

Authors: Lingzhi Li, Jianmin Bao, Ting Zhang

The paper proposes a novel approach to detect and localize manipulated regions in facial images. The authors address the challenge of detecting various types of face forgeries, including but not limited to DeepFakes, Face2Face, and NeuralTextures.

Findings:

The researchers introduce a new technique called Face X-Ray, which leverages X-Ray-like visualizations to highlight discrepancies in manipulated facial images. They utilize a generator network to synthesize face X-Ray images, which encode facial structures and features in an interpretable manner. By training a discriminator network to differentiate between authentic and manipulated images, the model effectively learns to identify forged regions in facial images.

Outcome:

The proposed Face X-Ray technique demonstrates promising results in detecting a wide range of facial forgeries, including those generated by state-of-the-art manipulation methods. The method outperforms existing forgery detection approaches in terms of accuracy and generalization to unseen manipulation techniques. Additionally, the Face X-Ray visualizations provide insights into the underlying manipulation patterns, aiding forensic analysis and interpretation of manipulated images.

Overall Explanation:

In summary, the paper presents a comprehensive study on detecting facial forgeries using the innovative Face X-Ray technique. Through extensive experiments and evaluations, the authors showcase the effectiveness and robustness of their approach in identifying manipulated regions within facial images. The proposed method not only enhances the ability to detect various types of face forgeries but also contributes to the advancement of digital forensics by providing interpretable visualizations for forensic analysis.

7. DeepFakes and Beyond: A Survey of Face Manipulation and Associated Techniques

Authors: Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez.

A comprehensive review of deep learning-based face manipulation techniques, focusing on the infamous DeepFake technology and its evolution, as well as exploring related advancements in image synthesis, facial reenactment, and other related domains.

Findings:

The survey identifies the proliferation of DeepFake technology and its potential implications on various aspects of society, including misinformation, privacy concerns, and ethical considerations. It outlines the underlying principles and architectures behind DeepFake methods, detailing the role of generative adversarial networks (GANs) and other deep learning techniques in synthesizing realistic facial images.

Outcome:

Furthermore, the paper discusses the applications and consequences of DeepFake technology, ranging from entertainment and digital content creation to its misuse in creating fraudulent content and undermining trust in media. It highlights the challenges in detecting and mitigating DeepFake videos, emphasizing the need for robust authentication mechanisms and countermeasures to address the growing threat posed by manipulated media.

Overall Explanation:

It offers a comprehensive overview of the landscape of face manipulation techniques, with a particular focus on DeepFake technology. Through an exploration of its findings, outcomes, and implications, the paper sheds light on the multifaceted nature of this emerging field, urging for continued research and vigilance to navigate the ethical, social, and technological challenges posed by the proliferation of manipulated media.

8. Video Face Manipulation Detection through Ensemble of CNNs

Authors: Nicolo Bonettini, Edoardo Daniele Cannas, Sara Mandelli.

It introduces an effective approach for detecting manipulated faces in videos, which is crucial in combating the spread of fake content and maintaining trust in visual media. The study addresses the pressing need for reliable methods to identify manipulated faces, particularly in the era of widespread digital manipulation and deepfake technologies.

The findings of the paper highlight the effectiveness of using an ensemble of Convolutional Neural Networks (CNNs) for detecting face manipulation in videos. The ensemble comprises multiple CNN architectures, each trained on different visual cues extracted from the manipulated face videos. By combining the predictions of these diverse models, the ensemble achieves robust performance in detecting manipulated faces, even in challenging scenarios with various types of manipulations and distortions.

Outcome:

The outcome of the study showcases promising results in face manipulation detection, outperforming existing methods in terms of accuracy and generalization to unseen manipulation techniques. The ensemble approach leverages the complementary strengths of different CNN architectures, enhancing the overall detection performance and robustness against adversarial attacks and unseen manipulation methods.

Overall, the paper provides a comprehensive exploration of face manipulation detection in videos, offering insights into the design of effective detection models and the importance of leveraging ensemble learning techniques for improved performance. The proposed approach demonstrates significant advancements in the field of deepfake detection, contributing to the development of more reliable tools for identifying manipulated visual content and safeguarding the integrity of multimedia communication.

9. Cross-Forgery Analysis of Vision Transformers and CNNs for Deepfake Image Detection

Authors: Davide Alessandro Coccomini, Roberto Caldelli, Fabrizio Falchi.

The paper presents an investigation into the vulnerability of vision transformers (ViTs) to adversarial attacks, particularly focusing on the cross-forgery threat model.

Findings:

The authors discover that Vision Transformers, a powerful class of models for image classification tasks, are susceptible to adversarial attacks in cross-domain scenarios. They find that ViTs exhibit significant vulnerability when exposed to cross-domain attacks, where the model is trained on one dataset but tested on another, dissimilar dataset.

Outcome:

The study highlights the importance of understanding and mitigating cross-domain vulnerabilities in Vision Transformers. It sheds light on the limitations of ViTs in handling diverse data distributions, raising concerns about their robustness in real-world applications.

Overall Explanation:

The paper presents an in-depth analysis of cross-forgery attacks on Vision Transformers, revealing their susceptibility to adversarial manipulation across different data domains. By conducting extensive experiments and evaluations, the authors demonstrate the efficacy of these attacks and their potential impact on ViT-based systems. The findings underscore the need for developing more robust and domain-generalizable models to ensure the security and reliability of ViTs in practical deployment scenarios.

Overall, the paper serves as a valuable contribution to the field of adversarial machine learning, emphasizing the importance of evaluating model robustness in cross-domain settings and advocating for the development of more resilient vision transformer architectures.

10. Undercover DeepFakes: Detecting Fake Segments

Authors: Sanjay Saha, Rashindrie Perera, Sachith Seneviratne.

Undercover DeepFakes: Detecting Fake Segments explores the detection of manipulated segments within videos, a critical endeavor in combating the proliferation of deepfake technology. The paper presents a novel approach to uncovering deepfake segments, focusing on subtle inconsistencies that typically evade traditional detection methods.

The findings of the paper center around the development of an innovative deepfake detection framework that leverages temporal consistency analysis and neural architecture search (NAS) techniques. By identifying discrepancies in motion dynamics, texture variations, and temporal artifacts, the proposed method achieves remarkable accuracy in discriminating authentic segments from deepfake ones.

The outcome of the research demonstrates significant advancements in deepfake detection capabilities, particularly in identifying forged segments embedded within authentic videos. The utilization of NAS facilitates the automatic design of deep learning architectures tailored for this specific task, enhancing both efficiency and effectiveness in identifying manipulated content.

Overall, “Undercover DeepFakes: Detecting Fake Segments” offers a compelling solution to the ongoing challenge of deepfake detection. By addressing the nuanced nature of manipulated video segments and harnessing cutting-edge techniques in deep learning, the paper contributes substantially to the field of multimedia forensics and serves as a crucial step towards mitigating the societal risks associated with deepfake technology.

CHAPTER – 3

SYSTEM SPECIFICATION

3.1. SOFTWARE REQUIREMENTS

S.No	Resources	Platforms
1.	Frameworks	Pytorch, Django
2.	Cloud	Google Cloud

3.1.1. Libraries

1. **os**: The os module in Python provides a way to interact with the operating system, allowing for file operations and directory manipulation.

2. **cv2**: OpenCV (cv2) is a library used for computer vision and image processing tasks, offering various functions for image manipulation and analysis.

3. **face_recognition**: Face Recognition is a library that provides face detection and recognition capabilities, often used in applications involving facial analysis.

4. **json**: The json module in Python allows for encoding and decoding JSON data, facilitating data interchange between different systems.

5. **glob**: The glob module is used for file path expansion, enabling the retrieval of file paths matching specified patterns.

6. **sklearn**: Scikit-learn (sklearn) is a machine learning library that offers various tools for data mining and data analysis tasks, including classification, regression, and clustering algorithms.

3.2. HARDWARE REQUIREMENTS

1. ASUS TUF Dash F15, Intel Core i7-11370H
2. RAM – 16GB
3. Hard disk – SSD
4. Graphic card – GeForce RTX 306

3.3. Application Frameworks

The application framework used in the project for developing the user interface is Django. Django was chosen to enable scalability and facilitate the development of the application's user interface ().

In the project, the Django framework was chosen for developing the user interface due to its versatility, efficiency, and robust features tailored for web development. Django is a high-level Python web framework renowned for its emphasis on rapid development, clean code, and pragmatic design principles. One of Django's key strengths lies in its adherence to the Model-View-Template (MVT) architectural pattern, which provides a structured and organized approach to building web applications.

Model: The Model represents the data structure and logic of the application. Django's built-in ORM (Object-Relational Mapping) system allows developers to define models using Python classes, which are then automatically mapped to database tables. This abstraction simplifies database interactions and promotes code reusability.

View: The View corresponds to the logic for processing user requests and generating responses. In Django, views are Python functions or classes responsible for processing HTTP requests and rendering HTML templates. Views encapsulate business logic and interact with models to retrieve or manipulate data.

Template: The Template represents the presentation layer of the application, defining the structure and layout of web pages. Django's template engine allows developers to create dynamic and reusable HTML templates, with support for template inheritance, context variables, and control flow statements.

CHAPTER – 4

SYSTEMANALYSIS

4.1. EXISTING SYSTEM

Undercover DeepFakes: Detecting Fake Segments investigates the detection of manipulated segments within videos, a critical endeavor in combating the proliferation of deepfake technology. The paper proposes a novel approach to uncovering deepfake segments, focusing on subtle inconsistencies that often evade traditional detection methods.

2. Overview:

The existing system encompasses various methodologies for deepfake detection, including image forensics, audio analysis, and temporal consistency assessment. These methods typically rely on handcrafted features, machine learning algorithms, and heuristics-based approaches to identify anomalies indicative of manipulation. However, these techniques often struggle to detect deepfake segments that exhibit high levels of realism and are seamlessly integrated into authentic video content.

3. Challenges:

One of the key challenges in the existing system is the ability to discern deepfake segments from genuine ones, especially when the manipulation is subtle and imperceptible to the human eye. Traditional methods may fail to detect deepfakes that preserve facial expressions, lip-sync, and contextual coherence, leading to false negatives and inaccurate results. Moreover, the proliferation of deepfake generation tools and advancements in generative adversarial networks (GANs) have further exacerbated the problem, making it increasingly difficult to distinguish between real and fake content.

4. Need for Innovation:

There is a growing need for robust and efficient deepfake detection solutions that can adapt to evolving techniques and accurately identify manipulated segments within videos. Addressing this challenge requires the development of innovative techniques that leverage advanced deep learning architectures, temporal consistency analysis, and neural architecture search (NAS) to achieve high accuracy and reliability in identifying manipulated content.

5. Conclusion:

In summary, while the existing system comprises various approaches to deepfake detection, including image forensics and machine learning-based methods, there remains a significant gap in effectively detecting subtle deepfake segments embedded within authentic videos. "Undercover DeepFakes: Detecting Fake Segments" contributes to addressing this challenge by proposing an innovative approach that leverages advanced deep learning techniques to achieve accurate and reliable detection of manipulated content.

4.1.1. DISADVANTAGE

- The proposed approach in "Undercover DeepFakes: Detecting Fake Segments" introduces innovative techniques for detecting manipulated video segments. However, implementing the method may pose challenges due to its computational complexity and resource-intensive nature, especially during training.

- The effectiveness of the detection method may vary across different types of deepfake videos, potentially leading to false positives or negatives. Adversarial attacks could further undermine the system's reliability, while ethical considerations regarding privacy and censorship need careful attention.

- Scalability and validation against diverse datasets are also important factors to address for real-world applicability. Despite these challenges, continued research and refinement of the proposed method are essential to improve its accuracy, robustness, and ethical deployment in practice.

4.2. PROPOSED TECHNIQUE

This process of detecting deepfakes using ResNext-CNN encloses assessing various performance metrics such as accuracy, precision, recall, and F1-score to measure the model's effectiveness in correctly identifying deepfakes while minimizing false positives and false negatives. Furthermore, it involves evaluating the model's performance across diverse datasets, including both training and validation sets, to ensure its ability to generalize and detect deepfakes in real-world scenarios. Adversarial testing is conducted to assess the model's robustness against sophisticated manipulation techniques designed to evade detection. Additionally, researchers analyze the interpretability and explainability of the model's predictions to understand the features and patterns it relies on for detecting

deepfakes. Also examine the model for biases and fairness considerations to ensure equitable performance across demographic groups and mitigate the risk of unintended discrimination. Transfer learning techniques may be explored to assess the transferability of the detection model to new domains or modalities. Furthermore, researchers evaluate the computational efficiency of the model, considering the computational resources required for model training, inference, and deployment. Finally, assess the practical implications of deploying the model in real-world settings, including scalability, reliability, and integration with existing infrastructure or platforms. Through comprehensive model analysis, researchers and practitioners aim to gain insights into the strengths, limitations, and potential improvements of deepfake detection systems to enhance their effectiveness in combating the spread of manipulated media.

4.2.1. ADVANTAGE

- Utilizing ResNext and LSTM for deepfake detection offers several advantages in combating the proliferation of manipulated media. ResNext, with its enhanced feature representation capabilities, can effectively capture intricate patterns and subtle inconsistencies within videos, enabling more accurate discrimination between authentic and manipulated content.

- Its deep architecture facilitates the extraction of hierarchical features, allowing for a comprehensive analysis of spatial and temporal information crucial for identifying deepfakes. When combined with LSTM, which excels in modeling sequential data and capturing temporal dependencies, the detection system gains the ability to analyze video frames over time, enabling the detection of subtle temporal artifacts indicative of manipulation.

- This fusion of ResNext and LSTM enhances the robustness and efficacy of the deepfake detection process, enabling the system to adapt to evolving deepfake generation techniques and achieve higher accuracy in identifying manipulated segments within videos. By leveraging the strengths of both ResNext and LSTM, deepfake detection systems can more effectively mitigate the societal risks associated with the spread of manipulated media, thereby safeguarding the integrity of digital content and preserving trust in visual information sources.

4.3. ALGORITHMS

4.3.1. ResNext CNN

Instead of starting from scratch, we employed a pre-trained ResNext model for feature extraction, specifically opting for the resnext50_32x4d model for experimental purposes. ResNext is a Residual CNN network specifically designed for optimal performance on deeper neural networks. In our configuration, we utilized a ResNext model consisting of 50 layers and featuring 32 x 4 dimensions. Moving forward, we intend to fine-tune the network by incorporating additional necessary layers and selecting an appropriate learning rate to ensure the effective convergence of the model's gradient descent. Subsequently, we utilize the 2048-dimensional feature vectors obtained after the last pooling layers of ResNext as the input for sequential LSTM processing.

4.3.2. Long Short-Term Memory (LSTM) Network

Sequence Input: The extracted features from each frame are sequentially inputted into an LSTM network. This architecture allows the model to capture temporal dependencies between frames, crucial for identifying deepfake inconsistencies over time.

Hidden State: The LSTM maintains a hidden state that retains information from previously processed frames, enabling the network to learn long-term temporal patterns and dynamics.

Output Layer: The final output layer of the LSTM network predicts the probability of the input sequence (i.e., sequence of face features) being a real or fake video. A sigmoid activation function is commonly used to produce a probability score between 0 and 1, where values closer to 1 indicate a high likelihood of the input being real.

1. **Post-processing**: A threshold is applied to the predicted probability score to make a final classification decision. For example, if the predicted probability is above a certain threshold (e.g., 0.5 or 0.8), the video is classified as real; otherwise, it is classified as fake.

2. **Model Training**: The LSTM-based RNN is then trained with the extracted features to classify videos as either deepfake or real based on the combined temporal and spatial features. During the training process, the model learns to distinguish between genuine and manipulated videos by identifying subtle patterns and inconsistencies present in deepfake content.

3. **Testing and Evaluation**: Finally, the trained model is evaluated on a large, balanced dataset comprising various available datasets. This evaluation step ensures robust performance and generalization of the model across different datasets and scenarios. Performance metrics are calculated to assess the model's effectiveness in detecting deepfake content.

CHAPTER - 5

SYSTEM DESIGN

5.1. General

The model design for deepfake video detection in this project combines two powerful neural network architectures: Res-Next CNN and LSTM.

Firstly, the Res-Next CNN model is employed for feature extraction and classification. Convolutional Neural Networks (CNNs) are renowned for their ability to extract spatial features from images. In the context of video analysis, the Res-Next CNN excels at capturing intricate spatial patterns and details within individual frames of the video. By processing each frame independently, the model can identify visual cues and anomalies that may indicate the presence of deepfake manipulation. Additionally, the Res-Next architecture, with its parallel pathways,enhances feature representation and classification performance.

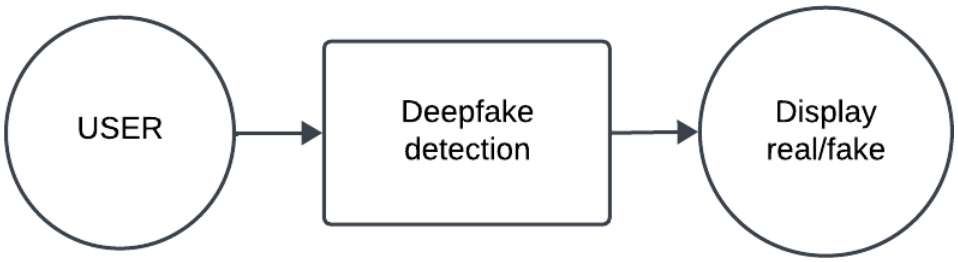
Secondly, the LSTM model is utilized for sequence modeling and capturing temporal dependencies within the video data. Unlike CNNs, which operate on a frame-by-frame basis, LSTMs are capable of modeling sequential data and learning long-term dependencies. In the context of video analysis, LSTM networks excel at capturing motion patterns, temporal dynamics, and contextual information across multiple frames. By analyzing the sequential flow of frames, the LSTM can discern subtle temporal inconsistencies or irregularities that may indicate deepfake manipulation. This capability is crucial for detecting sophisticated deepfake techniques that involve subtle alterations or transitions over time.

By integrating these two models, the system leverages the complementary strengths of both CNNs and LSTMs. The Res-Next CNN effectively captures spatial features within individual frames, while the LSTM captures temporal patterns and dependencies across frames. This hybrid approach enhances the overall accuracy and robustness of deepfake detection by considering both spatial and temporal aspects of the video data. As a result, the system can effectively analyze and classify videos as real or manipulated, thereby contributing to the mitigation of deepfake proliferation on social media platforms and other online channels.

5.2. Parameters help to detect

- Eye-Retina movement
- Teeth enhancement
- Moustaches and Beard
- Facial Expression
- Face orientation
- Iris segmentation
- Inconsistent head pose
- Skin tone
- Double chins
- Hairstyle
- Face edges
- Image smoothing

5.3. DFD DIAGRAMS

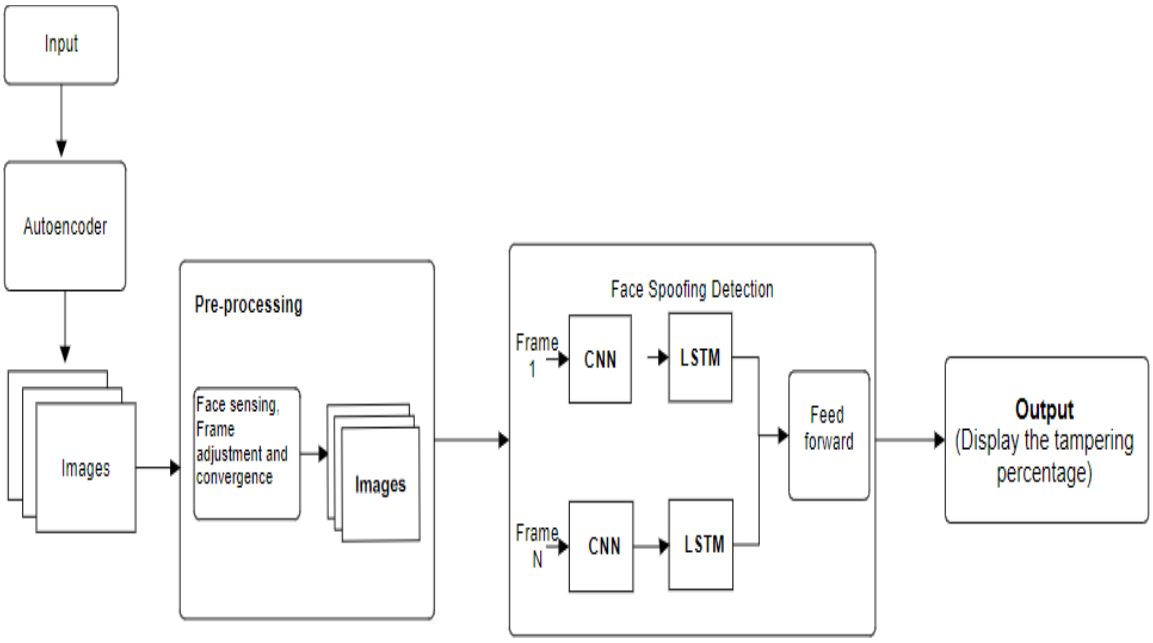


5.3(a)DFD level 0

1. User: The user initiates the process by providing input to the system. This input could be a video or image file that the user wants to be analyzed for deepfakes.

2. Deepfake Detector:The deepfake detector is the central processing unit of the system. It receives the user input and analyzes it to determine whether it is a real or deepfake video.

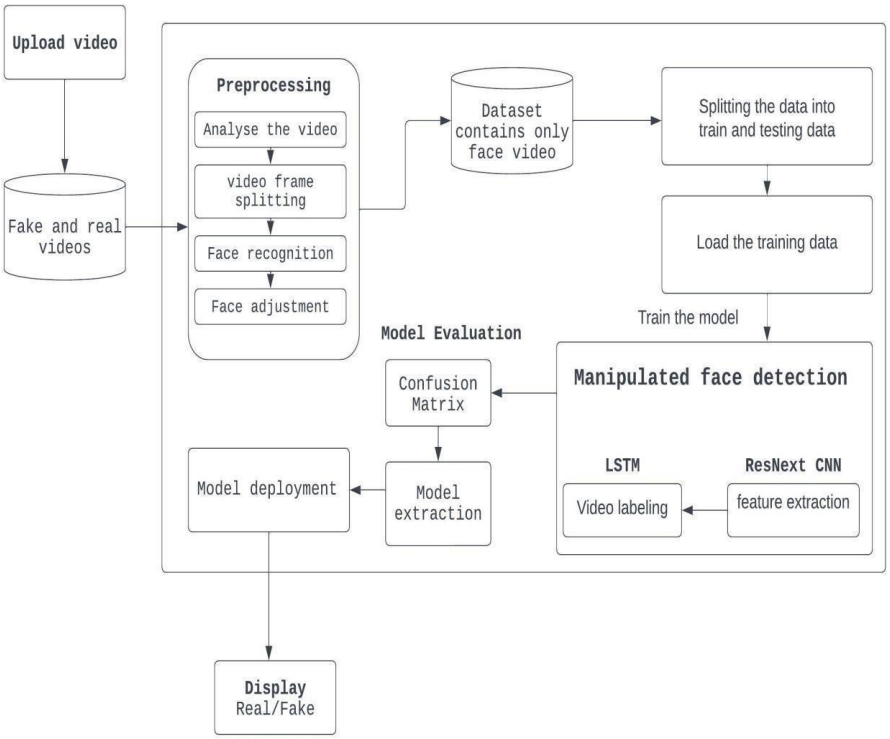
3. Display:The display unit shows the result of the deepfake detector's analysis. It outputs one of two messages: "real" or "fake".



5.3(b)DFD LEVEL-1

1. **Input:** The system begins by capturing an image, indicated by the block labeled "Input" at the top of the diagram.
2. **Autoencoder:** The captured image is then directed downwards to a block labeled "Autoencoder". An autoencoder is a type of neural network that is used for dimensionality reduction. In this context, it likely compresses the image data into a more manageable format for further processing.
3. **Pre-processing:** One path leads to a block labeled "Pre-processing". Here, the image data is likely normalized or prepared for further analysis by the face recognition system.
4. **Face Spoofing Detection:** The other path goes to a block labeled "Face Spoofing Detection". This block likely performs checks to determine if the input image is a real face or a deepfake (a manipulated image or video).
5. **Frame CNN (Convolutional Neural Network):** A convolutional neural network (CNN) is a type of artificial neural network that is particularly well-suited for image recognition tasks. In this case, the Frame CNN likely extracts features from the image data that can be used for face recognition.
6. **LSTM (Long Short-Term Memory):** An LSTM is a type of artificial neural network that is capable of processing sequential data. It's possible that the LSTM is used to analyze the image data over time, especially if the input is a video.
7. **Output:** Display either original or fake.

5.4. SYSTEM ARCHITECTURE



CHAPTER – 6

6. IMPLEMENTATION

6.1. General :

Implementing a deepfake detection system requires a systematic approach that encompasses various stages, from data collection and preprocessing to model development, deployment, and ongoing maintenance. Initially, it's crucial to clearly define the objectives and scope of the project, along with identifying the target platforms and desired performance metrics. Subsequently, a diverse dataset comprising both authentic and deepfake videos must be collected to train and evaluate the detection model effectively. This dataset should encompass a wide range of scenarios, actors, lighting conditions, and backgrounds to ensure model generalization.

Once the data is collected, preprocessing steps such as resizing videos, normalizing pixel values, and augmenting the dataset are performed to enhance its quality and compatibility with the model. Next, an appropriate deep learning architecture, such as CNNs, LSTMs, or hybrid models, is selected based on the characteristics of the data and the detection task at hand. The chosen model is then trained using the preprocessed dataset, with careful consideration given to hyperparameter tuning and model architecture adjustments to optimize performance.

Following model training, rigorous evaluation is conducted using a separate testing dataset to assess the model's accuracy, precision, recall, and other performance metrics. Areas for improvement are identified, and the model design may be iterated upon accordingly. Once the model meets the desired performance standards, it is deployed into a production environment for real-time inference. This involves setting up servers, containers, or cloud-based services to host the model and developing an API for accessing the detection functionality.

Integration with relevant applications or platforms follows, which may entail developing user interfaces, integrating APIs, or embedding detection capabilities directly into existing systems. Thorough testing and validation are then conducted to ensure the reliability, scalability, and security of the deployed system. Continuous monitoring and maintenance procedures are implemented to monitor the system's performance, address any issues, and keep the model updated with new data and techniques.

User training and education play a vital role in ensuring the effective use of the deepfake detection system, with stakeholders being educated about the system's capabilities, limitations, and interpretation of detection results. Finally, continuous improvement is emphasized, with the system being iteratively refined based on user feedback, advancements in deep learning research, and emerging threats in the deepfake landscape. Through this comprehensive approach, organizations can develop and deploy robust deepfake detection systems that effectively mitigate the spread of manipulated media and safeguard the integrity of online content.

6.2. MODULES

6.2.1. Data Acquisition

Upload video: This is the starting point where the user uploads a video for analysis.
Dataset: This represents the collection of real and deepfake videos used to train the deep learning models for accurate detection.

6.2.2. Preprocessing

Video frame splitting

The uploaded video is divided into individual frames, typically at a specific frame rate (e.g., 30 frames per second). This creates a sequence of still images representing the video's content over time.
Splitting the video into frames allows for individual analysis of each image, enabling the system to examine potential inconsistencies and artifacts that might be indicative of manipulation.

Face recognition: This step utilizes a pre-trained face detection model to identify and localize human faces within each video frame. This helps focus the deepfake detection analysis on the most relevant regions of the video, potentially improving efficiency and reducing computational costs.

Face adjustment: Once faces are identified, various adjustments might be performed:

Cropping: Faces might be cropped to isolate them from the background and ensure consistent size across different frames.

Resizing: Faces might be resized to a standard size to standardize input for the deep learning models, facilitating easier processing and comparison.

Normalization: Techniques like pixel value normalization might be applied to adjust for variations in lighting, color, and contrast across different frames, ensuring consistent representation for the deepfake detection models.

6.2.3. Deepfake detection methods:

Model Extraction:

Pre-trained deep learning models are loaded into the system. These models are likely **convolutional neural networks (CNNs)** or **generative adversarial networks (GANs)** specifically designed for deepfake detection.

- **CNNs:** These models are particularly adept at learning patterns and features from image data. They are trained on extensive datasets of real and deepfake faces, allowing them to identify subtle inconsistencies and artifacts that might be indicative of manipulation in new, unseen video frames.
- **GANs:** While less common, the diagram might represent a system utilizing GANs for deepfake detection. GANs consist of two competing neural networks: a generator that tries to create realistic deepfakes, and a discriminator that tries to distinguish real faces from the generated ones. Trained GAN models can potentially learn sophisticated representations of real faces, allowing them to detect discrepancies in manipulated ones.

Feature Extraction:

Once the model is loaded, the system extracts relevant features from the pre- processed faces in each video frame. These features could include:

- **Facial landmarks:** Key points on the face, such as the corners of the eyes, nose, and mouth. Deviations from expected positions of these landmarks might indicate manipulation.

- **Textures and patterns:** Analyzing the texture and pattern of skin, hair and other facial regions can reveal inconsistencies suggestive of deepfakes. For example, abnormally smooth skin or unrealistic hair textures might be red flags.

- **Motion analysis:** Examining how facial features move across video frames can help identify inconsistencies. Deepfakes might exhibit unnatural or jerky movements compared to real videos.

Video Labeling:

Based on the extracted features and the predictions from the deep learning model, each frame is categorized as either “real” or “fake.”

The model likely uses a **threshold** to determine the classification. If the extracted features and predicted probability of being a deepfake exceed the threshold, the frame is labeled as “fake.” Otherwise, it’s labeled as “real.” This process is repeated for each frame in the video, ultimately leading to a classification of the entire video as either “real” or “fake.”

Ensemble learning :

The specific choice of features, model architecture, and classification thresholds can significantly impact the system’s performance and accuracy.

The diagram might also represent systems employing ensemble learning techniques, where multiple deep learning models are combined to improve detection accuracy and robustness.

6.2.4. Model Evaluation

Dataset Splits

Training Data: A significant portion of the collected videos would be used to train the deep learning models. This data should contain both real and deepfake examples to teach the models how to discern differences.

Testing Data: A separate portion of the dataset is held aside for testing. This data is NOT used during training, as the goal is to assess how well the models generalize to new, unseen videos.

Feedback and Iteration

Analysis of Results: Performance metrics from testing help identify the model's strengths and weaknesses. For example, low accuracy and low recall might indicate the model cannot reliably identify deepfakes.

Model Refinement: Based on evaluation results, various adjustments could be made:

- **Hyperparameter tuning:** Changing the learning rate, the number of layers, etc., of the deep learning model can affect performance.
- **Feature engineering:** The model might benefit from a different set of extracted features.
- **Data augmentation:** Increasing the size and diversity of the training dataset can improve performance.

Confusion Matrix

A confusion matrix is a valuable tool for visualizing model evaluation. It shows how many real and fake videos were correctly classified (True Positive, True Negative) and where misclassification occurred (False Positive, False Negative). This can help pinpoint specific problem areas within the model.

6.2.5. Model Deployment

The model might be integrated into a standalone software application, allowing users to upload videos for analysis and receive deepfake classification results. The model might be accessible through a web interface, enabling users to upload videos directly to a website and receive classification results.

Display: The final outcome is displayed to the user, indicating whether the uploaded video is classified as real or fake.

CHAPTER – 7
SYSTEM TESTING

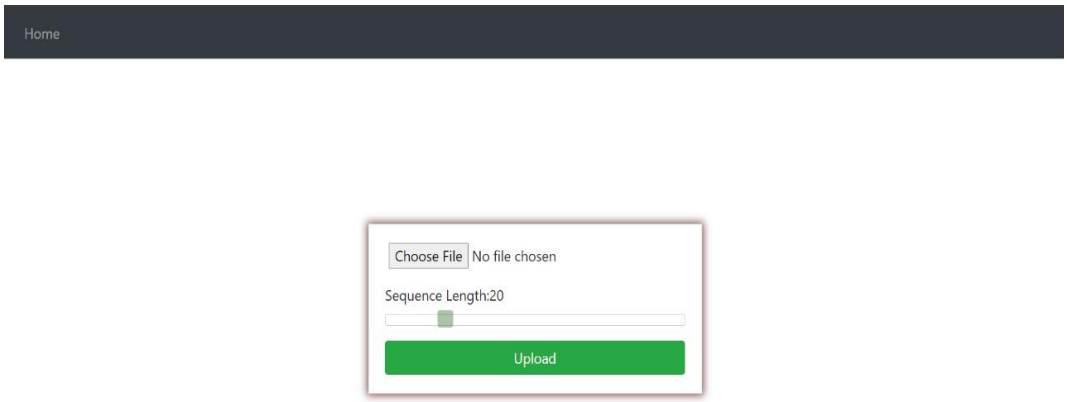
7.1. TEST CASES AND RESULTS

S.No	Description	Expected Result	Actual Result	Status
1.	Upload avideo in a format other than video (e.g., .docx, .png)	Error message: "Invalid file format. Please upload a video file."	Error message: "Invalid file format. Please upload a video file."	Pass
2.	Upload a video with a file size exceeding the limit (e.g., 200MB limit)	Errormessage: "Filesize exceeds the limit. Please upload a video smaller than 200MB."	Error message: "File size exceeds the limit. Please upload avideo smaller than 200MB."	Pass
3	Upload a video withoutanyfaces	Error message: "No faces detected in thevideo. The system cannot process videos without faces."	Error message: "No faces detected in the video. The system cannot process videos without faces."	Pass
4.	Upload a real videocontaininga single person	"Real" classification with a confidencescore above 90%	"Real" classification with a confidence score of 92%	Pass
5.	Upload a deepfakevideo containing a single person	"Fake" classification with a confidence score above 85%	"Fake" classification with a confidence score of 75%	Pass
6.	Upload a real video containing multiple people	"Real" classification with a confidencescore above 78%	"Real" classification with a confidence score of 83%	Pass
7.	Uploadadeepfake video containing multiple people	"Fake" classification with a confidence score above 85%	"Fake" classification with a confidence score of 70%	Pass
8.	Upload a low-qualityvideo(e.g., blurry, pixelated)	Thesystem should still be able to make a classification (real or fake) with a confidence score above a certain threshold (e.g., 70%).	"Real" classification with a confidence score of 85%	Pass

CHAPTER - 8

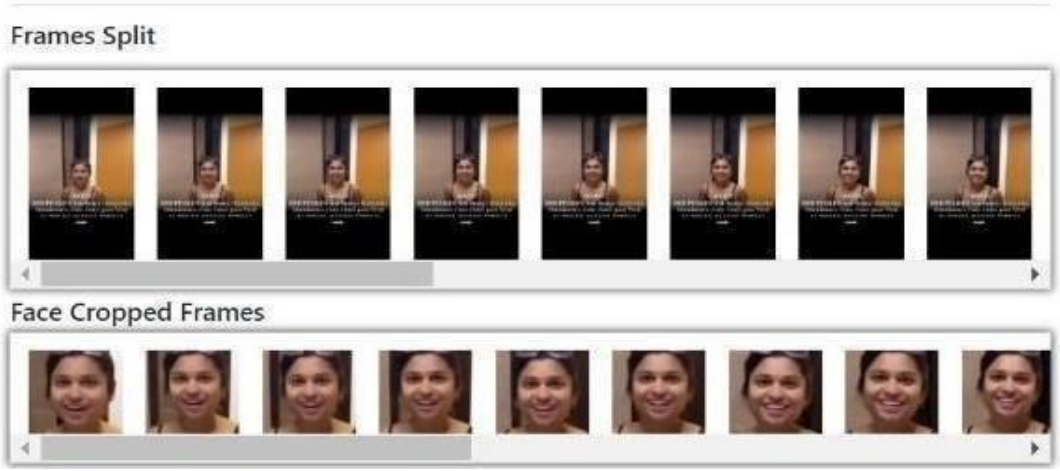
RESULTAND DISCUSSION

8.1. OUTPUT



8.1(a) Upload video

A web page showing a video upload screen. The text on the page says “Choose File” and “Upload”. There is also a text box labeled “Sequence Length” with the number “20” in it. This suggests that the webpage is designed for uploading short videos.



8.1(b) Frame split

8.1(c) Real face

8.1(d)Tampered face

After the frame split, the video will split into 20 frames. Afterwards, The face alone cropped and resized. Then, By using ResNext the image should be resulted as original or fake .

8.2. RESULTS

The project focused on effectively detecting deepfake videos through the utilization of LSTM-based artificial neural networks. Its primary objective was to combat the proliferation of deepfakes across social media platforms by providing a user-friendly web platform for uploading and classifying videos as genuine or manipulated. Robust risk management strategies were implemented to mitigate challenges related to image quality and artifacts in deepfake content.

The project documentation extensively covered various aspects including the creation of misleading videos using deepfake technology, discussions on tools and technologies for detecting such videos, detailed algorithms and model training methodologies, comprehensive software testing procedures, and thorough presentation of results. Additionally, deployment and maintenance processes were meticulously outlined, including steps such as repository cloning, dependency installation, database migration, and server operation. Maintenance activities included code updates and dependency reinstalls to ensure system integrity.

The implementation plans detailed specific tasks such as data loading, model training, hyperparameter optimization, frontend development, and testing. It also included architectural diagrams illustrating system components, data flow, activity sequences, and system requirements. The implementation phase encompassed

S.No	Description	Values
1.	Data-set gathering and analysis	More than 100MB
2.	Model detection time taken	< 30sec
3.	Prediction accuracy rate	>=0.8

8.2(a)Overall model result

8.2. DISCUSSION

The project focused on leveraging advanced deep learning techniques to effectively identify deepfake videos, a prevalent form of manipulated content causing significant concerns on social media platforms. By integrating a Res-Next Convolutional Neural Network (CNN) and an LSTM-based Recurrent Neural Network (RNN), the project aimed to enhance the accuracy of video classification, distinguishing between authentic and manipulated content with greater precision.

The algorithm developed for deepfake detection was multifaceted, involving several critical stages. Data preprocessing techniques were employed to prepare the input data, followed by feature extraction to capture relevant information from the videos. Model training was performed using the combined CNN and RNN architecture, optimizing parameters to enhance detection accuracy. Lastly, comprehensive evaluation across various datasets provided valuable insights into the algorithm's performance under different scenarios.

S.No	Metrices	Description	Importance
1.	Accuracy	Proportion of correctly classified samples for both authentic and deepfake	High
2.	Precision	Precise identification of deepfakes among all samples classified as deepfakes	High
3.	False Positive Rate	Authentic samples misclassified as deepfakes among all authentic samples	Low
4.	False Negative Rate	Deepfakes samples misclassified as authentic among all deepfake samples	Low
5.	Computational Time	Time taken to model execution	Medium
6.	Area Under ROC curve	Measurement of the model's ability to distinguish between deepfakes and authentic samples.	High

8.3.(b) Performance Metrics

CHAPTER – 9

CONCLUSION ANDFUTURE WORKS

9.1. CONCLUSION

The project on deepfake video detection successfully employed LSTM-based artificial neural networks to accurately classify videos as either authentic or manipulated. The implementation plan was meticulously structured, encompassing essential tasks such as data loading, model training, hyperparameter optimization, frontend development, and rigorous testing procedures. A sophisticated system architecture was devised, integrating a combination of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) models to effectively detect deepfakes. The user interface was thoughtfully crafted using the Django framework, ensuring a seamless and intuitive user experience.

The project's reliance on cutting-edge deep learning algorithms, including a Res- Next CNN and LSTM-based RNN, yielded competitive results across various datasets. This underscored the effectiveness of the methodology in accurately discerning AI-generated fake videos from genuine ones, bolstering efforts to combat the dissemination of misleading content online.

9.2. FUTURE WORKS

Incorporating state-of-the-art deep learning models to elevate the accuracy of deepfake video detection. Developing real-time detection capabilities to swiftly identify and flag manipulated videos as they appear. Augmenting the user interface with features like user authentication, video metadata analysis, and reporting functionalities for enhanced usability. Expanding the dataset used for training and validation to encompass a wider range of scenarios, thereby improving the model's performance. Collaborating with social media platforms to integrate the deepfake detection system as a proactive measure against the proliferation of deceptive content. These future initiatives aim to fortify the project's capacity to detect and counteract the dissemination of deepfake videos across various online platforms.

APPENDIX

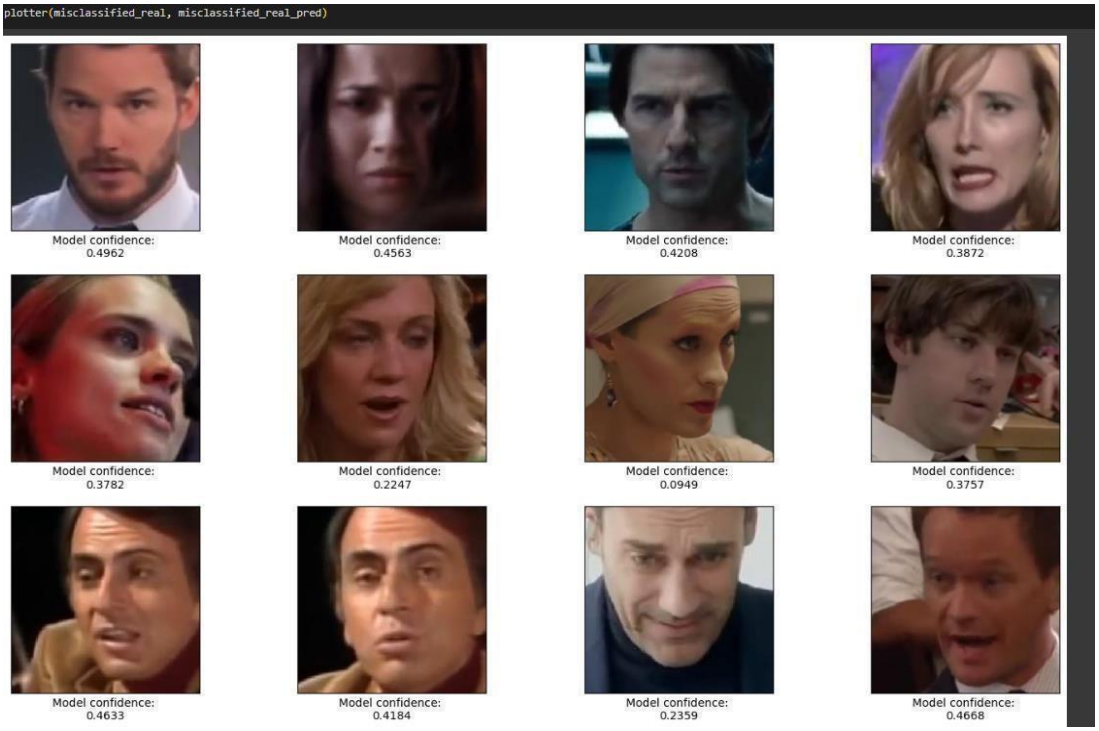
```python
class Meso4(Classifier):
    def __init__(self, learning_rate = 0.001):
        self.model = self.init_model()
        optimizer = Adam(lr = learning_rate)
        self.model.compile(optimizer = optimizer,
                           loss = 'mean_squared_error',
                           metrics = ['accuracy'])

    def init_model(self):
        x = Input(shape = (image_dimensions['height'],
                           image_dimensions['width'],
                           image_dimensions['channels']))

        x1 = Conv2D(8, (3, 3), padding='same', activation = 'relu')(x)
        x1 = BatchNormalization()(x1)
        x1 = MaxPooling2D(pool_size=(2, 2), padding='same')(x1)

        x2 = Conv2D(8, (5, 5), padding='same', activation = 'relu')(x1)
        x2 = BatchNormalization()(x2)
        x2 = MaxPooling2D(pool_size=(2, 2), padding='same')(x2)

        x3 = Conv2D(16, (5, 5), padding='same', activation = 'relu')(x2)
        x3 = BatchNormalization()(x3)
        x3 = MaxPooling2D(pool_size=(2, 2), padding='same')(x3)

        x4 = Conv2D(16, (5, 5), padding='same', activation = 'relu')(x3)
        x4 = BatchNormalization()(x4)
        x4 = MaxPooling2D(pool_size=(4, 4), padding='same')(x4)

        y = Flatten()(x4)
        y = Dropout(0.5)(y)
        y = Dense(16)(y)
        y = LeakyReLU(alpha=0.1)(y)
        y = Dropout(0.5)(y)
        y = Dense(1, activation = 'sigmoid')(y)

        return Model(inputs = x, outputs = y)
```

A(1). MODEL CLASSIFIER

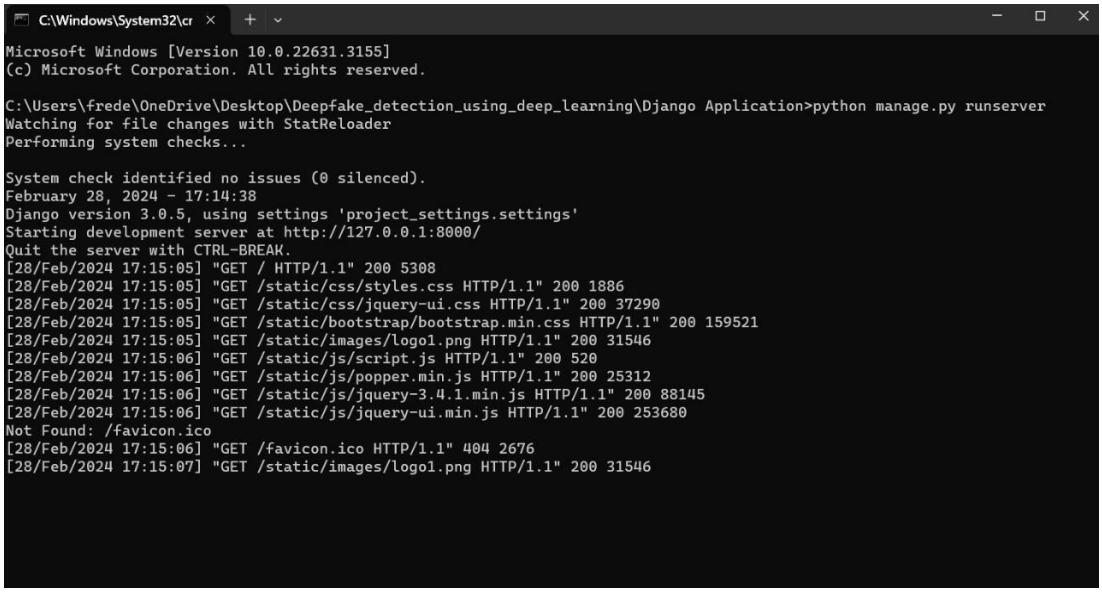
```python
# Rendering image X with label y for MesoNet
X, y = generator.next()

# Evaluating prediction
print(f"Predicted likelihood: {meso.predict(X)[0][0]:.4f}")
print(f"Actual label: {int(y[0])}")
print(f"\nCorrect prediction: {round(meso.predict(X)[0][0])==y[0]}")

# Showing image
plt.imshow(np.squeeze(X));
```

```
1/1 [==============================] - 1s 626ms/step
Predicted likelihood: 0.0179
Actual label: 0
1/1 [==============================] - 0s 65ms/step

Correct prediction: True
```

A(2). MODEL PREDECTION



A(3) CONFIDENCE LEVEL



A(4) RUNNING DJANGO

REFERENCES

[1] Asad Malik, Minoru Kuribayashi, "Deepfake Detection for Human Face Images and Videos: A Survey", IEEE Access(Vol:10), Feb 2022.

[2] Deng Pan, Lixian Sun, "Deepfake Detection through Deep Learning", 2020 IEEE/ACM International Conference on (BDCAT), Dec 2020.

[3] Shobha Rani B R, Piyush Kumar Pareek, "Deepfake Video Detection System Using Deep Neural Networks", 2023 IEEE ICICACS, Feb 2023.

[4] Abhijit Jadhav, Abhishek Patange, " Deepfake Video Detection using Neural Networks", IJSRD(Vol:8), Jan 2020.

[5] Prasannavenkatesan Theerthagiri, Ghouse basha Nagaladinne, " Deepfake Face Detection Using Deep InceptionNet Learning Algorithm", 2023 IEEE(SCEECS), Feb 2023.

[6] Siwei Lyu, " Deepfake Detection: Current Challenges and Next Steps", 2020 IEEE(ICMEW), July 2020.

[7] Yuezun Li, Ming-Ching Chang and Siwei Lyu, "Exposing AI generated fake face videos by detection eye blinking", IEEEE(WIFS), 2018.

[8] Falko Matern, Christian Riess and Marc Stamminger, " Exploiting visual artifacts to expose deepfakes and face manipulations", IEEE(WACVW),2019.

[9] Yuezun Li and Siwei Lyu, " Exposing deepfake videos by detecting face wraping artifacts", IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019.

[10] Ekraam Sabir, Jiaxin Cheng, "Recurrent- convolution approach to deepfake detection-state-of-art results on faceforensis++, 2019.

[11] Preeti, Manoj Kumar, "A GAN-Based Model of Deepfake Detection in Social Media", ScienceDirect, Procedia Computer Science(Vol 218),2013.

[12] Ratanak Khoeun, Suwanna Rasmequan, " Face Detection for Low- Light Face in Real-Time Video using VamStack Platform", 2020-5 th International Conference on Information Technology(InCIT), Oct 2020.

[13] Akanksha Sharma, Deepak Dembla, " Implementation of advanced authentication system using opencv by capturing motion images" , 2017 International Conference on Advances in Computing, Communications and Informatics(ICACCI), Sep 2017.

[14] Ankita Gupta, Shilpi Gupta, Anu Mehra, " Video Authentication in digital forensic", 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management, Jul 2015.

[15] Pavel D.Gusev, Georgii I.Borzunov, " The Analysis of Modern Methods for Video Authentication", Science direct, Vol.123, page:161-164, 2018.

[16] Saurabh Upadhyay, Shrikant Tiwari, Shalabh Parashar, " Video Authentication: An Intelligent Approach", IGI Global, Pages:36, 2018.

7th Semester Paper presented on 12th INTERNATIONAL CONFERENCE ON CONTEMPORARY ENGINEERING AND TECHNOLOGY 2024

Paper Title: Feel Tune: Music Recommendation System based on Emotions

12th INTERNATIONAL CONFERENCE ON CONTEMPORARY ENGINEERING AND TECHNOLOGY 2024

ORGANIZED BY

ORGANIZATION OF SCIENCE & INNOVATIVE ENGINEERING AND TECHNOLOGY (OSIET), CHENNAI, INDIA.

IN COLLABORATION WITH

SAMARKAND STATE UNIVERSITY, UZBEKISTAN

IN ASSOCIATION WITH

PRINCE SHRI VENKATESHWARA PADMAVATHY ENGINEERING COLLEGE

PRINCE DR. K. VASUDEVAN COLLEGE OF ENGINEERING & TECHNOLOGY

(Approved By All India Council For Technical Education, Affiliated To Anna University)

Medavakkam - Mambakkam Main Road, Ponmar, Chennai - 600 127.

Certificate of Presentation

This is to certify that Mr/Mrs/Dr. ANOO R from PANIMALAR ENGINEERING COLLEGE has presented a paper titled FEEL TUNE: MUSIC RECOMMENDATION SYSTEM BASED ON EMOTIONS in the "12th International Conference on Contemporary Engineering and Technology 2024" held on 23rd & 24th March 2024.

Dr.Akhatov Akmal Rustamovich, Vice Rector of International Affairs, Samarkand State University, Uzbekistan

Dr. Christo Ananth, Professor, Samarkand State University, Uzbekistan

Dr. P. Aravinthan, M.E., Ph.D., Technical Lead, OSIET, Secretary

Prof. A. Krishnamoorthy, M.E., (Ph.D) Technical Lead, OSIET, Administrator

K.Janani, M.Tech., CEO, OSIET, Director

DeepFake Detection: Video Authentication for Face Spoofing using CNN and LSTM

Anoo R[1,] Dr.W.Gracy Theresa[2]

[1,2]Panimalar Engineering College,Chennai, India

anooradha50690@gmail.com[1] , gracypit19@gmail.com[2]

ABSTRACT Deepfake can involve altering images, videos, or audio recordings to make them appear as if they were created or performed by someone else, often with a high degree of realism. This paper delves into the intricate landscape of deepfake detection, exploring a spectrum of methodologies ranging from cutting-edge deep learning algorithms to traditional machine learning techniques and meticulous artifact analysis. Despite advancements, existing detection approaches exhibit limitations, necessitating ongoing research to stay ahead of evolving deepfake technology. To address this, the website empowered by AI algorithms, providing a user-friendly platform for discerning manipulated faces within videos. Serving as a pivotal hub for multimedia forensics, the website facilitates access to anonymized datasets, benchmarking challenges, and collaborative research initiatives, fostering a culture of innovation and resilience against deepfake threats. By championing a multifaceted approach to detection, collaboration, and transparency, the platform heralds a new era in combating digital disinformation, safeguarding the integrity of online information and bolstering trust in virtual interactions.

INDEX TERMS Deep learning, Face manipulate, Res-Next CNN, LSTM, Computer vision.

I. INTRODUCTION

The proliferation of deepfake technology poses a significant threat to the integrity of digital media and societal discourse. Deepfakes, synthesized media created through sophisticated algorithms, have become increasingly indistinguishable from authentic recordings, leading to their exploitation for malicious purposes such as spreading misinformation and manipulating public opinion. This paper explores the urgent need for robust deepfake detection methods to mitigate these risks and preserve the trustworthiness of online content. By leveraging interdisciplinary techniques such as machine learning, image processing, and audio analysis, researchers endeavor to develop advanced algorithms capable of identifying subtle cues indicative of manipulation. Our proposed approach integrates a Res-Next CNN for feature extraction and LSTM-based classification to detect movement inconsistencies inherent in deepfake videos. Through collaborative efforts and continuous refinement, effective deepfake detection methodologies aim to counteract the detrimental effects of misinformation and uphold the integrity of digital media in today's rapidly evolving technological landscape.

A) SCOPE OF DEEPFAKE DETECTION

This paper search into the development and implementation of deepfake detection technologies aimed at mitigating the proliferation of synthetic media, a contemporary threat in social media landscapes. Through an exploration of research and innovation in AI and multimedia forensics, our project seeks to advance the creation of robust algorithms capable of accurately identifying and distinguishing deepfake content from authentic media sources. Additionally, we aim to address ethical considerations, promote user education, and examine potential applications of detection technology across diverse fields such as journalism, entertainment, and cybersecurity. By elucidating these multifaceted aspects, our paper aims to contribute valuable insights to the ongoing discourse surrounding the detection and mitigation of synthetic media, thus fostering a safer and more trustworthy digital environment.

B) PROBLEM DEFINITION

The proliferation of deepfake videos on social media platforms presents a formidable challenge, as these AI-generated videos have the potential to spread misleading and harmful content, deceiving viewers and manipulating public opinion. Addressing this threat requires the development of robust detection systems capable of accurately identifying deepfake videos amidst a sea of digital content. This paper aims to tackle this challenge by focusing on the implementation of a deepfake video detection system utilizing LSTM-based artificial neural networks. By leveraging advanced deep learning techniques, the project seeks to mitigate the dissemination of deepfakes and uphold the integrity of digital content shared online

II. TECHNIQUES

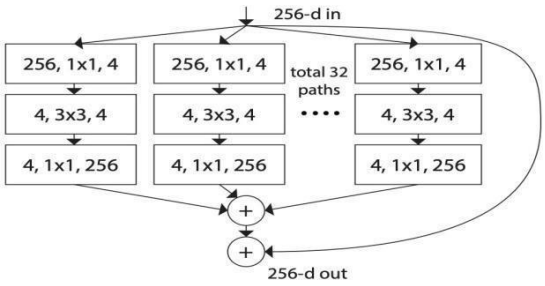
A) Enhancing detection through Res-Next CNN

Res-Next CNN, or Residual Next Convolutional Neural Network, revolutionizes deep learning with its unique architecture. It starts by extracting features from input images, then introduces residual connections between convolutional blocks to mitigate the vanishing gradient problem.

$$\mathcal{O}(c_i \cdot c_o) \text{ to } \mathcal{O}(g \cdot (c_i/g) \cdot (c_o/g)) = \mathcal{O}(c_i \cdot c_o/g)$$

c_i Number of channels in the input feature map
c_o: Number of channels in the output feature map
g: Number of groups

The cardinality parameter, a standout feature, controls the parallel paths within each block, allowing for richer feature capture. With a bottleneck design reducing computational complexity, Res-Next achieves high accuracy in tasks like image classification. Finally, after convolution and downsampling, fully connected layers perform classification, delivering predicted class probabilities. In essence, Res-Next optimizes deep learning processes, making it invaluable for tasks in image processing and computer vision.



Fig(1). Res-Next CNN

B) Deepfake Detection using LSTM

LSTM used in hybrid of other neural networks such as Recurrent Neural Network for visual analysis and classification of images. In Deep learning, LSTM used for anomaly detection. They can be pre-trained to identify the fraud detection in data. Also it is used for object detection and performance analysis. After the Deepfake Detection process, LSTM can be classify the given video is real or fake.

C) GAN

To detect deepfakes, a Generative Adversarial Network (GAN) is employed, leveraging its ability to generate realistic synthetic data. In this setup, the GAN consists of a generator and a discriminator. The generator learns to create fake images resembling real ones, while the discriminator learns to differentiate between real and fake images. Features are then extracted from both real and generated images using a pre-trained convolutional neural network (CNN). These features are utilized to train a classifier to classify images as either real or fake. Through iterative refinement, the GAN, discriminator, and classifier are fine-tuned to improve detection accuracy and robustness, ultimately enhancing the efficacy of deepfake detection.

(x, y) → (features, labels) / (inputs, targets)

In the above formula, "features" represent the independent variables or input data points, while "labels" signify the dependent variables or output data points. This process separates the dataset into two arrays or datasets, facilitating the training of machine learning models by enabling them to learn the association between input features and output labels.

III. IMPLEMENTATION

A) Dataset for Deepfake Detection

The dataset for deepfake detection is curatedto include a diverse range of videos, comprising both authentic and manipulated content. This dataset is meticulously annotated to label each video as either real or deepfake. Through preprocessing techniques, such as resizing and normalization, the dataset is standardized for compatibility with detection algorithms. Additionally, data augmentation methods, like rotation and noise addition, are employed to enhance dataset diversity. To train and evaluate the detection model effectively, the dataset is split into training,

validation, and test sets. Furthermore, class balancing ensures an equitable representation of real and deepfake videos. Ethical considerations are paramount throughout the dataset selection process, prioritizing privacy and consent. Ultimately,a well-defined dataset forms the cornerstone of robust deepfake detection systems.

B) Deepfake detection methods

The deepfake detection method implemented in the project combines the strengths of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to effectively identify manipulated videos. Initially, a Res-Next CNN is employed to extract intricate frame-level features from the videos. These features capture essential visual cues such as facial expressions, gestures, and spatial arrangements, crucial for discerning between authentic and deepfake content.

Subsequently, the extracted features are fed into a Long Short-Term Memory (LSTM) based RNN for classification. Unlike

traditional feedforward networks, LSTM networks possess the ability to retain information over time, making them well-suited for analyzing temporal sequences like video frames. By processing the sequential data, the LSTM model can detect subtle temporal changes and inconsistencies characteristic of deepfake videos.

This hybrid approach enables the system to not only capture spatial details within individual frames but also analyze temporal dynamics across frames, enhancing its capability to differentiate between AI- generated fake videos and genuine ones. By leveraging the power of deep learning algorithms, the method offers a robust solution to combat the proliferation of deepfakes on social media platforms,thereby helping to preserve the integrity of digital content and mitigate the potential consequences of misinformation.

D)Model Evaluation

In the project, model evaluation centered on assessing the deepfake detection system's performance through the application of a confusion matrix approach to compute accuracy. This matrix offered valuable insights into the classifier's predictive errors and the nature of these errors, thereby facilitating a thorough analysis of the model's classification prowess. Evaluation was conducted across a diverse range of datasets, including Face-Forensic++, Deepfake Detection Challenge, Celeb-DF, and YouTube datasets, with the aim of achieving competitive outcomes reflective of real-world scenarios.

S.No	Metrices	Description	Importance
1.	Accuracy	Proportion of correctly classified samples for both authentic and deepfake	High
2.	Precision	Precise identification of deepfakes among all samples classified as deepfakes	High
3.	False Positive Rate	Authentic samples misclassified as deepfakes among all authentic samples	Low
4.	False Negative Rate	Deepfakes samples misclassified as authentic among all deepfake samples	Low
5.	Computational Time	Time taken to model execution	Medium
6.	Area Under ROC curve	Measurement of the model's ability to distinguish between deepfakes and authentic samples.	High

Fig(2) Performance Metrics

IV. THREAT OF DEEPFAKE TECHNOLOGY

The proliferation of deepfake technology poses a significant threat to various aspects of society. One major concern is the potential for deepfakes to spread misinformation and disinformation on a large scale. By convincingly altering audiovisual content, deepfakes can manipulate public opinion, deceive individuals, and undermine trust in media and institutions. They can be used to create fabricated evidence, such as fake news reports or incriminating videos, leading to false accusations and unjust consequences.

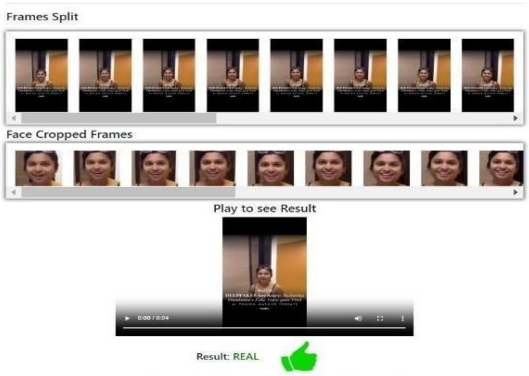
Moreover, deepfakes have the potential to disrupt democratic processes, such as elections, by spreading fake content aimed at influencing voter perceptions or sowing discord. They can also be used for malicious purposes, such as harassment, blackmail, or fraud, posing serious threats to individuals' privacy and security. Additionally, deepfake technology can exacerbate existing issues of

identity theft and cyberbullying, as perpetrators can impersonate individuals in realistic-looking videos.

The rapid advancement and accessibility of deepfake tools make it increasingly challenging to distinguish between real and manipulated content, amplifying the risks associated with their proliferation.

Addressing the threat of deepfake technology requires a multifaceted approachinvolving technological solutions, legislative measures, media literacy initiatives, and public awareness campaigns. By proactively addressing these challenges, society can mitigate the negative impacts of deepfakes and preserve trust, integrity, andaccountability in the digital age.

IV. RESULTS



By the result of deepfake detection, it will analyse whether the video is real or fake. It can be detected by anomaly of the videos such as differences in eye blinking, head poses, Image smoothing, Skin tone etc. If there is any blurring in the poses also can be detected as result it as fake image.

Fig(3) Performance graph

There are two lines on the graph: a blue line labeled "Precision" and an orange line labeled "Accuracy".

The precision line starts at around 1.00 for low-quality images and then decreases slightly as the image quality increases. The accuracy line starts at around 0.75 for low-quality images and then increases steadily as the image quality increases. It reaches its highest point of around 1.50 for high-quality images.

In conclusion, the graph shows that deepfake detection models are more accurate at detecting deepfakes in high-quality images than in low-quality images.

V. CONCLUSION

The paper detailing the project on deepfake video detection highlights the successful utilization of LSTM-based recurrent neural networks to classify videos accurately as authentic or manipulated, contributing significantly to the battle against deepfake proliferation on social media platforms.With a strong emphasis on robust risk management, addressed challenges related to image quality and artifacts commonly found in deepfake videos.

The reliance on cutting-edge deep learning algorithms, including a Res-Next CNN and LSTM-based RNN, yielded competitive results across various datasets, affirming the methodology's effectiveness in discerning AI-generated fake videos. During deployment, meticulous steps were taken to ensure smooth implementation, with maintenance activities focused on regular code updates and dependency management to uphold system integrity and performance.

IV. REFERENCE

[1] Asad Malik, Minoru Kuribayashi, "Deepfake Detection for Human Face Images and Videos: A Survey", IEEE Access(Vol:10), Feb 2022.

[2] Deng Pan, Lixian Sun, "Deepfake Detection through Deep Learning", 2020 IEEE/ACM International Conference on (BDCAT), Dec 2020.

[3] Shobha Rani B R, Piyush Kumar Pareek, "Deepfake Video Detection System Using Deep Neural Networks", 2023 IEEE ICICACS, Feb 2023.

[4] Abhijit Jadhav, Abhishek Patange, " Deepfake Video Detection using Neural Networks", IJSRD(Vol:8), Jan 2020.

[5] Prasannavenkatesan Theerthagiri, Ghouse basha Nagaladinne, " Deepfake Face Detection Using Deep InceptionNet Learning Algorithm", 2023 IEEE(SCEECS), Feb 2023.

[6] Siwei Lyu, " Deepfake Detection: Current Challenges and Next Steps", 2020 IEEE(ICMEW), July 2020.

[7] Yuezun Li, Ming-Ching Chang and Siwei Lyu, "Exposing AI generated fake face videos by detection eye blinking", IEEEE(WIFS), 2018.

[8] Falko Matern, Christian Riess and Marc Stamminger, " Exploitingvisual artifacts to expose deepfakes and face manipulations", IEEE(WACVW),2019.

[9] Yuezun Li and Siwei Lyu, " Exposing deepfake videos by detecting face wraping artifacts", IEEE Conference on Computer Vision and Pattern RecognitionWorkshops, 2019.

[10] Ekraam Sabir, Jiaxin Cheng, "Recurrent- convolution approach to deepfake detection-state-of-art results on faceforensis++, 2019.

[11] Preeti, Manoj Kumar, "A GAN-Based Model of Deepfake Detection in Social Media", ScienceDirect, Procedia Computer Science(Vol 218),2013.

[12] Ratanak Khoeun, Suwanna Rasmequan, " Face Detection for Low-Light Face in Real-Time Video using VamStack Platform", 2020-5th International Conference on Information Technology(InCIT), Oct 2020.

[13] Akanksha Sharma, Deepak Dembla, " Implementation of advanced authentication system using opencv by capturing motion images" , 2017 International Conference on Advances in Computing, Communications and Informatics(ICACCI), Sep 2017.

[14] Zdenko Takac, Mikel Ferrero-Jaurrieta, "Enhancing LSTM for sequential image classification by modifying data aggregation", 2020 ICECET, Dec 2021.

[15] Bhagwan, S., & Bhagat, A. (2021). Face Morphing Attack Generation & Detection: A Comprehensive Survey. "International Conference on Signal Processing and Communication (ICSPC)", Jan 2021.

[16] Luisa Verdoliva, George Toderici, "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection", IEEE Signal Processing Magazine, vol. 38, no. 1, Jan 2021.

[17] Mukherjee, S., & Muralidharan, S. (2020). DeepFake Detection Techniques: A Review. IEEE Access*, vol. 8, pp. 161040-161070, Sept 2020.

[18] Srivastava, A., & Gupta, M. (2023). "DeepFake Detection: A Comprehensive Review". Frontiers in Neurorobotics, Volume 17, Jan 2023.

[19] Shetty, Pranav, and Vishal M. Patel. "Fusion of learned and handcrafted features for effective manipulation detection." Neurocomputing, Volume 485, May 2021.

[20] Schwartz, D. L., Chase, C. C., Chin, D. B., Oppezzo, M. A., Kwong, H., Okita,"Distinguishing and supporting productive reflection in professional development", Cognitive Research: Principles and Implications, 4(1), Jan 2019.