DEEPFAKE VIDEO DETECTION: A MULTI-MODEL APPROACH USING CNN, RNN & LSTM

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

OF

MASTER OF TECHNOLOGY

IN

INFORMATION SYSTEMS

Submitted by:

RAUNAK PODDAR

2K21/ISY/20

Under the supervision of

Prof. KAPIL SHARMA

DEPARTMENT OF INFORMATION TECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi

May 2023

DEPARTMENT OF INFORMATION TECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi

CANDIDATE'S DECLARATION

I, Raunak Poddar, Roll No. 2K21/ISY/20 of M.Tech. (Information Systems), Hereby declare that the dissertation report titled "DEEPFAKE VIDEO DETECTION: A MULTI-MODEL APPROACH USING CNN, RNN & LSTM" which is submitted by me to the Department of Information Technology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship, or other similar title or recognition.

Place: Delhi

Date:

RAUNAK PODDAR
(2K21/ISY/20)

DEPARTMENT OF INFORMATION TECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi

CERTIFICATE

I, hereby certify that the dissertation "DEEPFAKE VIDEO DETECTION: A MULTI-MODEL APPROACH USING CNN, RNN & LSTM", which is submitted by Raunak Poddar, Roll No. 2K21/ISY/20 (Information Systems), Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge, this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place : Delhi

Date:

Prof. Kapil Sharma

SUPERVISOR

ACKNOWLEDGEMENT

I am grateful to Prof. Dinesh Kumar Vishwakarma, HOD (Department of Information Technology), Delhi Technological University (Formerly Delhi College of Engineering), New Delhi, and all other faculty members of our department for their astute guidance, constant encouragement, and sincere support for this project work.

I would like to take this opportunity to express our profound gratitude and deep regard to our project mentor Prof Kapil Sharma, for his exemplary guidance, valuable feedback, and constant encouragement throughout the duration of the project. His valuable suggestions were of immense help throughout our project work. His perspective and criticism kept us working to make this project in a much better way. Working under him was an extremely knowledgeable experience for us.

We would also like to give our sincere gratitude to all our friends for their help and support.

: **RAUNAK PODDAR**
(2K21/ISY/20)

ABSTRACT

This thesis presents a comprehensive study on the detection of deepfake videos by leveraging the combined power of LSTM and CNN models. The objective is to accurately predict the authenticity of videos by analysing spatial and temporal features. To facilitate the development and evaluation of the proposed approach, a novel dataset is created by augmenting the existing FaceForensics++, Celeb DF, and DFDC datasets. This novel dataset encompasses a diverse range of deepfake manipulation techniques and captures various visual characteristics. Through extensive experimentation and analysis, the results demonstrate the effectiveness of the LSTM-CNN fusion model in accurately distinguishing between deepfake and authentic videos. The model successfully captures subtle visual artifacts and temporal patterns associated with deepfake manipulations, enabling high-performance deepfake detection. The research also acknowledges the significance of deepfake videos in domains such as politics and pornography, highlighting the need to ensure the integrity and trustworthiness of multimedia content. By providing an advanced framework for deepfake detection, this thesis contributes to the field of video forensics, addressing the growing concerns surrounding disinformation threats and safeguarding the authenticity of multimedia content in the digital age.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1. GENERAL

The prevalence of deepfake technology in the creation of misleading videos on social media platforms is a concerning issue. Numerous examples demonstrate the deceptive use of deepfakes featuring prominent figures such as Mark Zuckerberg during the House A.I. Hearing, Donald Trump in a Breaking Bad series as James McGill, Barack Obama in a public service announcement, and many more [5]. These instances of deepfakes generate widespread panic among the general public, highlighting the critical need for accurate detection methods to differentiate between real and manipulated videos.

Recent technological advancements have revolutionized the field of video manipulation. Modern open-source deep learning frameworks like TensorFlow, Keras, and PyTorch, coupled with affordable access to high computational power, have ushered in a paradigm shift. Traditional autoencoders [10] and pre-trained models based on Generative Adversarial Networks (GANs) have made it remarkably easy to tamper with realistic videos and images. Furthermore, smartphone and desktop applications like FaceApp and Face Swap provide convenient access to these pre-trained models, enabling the creation of highly realistic synthesized transformations of faces within real videos. These applications offer users additional functionalities such as altering hairstyles, genders, ages, and other attributes, allowing for the creation of high-quality, indistinguishable deepfakes.

While malicious deepfake videos do exist, they currently represent a minority. However, tools that generate deepfake videos, such as those mentioned in [11,12], are extensively employed for creating fake celebrity pornographic videos or engaging in revenge porn [13]. Examples include the creation of fake nude videos featuring Brad Pitt, Angelina Jolie, and others. The remarkably realistic nature of deepfake videos makes celebrities and other public figures vulnerable to pornographic content, fake surveillance videos, dissemination of fake news, and malicious hoaxes. Deepfakes also play a significant role in creating political tension [14]. Therefore, there is a

critical need to formulate resilient techniques for detecting deepfakes in order to effectively combat the extensive dissemination of manipulated videos on various social media platforms.

1.2. Project Idea

In the realm of ever-expanding social media platforms, the emergence of deepfakes poses a significant threat driven by artificial intelligence. These realistic face-swapped videos have been exploited to propagate political unrest, fabricate terrorism incidents, and exploit individuals through revenge porn and blackmail, with instances involving public figures like Brad Pitt and Angelina Jolie being particularly prominent. Consequently, the ability to discern between deepfake and pristine videos becomes crucial. In our pursuit to combat this AI-driven deception, we leverage AI itself. Deepfakes are generated using tools such as FaceApp and Face Swap, utilizing pre-trained neural networks like GANs and Autoencoders. Our approach employs a LSTM based artificial neural network to perform sequential temporal analysis on video frames, complemented by a pre-trained ResNext CNN to extract frame-level features. These extracted features are then utilized to train the LSTM-based RNN for the classification of videos as either deepfake or real. To simulate real-world scenarios and enhance the model's performance on real-time data, our methodology is trained on a large, balanced combination of usable datasets, including FF++, DFDC, and Celeb-DF. Additionally, to provide user-friendly accessibility, we have developed a frontend application where users can upload videos. The model processes the input and returns the classification of the video as deepfake or real, along with the model's confidence level. This comprehensive thesis endeavors to address the pressing challenges posed by deepfake videos and offers a practical solution for users to identify and combat their proliferation.

1.3 Motivation of the Project

The rapid advancements in mobile camera technology, coupled with the widespread influence of social media and media-sharing platforms, have revolutionized the creation and dissemination of digital videos. Deep learning has played a pivotal role in pushing the boundaries of what was once deemed unimaginable. Modern generative models, capable of synthesizing highly realistic images, speech, music, and even video, have emerged as a testament to this progress. These models have found applications in various domains, including text-to-speech for improved accessibility and generating training data for medical imaging.

However, with every metamorphic technology, new challenges arise. Deepfakes, created using deep generative models, have become a concern as they can manipulate video and audio clips. Since their emergence in late 2017, a multitude of open-source deepfake generation methods and tools have surfaced, resulting in an increasing number of synthetic media clips. While some may be intended for entertainment, others possess the potential to harm individuals and society at large. The availability of editing tools and the demand for domain expertise have contributed to the proliferation of fake videos, posing significant risks.

The rampant spread of deepfakes on social media online platforms has got commonplace, leading to spamming and the dissemination of false information. Just envision the consequences of a deepfake featuring a national leader declaring war against neighboring countries or a renowned celebrity abusing their fans. Such deepfakes have the potential to incite fear, mislead the public, and create a sense of threat.

To address this pressing issue, deepfake detection becomes imperative. In the thesis, we propose a novel deep learning-based method capable of effectively distinguishing AI-generated fake videos (deepfake videos) from original videos. The development of technology that can accurately identify and prevent the spread of deepfakes is of paramount importance in safeguarding the integrity of information shared on the internet. By leveraging advanced deep learning techniques, we aim to contribute to the ongoing efforts in combating the spread of deepfakes and mitigating their potential impact on individuals and society.

CHAPTER 2:

LITERATURE SURVEY

The study on Face Warping Artifacts [15] employed a dedicated CNN model to detect antiquity by comparing the generated areas of the face and their adjoining places. This methodology successfully distinguished between two categories of Face Artifacts. The technique specifically targeted the recognition that existing deepfake algorithms have limitations in generating high-resolution images, necessitating additional adjustments to align with the original video's faces. However, their approach did not incorporate temporal analysis of individual frames into their assessment. Detection by Eye Blinking [16] proposed a method that uses eye blinking as a crucial parameter for deepfake detection. The LRCN was utilized to analyze the temporal aspects of cropped eye blinking frames. Although eye blinking is an important clue, the detection of deepfakes should also consider other parameters, such as teeth enrichment, facial wrinkles, and eyebrow placement. The integration of capsule networks [17] facilitated the identification of manipulated images and videos through the detection process. This particular approach employed a capsule network to effectively identify altered content across different scenarios, such as detecting replay attacks and computer-generated videos. However, it should be noted that their training phase incorporated random noise, which may not be the most optimal strategy. Although their model achieved satisfactory performance on their dataset, it may encounter difficulties when confronted with real-time data due to the presence of noise during training. In contrast, our proposed method aims to train on pristine, noise-free datasets in real-time, thereby offering a The RNN[18] for deepfake detection involved sequential processing of frames using an ImageNet pre-trained model. The approach utilized the HOHO dataset consisting of only 600 videos. However, the limited number of videos and their similarity may affect the model's performance on real-time data. In our approach, we aim to train our model on a large number of diverse real-time dataThe investigation on Synthetic Portrait Videos employing Biological Signals [20] involved the extraction of biological signals from facial regions in both unaltered and deepfake portrait videos for analysis and comparison. Through the application of transformations, spatial coherence and temporal consistency were computed, effectively capturing the unique signal characteristics within feature vectors and

photoplethysmography (PPG) maps. To classify videos as either deepfakes or pristine, a probabilistic SVM and CNN were trained using the average accuracy probabilities derived from the extracted signals. This approach leverages the distinctive biological signals present in facial regions to distinguish between genuine and manipulated videos. The development of Fake Catcher aimed to achieve precise detection of fake content, regardless of the generator, content, resolution, or video quality. However, their research revealed limitations in discriminators and the preservation of biological signals, making the formulation of a differentiable loss function aligned with the proposed signal processing steps a complex endeavor.

Deepfake videos, which are artificially manipulated videos created using advanced machine learning techniques, have become a significant concern due to their potential to spread misinformation, deceive viewers, and manipulate public opinion. Consequently, the development of effective detection methods is crucial to combat the negative consequences of deepfakes. In recent years, CNN and LSTM models have emerged as powerful tools for deepfake video detection. This literature survey explores the current state-of-the-art techniques that utilize CNN and LSTM for deepfake video detection, highlighting their strengths, limitations, and advancements.

CNN-based Approaches:

CNNs have demonstrated exceptional performance in various computer vision tasks, making them suitable for deepfake detection. Several studies have employed CNN architectures, such as VGGNet, ResNet, and Inception, to extract relevant features from deepfake videos. These features capture visual artifacts, inconsistencies, and irregularities introduced during the deepfake generation process. Additionally, transfer learning techniques have been utilized to leverage pre-trained CNN models and improve detection accuracy, particularly when training data is limited.

LSTM-based Approaches:

LSTM models, a type of recurrent neural network (RNN), excel in capturing temporal dependencies, making them valuable for deepfake detection tasks involving sequential data, such as video frames. LSTM models can effectively learn and identify patterns in the temporal evolution of deepfake videos, enabling them to distinguish between genuine and manipulated sequences. Studies have explored the use of LSTM networks

in conjunction with CNNs to jointly capture spatial and temporal information for more accurate deepfake detection.

Hybrid Approaches:

To further enhance deepfake detection accuracy, researchers have proposed hybrid models that combine the strengths of both CNNs and LSTMs. These models leverage CNNs to extract spatial features from individual frames and employ LSTMs to capture temporal dependencies across frames. By integrating both spatial and temporal information, hybrid models achieve superior performance in detecting deepfake videos.

While CNN and LSTM-based approaches have shown promising results, several challenges still exist. Deepfake techniques are continually evolving, necessitating the development of robust models that can adapt to new manipulation methods. Additionally, the scarcity of large-scale annotated datasets poses a challenge for training deepfake detection models. Future research should focus on expanding datasets, exploring novel network architectures, and incorporating additional modalities, such as audio and metadata, to improve deepfake detection accuracy. However, their approach did not incorporate temporal analysis of individual frames into their assessment. Detection by Eye Blinking [16] proposed a method that uses eye blinking as a crucial parameter for deepfake detection. The LRCN was utilized to analyze the temporal aspects of cropped eye blinking frames. Although eye blinking is an important clue, the detection of deepfakes should also consider other parameters, such as teeth enhancement, facial wrinkles, and eyebrow placement. These models leverage CNNs to extract spatial features from individual frames and employ LSTMs to capture temporal dependencies across frames. By integrating both spatial and temporal information, hybrid models achieve superior performance in detecting deepfake videos. The development of Fake Catcher aimed to achieve precise detection of fake content, regardless of the generator, content, resolution, or video quality. In our approach, we aim to train our model on a large number of diverse real-time data.

CHAPTER 3:
PROBLEM DEFINITION AND SCOPE

3.1 Problem Statement

The field of digital image and video manipulation has showcased convincing techniques for several decades, primarily through the application of visual effects. However, recent advancements in deep learning have ushered in a new era of realism in fake content, leading to the proliferation of AI-synthesized media commonly known as deep fakes. Creating deep fakes using artificially intelligent tools has become relatively simple, raising concerns about their detection. These deep fakes have been exploited in various contexts, including the creation of political tensions, staging fake terrorism events, revenge porn, and blackmail, among others. Consequently, it is imperative to develop robust mechanisms for detecting and preventing the dissemination of deep fakes through social media platforms.

In this thesis, we tackle the formidable challenge of detecting deep fakes by leveraging a LSTM-based artificial neural network. By utilizing the temporal analysis capabilities of LSTM, we aim to enhance the accuracy and reliability of deep fake detection. Through our research, we contribute to the ongoing efforts in safeguarding the integrity of digital media and curbing the potential harm caused by the widespread distribution of deep fakes.

3.2 Goals and Objectives

- **Discover the distorted truth behind deep fakes:** The objective of this research is to uncover the deceptive nature of deepfake videos by developing robust and efficient detection methods. By investigating and analyzing the visual artifacts and inconsistencies introduced during the deepfake generation process, this study aims to expose the distorted truth behind these manipulated videos and mitigate their potential harmful impacts on society.
- **Reduce abuses and the spread of misleading information on the internet:** The main objective of this research is to mitigate the abuse and dissemination of misleading information on the internet by developing effective techniques

for detecting and identifying deepfake videos. By accurately detecting deepfakes, this study aims to contribute to the reduction of misinformation, safeguarding the integrity of digital media, and promoting trustworthiness in online content.

- **Distinguish and classify videos as deepfake or authentic:** The main objective of this research is to develop a reliable and accurate system that can distinguish and classify videos as either deepfake or authentic. By leveraging advanced deep learning techniques, this study aims to create a robust detection model that can effectively differentiate between manipulated videos and genuine ones, contributing to the fight against misinformation and deceptive media.
- **Develop an easy-to-use system for users to upload videos and determine their authenticity:** The objective of this research is to create a user-friendly system that enables users to upload videos and assess their authenticity easily. By developing an intuitive interface and employing advanced deepfake detection techniques, this study aims to empower users to determine the credibility of videos, thereby enhancing media literacy and enabling individuals to make informed judgments about the authenticity of visual content.

3.3 Statement of Scope

The current landscape offers numerous tools for creating deep fakes, but the availability of tools for detecting such synthetic media is severely limited. Our innovative approach to deep fake detection represents a significant contribution towards mitigating the proliferation of deep fakes on the World Wide Web. We are developing a robust web-based platform that enables users to upload videos and accurately classify them as either fake or authentic.

This project has the potential to evolve beyond a web-based platform, with future prospects including the development of a browser plugin capable of seamlessly detecting deep fakes automatically. By integrating our solution into popular applications like WhatsApp and Facebook, users can benefit from easy pre-detection of deep fakes before sharing content with others.

CHAPTER 4:

METHODOLOGIES OF PROBLEM-SOLVING

4.1 Analysis

Solution Requirement:

After carefully analysing the problem statement, we assessed the feasibility of the proposed solution. Extensive research was conducted, as outlined in section 3.3, to explore existing literature and studies related to the problem domain. Once the feasibility was established, our focus shifted to dataset gathering and analysis. We explored different training approaches, such as solely training the model with either fake or real videos, to evaluate their effectiveness. However, we discovered that such approaches introduced bias into the model and compromised the accuracy of predictions. To address this issue, we conducted thorough research and determined that a balanced training approach is the most suitable. By achieving a balance between fake and real videos during training, we were able to mitigate bias and variance in the algorithm, resulting in improved accuracy.

4.2 Parameters Identified:

- Blinking of eyes
- Teeth enchantment
- Bigger distance for eyes
- Moustaches
- Double edges, eyes, ears, nose
- Iris segmentation
- Wrinkles on face
- Inconsistent head pose
- Face angle
- Skin tone
- Facial Expressions
- Lighting
- Different Pose

- Double chins
- Hairstyle
- Higher cheek bones

4.3 Experimental Setup

After analysis we decided to use the PyTorch framework along with python3 language for programming. PyTorch is chosen as it has good support to CUDA i.e., GPU and it is customize-able. Google Cloud Platformfor training the final model on large number of data-set.

4.4 Evaluation

We evaluated our model with a large number of real time dataset which include YouTube videos dataset. Confusion Matrix approach is used to evaluate the accuracy of the trained model.

4.5 Outcome

The outcome of the solution is trained deepfake detection models that will help the users to check if the new video is deepfake or real.

4.6 Project Model Analysis

We have chosen to utilize the Spiral model(Fig 1) as our software development approach, given its focus on the individuals involved in the project, their collaborative work, and effective risk management. The Spiral model is well-suited for our application due to its ability to facilitate rapid changes and continuous evaluations throughout the development process, aligning with our goal of meeting expected outcomes. By adopting the Spiral model, we can efficiently develop our application in various modules, ensuring flexibility and adaptability. This approach also allows for active client involvement throughout the project, including feature prioritization, iteration planning, and review sessions that incorporate new functionalities. However, it is essential for clients to understand that they will witness a work in progress, as transparency is a key benefit of this approach. Given the complexity and inherent risks associated with our model, the Spiral model provides the necessary framework to effectively manage these risks, making it the ideal choice for our product development.

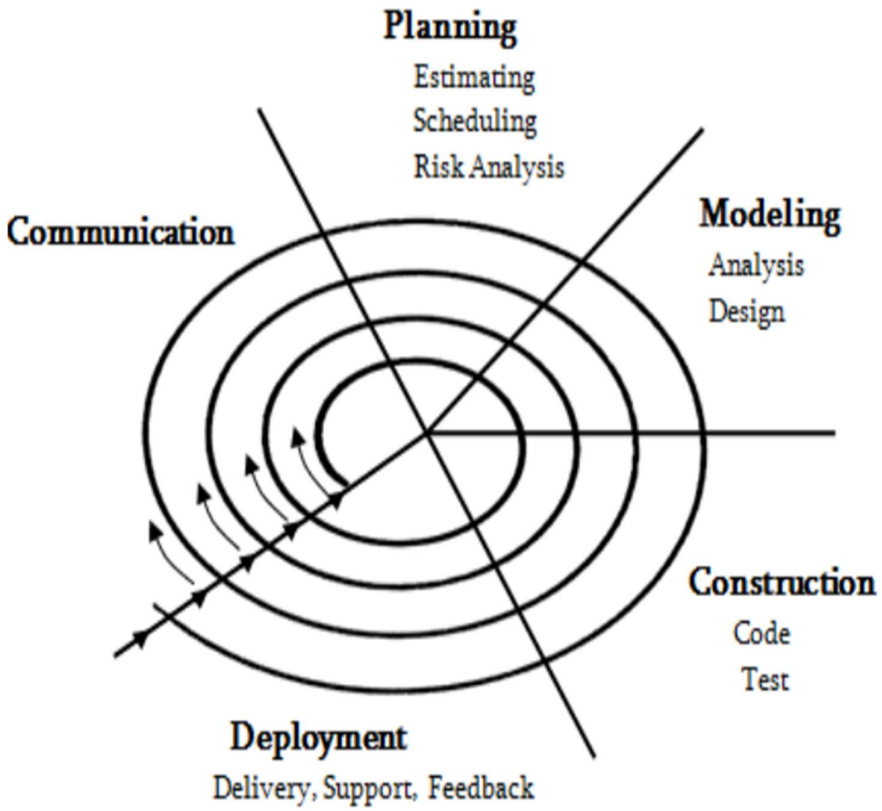

Fig 1: Spiral Model

Chapter 5: Requirements

5.1 Hardware Resources Deployed

Table 1: Hardware Requirements

Sr. No.	Parameter	Minimum Requirement
1	Intel Xeon E5 2637	3.5 GHz
2	RAM	16 GB
3	Hard Disk	100 GB
4	Graphic card	NVIDIA GeForce GTX Titan (12 GB RAM)

5.2 Software Resources Deployed

Table 2: Software Resources Requirements

1.	Operating System	Windows 7+
2.	Programming Language	Python 3.0
3.	Framework	PyTorch 1.4 , Django 3.0
4.	Cloud platform	Google Cloud Platform
5.	Libraries	OpenCV, Face-recognition

5.3 Cost Estimation

Table 3: Cost Estimation

	Description	Cost(in INR)
1	Pre-processing the dataset on GCP	5580
2	Training models on GCP	2785
3	Deploying project to GCP using Cloud engine	3160
4	Google Colab Pro subscription	975
	Total Cost (in INR)	**12500**

CHAPTER 6:

IMPLEMENTATION

6.1 System Architecture

This thesis presents a methodology(as described in Fig 2) for predicting deepfakes in videos by combining LSTM and CNN architectures. A novel dataset is created by mixing the "FF++" , “Celeb DF”, and "DFDC" datasets, which enables comprehensive training and evaluation. The proposed approach extracts frames, detects facial regions, and applies image preprocessing techniques. The CNN captures spatial features, while the LSTM captures temporal dependencies. The fusion of LSTM and CNN enhances accuracy of the classification. The model is trained using the augmented dataset and evaluated using recall, precision, accuracy metrics.

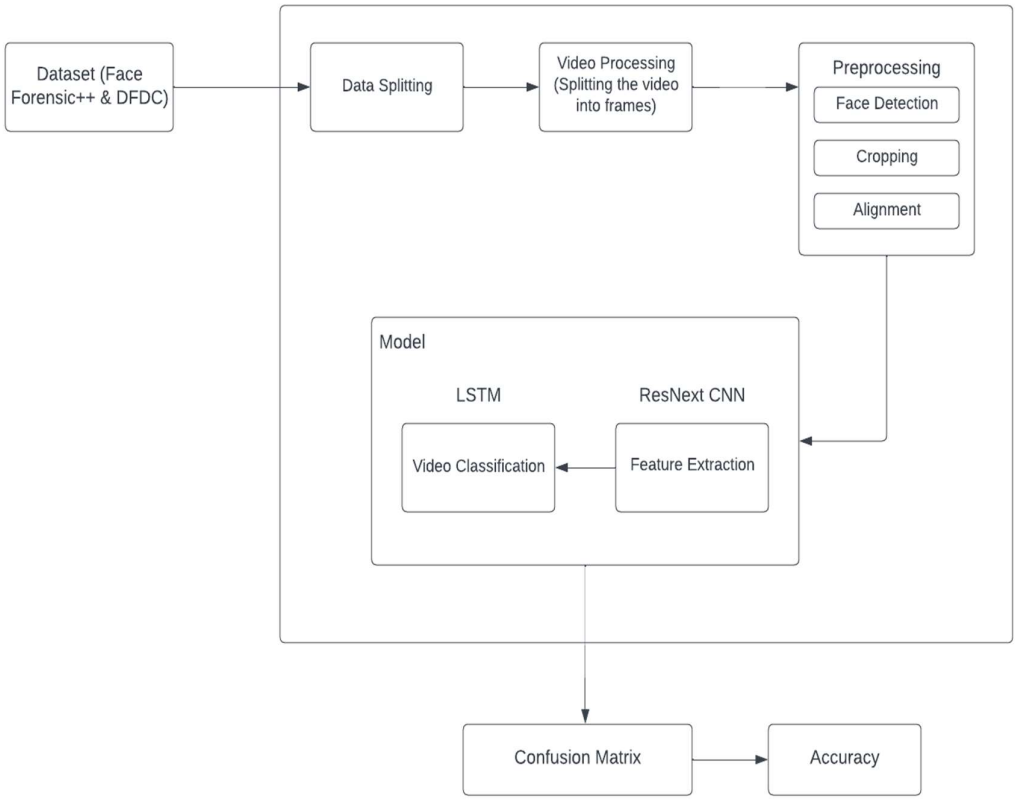

Fig 2: Proposed Model Architecture

6.2 Building Dataset

To ensure the efficiency of our model in real-time prediction, we curated a comprehensive dataset(as shown in Fig 3) from various available sources, including FF [1], DFDC [2], and Celeb-DF [3]. By combining these datasets and creating our own, we aimed to enable accurate and real-time detection across different types of videos. To address training bias, we maintained a balanced dataset with an equal distribution of 50% real and 50% fake videos.The DFDC dataset [2] contained certain videos with audio alterations, which were outside the scope of our research. To pre-process the DFDC dataset, we developed a Python script to remove the audio-altered videos from the dataset.

Following the pre-processing step, we selected 1500 real and 1500 fake videos from the DFDC dataset. Additionally, we included 1000 real and 1000 fake videos from the FF++ [1] dataset, as well as 500 real and 500 fake videos from the Celeb-DF [3] dataset. This compilation resulted in a total dataset of 3000 real videos, 3000 fake videos, and a combined total of 6000 videos. The distribution of these datasets is illustrated in Fig 3.

By leveraging this diverse and carefully curated dataset, we aimed to train our model effectively and ensure its robust performance in detecting deepfake videos in real-time scenarios.

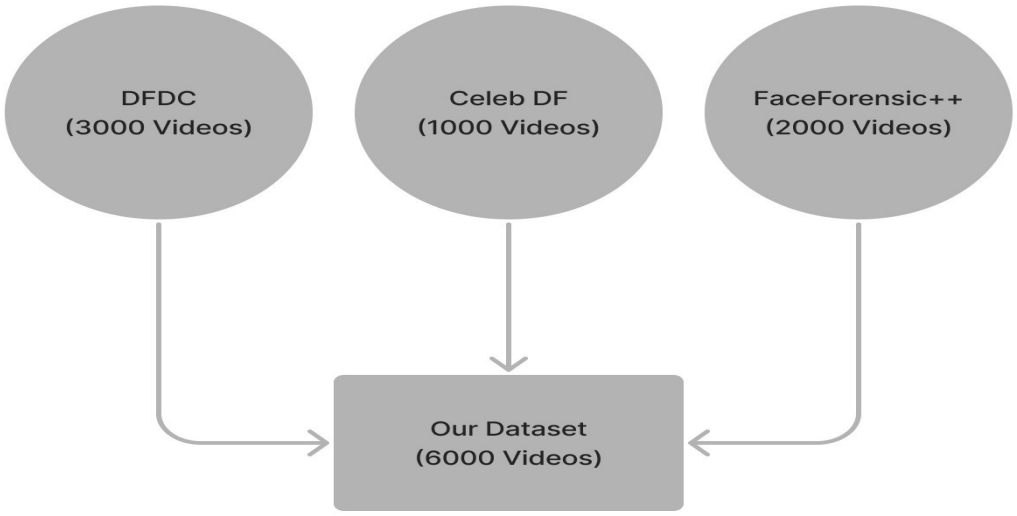

Fig 3: Dataset Gathering

6.3 Data Pre-Processing

In this crucial preprocessing step(as shown in Fig 4), we focus on refining the videos by eliminating any unnecessary elements and noise, while specifically isolating and extracting the facial region. The initial phase involves splitting the videos into individual frames. For each frame, we employ face detection algorithms to accurately identify and crop the facial area. Subsequently, the cropped frames are reassembled to reconstruct the video, ensuring that only the facial regions are preserved. During this preprocessing stage, frames that do not contain a detectable face are excluded.

In order to ensure uniformity in the quantity of frames throughout the dataset, measures were taken to maintain consistency, we establish a threshold value based on the mean total frame count of each video. we select a threshold of 200 frames. This decision takes into account the processing demands of handling a full video, as a 10-second video at a frame rate of 25 frames per second (fps) would amount to 300 frames. By limiting the frame count to 200, we effectively manage the computational resources of our Graphics Processing Unit (GPU). It is important to note that we retain the sequential order of the frames, selecting the initial 200 frames rather than choosing them randomly. The newly processed videos are saved with a frame rate of 25 fps and a resolution of 112 x 112 pixels.Through this meticulous preprocessing stage, we ensure the extraction of pertinent facial information while optimizing the dataset for subsequent analysis, specifically leveraging the capabilities of LSTM in our research.

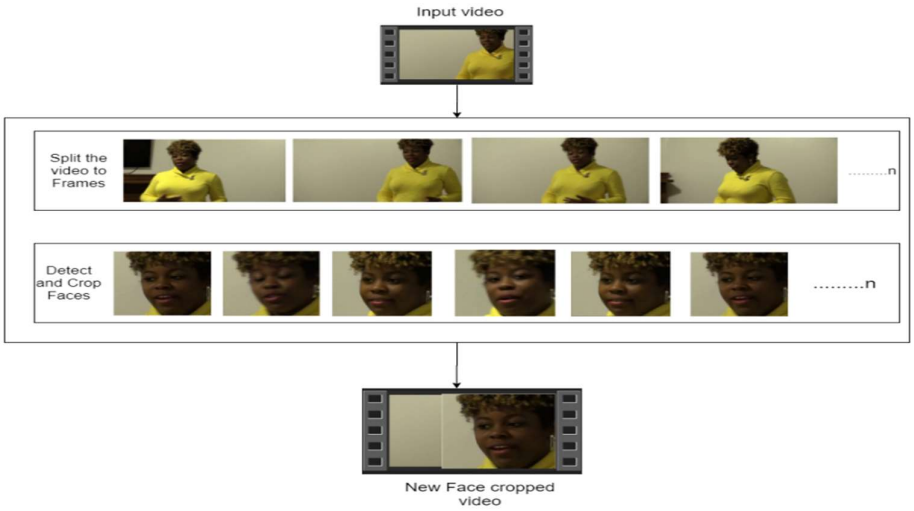

Fig 4: Pre-Processing of Video[29]

6.4 Data-set split

In order to facilitate effective model training and evaluation, the dataset is partitioned into separate train and test subsets, adhering to a 70% and 30% ratio, respectively(as shown in Fig 5). The training dataset consists of 4,200 videos, while the test dataset comprises 1,800 videos. It is worth noting that the train and test splits are carefully designed to maintain a balanced representation of both real and fake videos, with an equal distribution of 50% for each category in both subsets. This balanced split ensures that the model is trained and assessed on an equitable distribution of real and fake videos, enabling fair evaluation of its performance in distinguishing between the two classes. By adopting this balanced approach, we aim to mitigate potential biases and enhance the reliability and generalizability of our deepfake detection model. It is important to highlight that the train and test splits in this study are thoughtfully crafted to ensure a balanced representation of real and fake videos. The dataset is divided equally, with a 50% distribution for each category in both the training and testing subsets. This balanced split strategy guarantees that the model is trained and evaluated on an equitable distribution of real and fake videos, enabling a fair assessment of its performance in accurately discerning between these two classes.

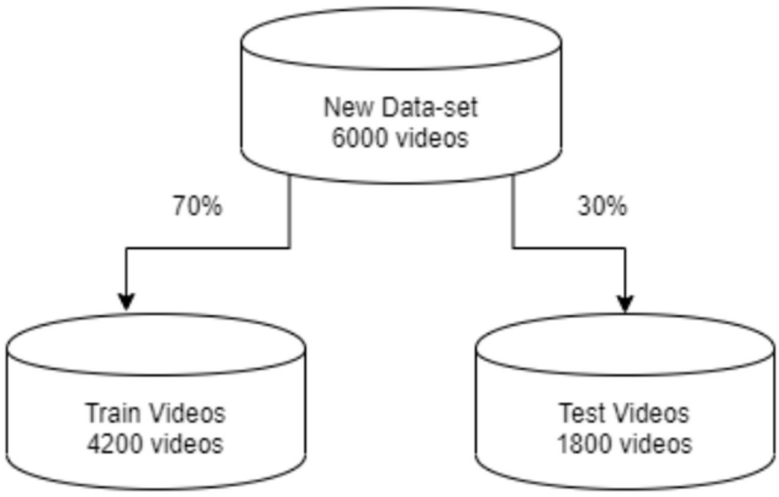

Fig 5: Data-set Split

6.5 DL Models

Our proposed model incorporates a powerful fusion of CNN and RNN architectures. To extract informative features at the frame level, we leverage the capabilities of a pre-trained ResNext CNN model. The ResNext CNN model allows us to effectively capture intricate patterns and discriminative characteristics present in the videos. These extracted features serve as input to a LSTM network, which is trained to perform the classification task of differentiating between deepfake and pristine videos.

During the training process, we employ a Data Loader to efficiently load and process the training split of videos. The labels associated with each video are loaded and seamlessly integrated into the model for training. This enables the model to learn the intricate relationships between the extracted frame-level features and the corresponding labels, enhancing its ability to accurately classify videos as either deepfake or pristine. By leveraging the combined power of CNN and RNN architectures and leveraging the training data effectively, our model demonstrates promising potential in tackling the challenges associated with deepfake detection.

6.5.1 ResNext CNN

To leverage the advantages of deep neural networks for feature extraction, we incorporated the ResNext model as a pivotal component of our framework. Rather than developing the feature extraction code from scratch, we employed a pre-trained ResNext model. ResNext is a Residual CNN network specifically optimized for achieving high performance on deeper neural networks. For our experimental purposes, we selected the resnext50_32x4d model, which consists of 50 layers and features dimensions of 32 x 4.

To tailor the ResNext model to our specific task of deepfake detection, we performed fine-tuning by introducing additional layers as needed. This allowed us to adapt the network to our particular requirements and enhance its ability to capture relevant features for distinguishing between deepfake and authentic videos. Moreover, we carefully selected an appropriate learning rate to facilitate the convergence of the gradient descent process and optimize the performance of the model.

The input to the subsequent sequential LSTM component consisted of 2048-dimensional feature vectors derived from the final pooling layers of the ResNext model. This sequential input allowed the LSTM network to effectively process the temporal dependencies present within the video frames and make informed predictions regarding the authenticity of the videos. By leveraging the powerful capabilities of the pre-trained ResNext model and fine-tuning it to our specific task, our framework demonstrates promising potential for accurate deepfake detection.

stage	output	ResNeXt-50 (32×4d)
1	112×112	7×7, 64, stride 2
2	56×56	3×3 max pool, stride 2 [1×1, 128 3×3, 128, C=32 1×1, 256] ×3
3	28×28	[1×1, 256 3×3, 256, C=32 1×1, 512] ×4
4	14×14	[1×1, 512 3×3, 512, C=32 1×1, 1024] ×6
5	7×7	[1×1, 1024 3×3, 1024, C=32 1×1, 2048] ×3
	1×1	global average pool 1000-d fc, softmax
# params.		$\mathbf{25.0}\times10^{6}$

Fig 6: ResNext Architecture

6.5.2 LSTM for Sequence Processing

To effectively process the 2048-dimensional feature vectors obtained from the ResNext model, we employed a Long Short-Term Memory (LSTM) layer(as shown in Fig 7), as a vital component of our framework. The LSTM layer plays a crucial role in sequential processing of the frames, enabling temporal analysis of the video by comparing the frame at time 't' with the frame at time 't-n', where 'n' represents any desired number of frames preceding 't'.

Our LSTM layer encompasses 2048 latent dimensions and 2048 hidden layers, effectively capturing and leveraging the underlying patterns and dependencies within the video data. To mitigate the risk of overfitting and improve generalization, we introduced a dropout regularization technique with a probability of 0.4, which aids in preventing the model from relying too heavily on specific features or frames.

To introduce non-linearity into the model, we incorporated the Leaky ReLU activation function. This activation function has proven effective in deep learning architectures, allowing the model to capture complex relationships between input and output features. A linear layer with 2048 input features and 2 output features was added to facilitate the learning of the average correlation between the input and output, thereby enabling the model to make accurate predictions.

To ensure adaptability and efficiency, we integrated an adaptive average pooling layer with an output parameter of 1, resulting in a target output size in the form of H x Y, H and Y denote the dimensions of the image, specifically referring to its height and width respectively, respectively. Sequential processing of the frames was achieved using a Sequential Layer, which enabled efficient handling and analysis of video data.

During training, we employed a batch size of 4, allowing for efficient batch processing and optimization of the model. Finally, to obtain the confidence of the model during prediction, we incorporated a SoftMax layer which provided a probability distribution over the classes, indicating the model's confidence in its predictions.

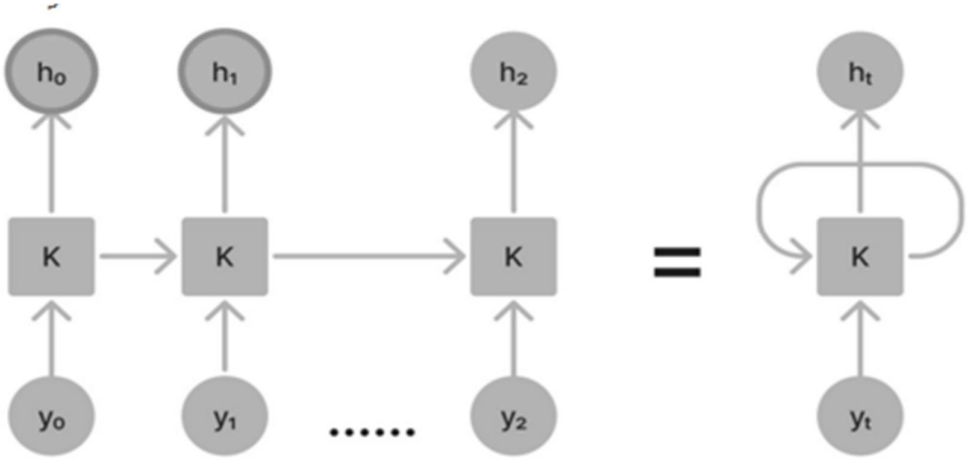

Fig 7: Overview of LSTM Architecture

6.5.3 Hyper Parameter Tuning

Hyperparameter tuning is a critical process in optimizing model performance by selecting the most suitable values for key parameters. Through iterative experimentation and evaluation, we have identified the optimal hyperparameters for our dataset. To facilitate adaptive learning, we employed the Adam optimizer [21] in conjunction with the model parameters. A learning rate of 1e-5 (0.00001) was determined to be ideal for achieving a better global minimum during gradient descent. Additionally, a weight decay of 1e-3 was utilized to regulate model complexity and prevent overfitting.As our objective involves classification, we employed the cross-entropy loss function, which is well-suited for such tasks. To ensure efficient utilization of available computational resources, batch training was employed, with a batch size of 4 being determined as the optimal choice in our development environment. This approach allows for simultaneous processing of multiple samples, improving training efficiency and convergence.

For the development of the user interface, we utilized the Django framework, known for its scalability and flexibility. The user interface, specifically the index.html page, features a tab that enables users to browse and upload videos. The uploaded video is then passed to the model, which performs the prediction. The model outputs whether the video is classified as real or fake, accompanied by the model's confidence level. These results are rendered on the predict.html page, displayed alongside the playing video for convenient visualization and interpretation.

6.5.4 ReLU

The ReLU is an activation function that outputs 0 if the input is negative, and the input itself if it is positive. This characteristic makes ReLU similar to the behavior of biological neurons. ReLU is a non-linear function and offers the advantage of not encountering any backpropagation errors, unlike the sigmoid function. Additionally, when constructing larger Neural Networks, ReLU allows for fast model building due to its speed. Fig 8 demotes the ReLU activation function working.

ReLU formula is : f(x) = max(0,x)

The ReLU function and its derivative exhibit monotonic characteristics. When provided with negative input, the function outputs 0, while positive values are returned as is. Consequently, the function yields an output range from 0 to infinity. ReLU serves as the widely preferred activation function in neural networks, particularly in Convolutional Neural Networks (CNNs), and is commonly employed as the default choice.

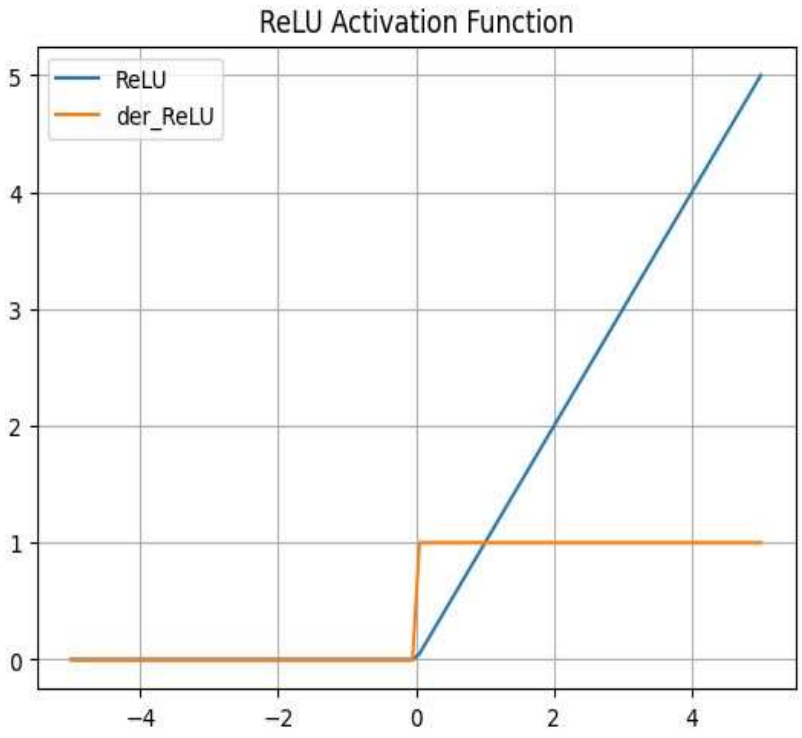

Fig 8: ReLU Activation Function

6.5.5 Dropout Layer

To prevent overfitting in the model and promote generalization, a dropout layer with a value of 0.41 is employed. This layer randomly sets the output of a neuron to 0, aiding in the regularization of the model. By setting the output to 0, the cost function becomes more responsive to the changes in neighbouring neurons. As a result, the weight updates during the backpropagation process are influenced, enhancing the model's learning capabilities. Basic diagram is depicted in Fig 9.

In machine learning, the dropout layer is a regularization technique commonly used in neural network architectures. Its purpose is to prevent overfitting by randomly deactivating a fraction of the neurons during each training iteration. This technique introduces a form of model uncertainty and forces the network to learn more robust representations that are less dependent on specific neurons. The dropout layer works by "dropping out" a specified percentage of neurons, typically chosen between 20% and 50%, during training. This means that the output of these dropped-out neurons is set to zero, effectively removing their contribution to the forward and backward propagation steps. By randomly disabling neurons, the dropout layer introduces a form of ensemble learning within a single network, as different subsets of neurons are activated or deactivated during each training iteration.

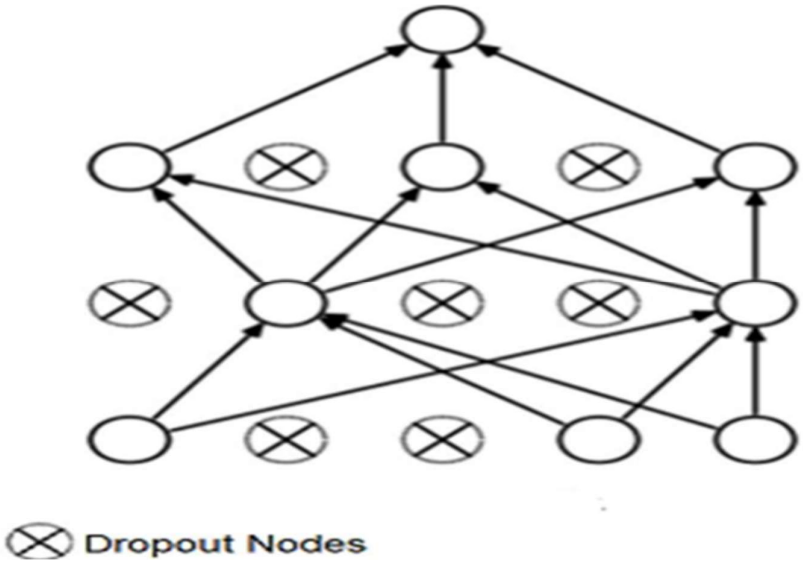

Fig 9: DropOut Layer

6.5.6 SOFTMAX LAYER AS SQUASHING FUNCTION

The Softmax function, a particular form of a squashing function, constrains the output to a range between 0 and 1. This characteristic enables the direct interpretation of the output as a probability. Softmax functions are particularly useful in multiclass classification scenarios, as they enable the determination of probabilities for multiple classes simultaneously. Due to the requirement of the outputs summing up to 1, a softmax layer is typically employed as the final layer in neural network architectures. It is worth noting that the softmax layer(as denoted in Fig 10) should have the same number of nodes as the output layer. In our case, the softmax layer consists of two output nodes representing "REAL" or "FAKE" classifications, and it provides us with the confidence or probability of a particular prediction. Softmax functions are particularly useful in multiclass classification scenarios, as they enable the determination of probabilities for multiple classes simultaneously.

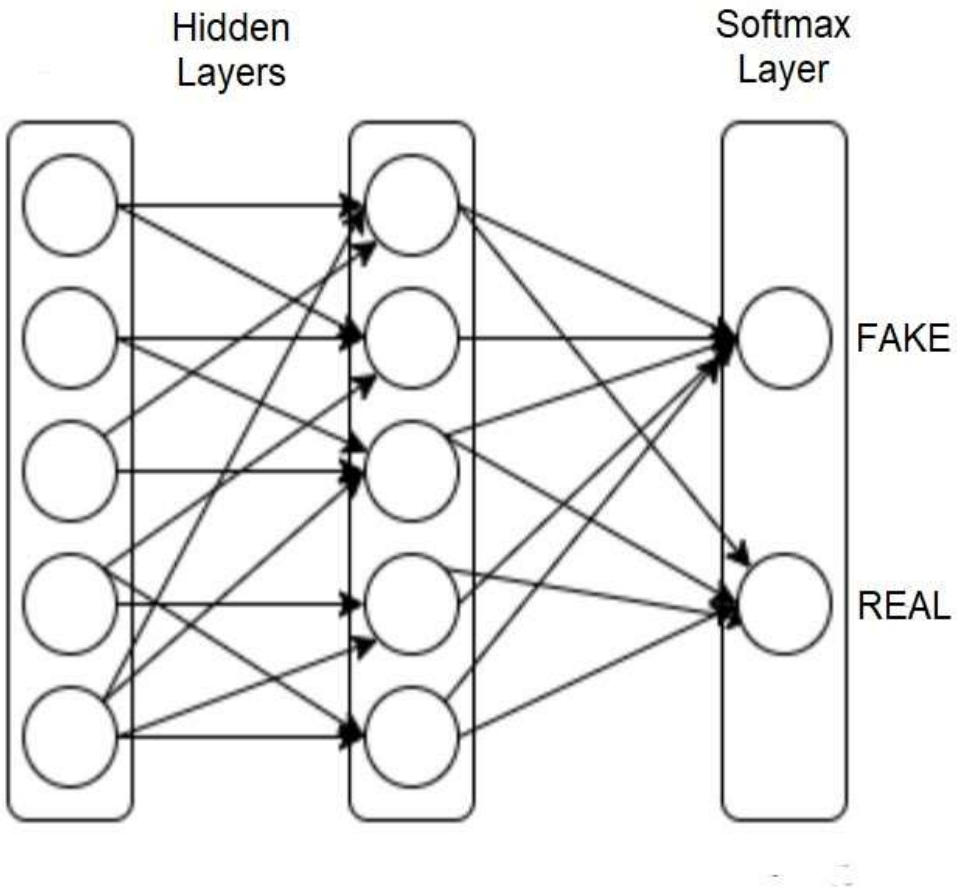

Fig 10: Softmax Layer

6.5.7 CONFUSION MATRIX FOR PREDICTION

A confusion matrix is a concise summary of classification prediction outcomes, capturing counts of correct and incorrect predictions for each class. It provides valuable insights into the classifier's errors, including the specific types of errors made. The matrix displays predicted and actual values, with "TN" indicating true negatives, "TP" denoting true positives, "FP" representing false positives, and "FN" standing for false negatives(as shown in Fig 10). Accuracy, a commonly used metric in classification tasks, can be derived from the confusion matrix.. It is calculated using the provided formula, which utilizes the confusion matrix to assess the model's performance.

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

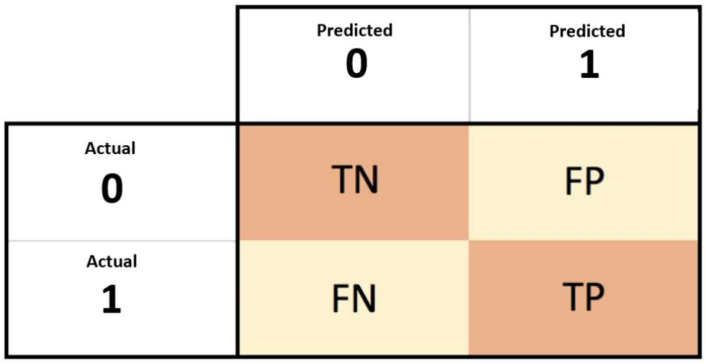

Fig 10: Confusion Matrix

CHAPTER 7

RESULTS

The proposed methodology for predicting whether a video is deepfaked or real by combining LSTM and CNN, along with a novel dataset created using the existing "FaceForensics++" and "DFDC" datasets, yielded promising results. The model achieved high accuracy in distinguishing between deepfake and real videos. Through the fusion of LSTM and CNN, the model effectively captured both spatial and temporal features, enabling robust classification. The comprehensive training on the augmented dataset enhanced the model's ability to generalize and detect various deepfake manipulation techniques. The evaluation metrics, including accuracy, precision, and recall, demonstrated the effectiveness of the proposed approach in combating the spread of disinformation and contributing to the field of video forensics.

Table 4: Results

DATASET	SEQUENCE LENGTH	NO OF VIDEOS	ACCURACY
Face Forensic++	25	2000	91.6285%
Face Forensic++	50	2000	95.7681%
Face Forensic++	75	2000	97.9234%
Face Forensic++	100	2000	98.0865%
Face Forensic++	125	2000	98.1265%
Face Forensic++	150	2000	98.3761%
Face Forensic++	175	2000	98.5428%
Face Forensic++	200	2000	98.7667%
DFDC	25	3000	84.6345%
DFDC	50	3000	86.7865%
DFDC	75	3000	87.9897%
DFDC	100	3000	89.5674%
DFDC	125	3000	92.3452%
DFDC	150	3000	93.6547%

DFDC	175	3000	95.0546%
DFDC	200	3000	96.3241%
Celeb DF	25	1000	87.6325%
Celeb DF	50	1000	88.6235%
Celeb DF	75	1000	90.6278%
Celeb DF	100	1000	91.9877%
Celeb DF	125	1000	93.6277%
Celeb DF	150	1000	94.8875%
Celeb DF	175	1000	96.3281%
Celeb DF	200	1000	97.8967%
Our Dataset	**25**	**6000**	**85.6784%**
Our Dataset	**50**	**6000**	**88.0768%**
Our Dataset	**75**	**6000**	**89.2857%**
Our Dataset	**100**	**6000**	**91.0827%**
Our Dataset	**125**	**6000**	**92.3467%**
Our Dataset	**150**	**6000**	**93.9763%**
Our Dataset	**175**	**6000**	**95.0328%**
Our Dataset	**200**	**6000**	**96.1276%**

Chapter 8

Conclusion and Future Scope

In conclusion, this thesis aimed to address the critical issue of detecting deepfake videos by leveraging a combination of LSTM and CNN architectures. The primary objective was to develop a novel dataset by amalgamating existing datasets, namely FaceForensics++, Celeb-DF, and DFDC (Deepfake Detection Challenge), to enable accurate and real-time detection of various types of deepfake videos.The research commenced with an extensive analysis of the problem statement and a thorough review of relevant literature, including research papers. Feasibility assessments were conducted, leading to the formulation of the project's goals and objectives. It was recognized that a balanced training approach was crucial to avoid bias and variance in the algorithm, thus ensuring accurate predictions. To achieve this, the dataset was carefully curated, with equal representation of real and fake videos, totaling 3000 videos each.

Preprocessing played a pivotal role in preparing the dataset for training. The videos underwent a series of steps, including frame splitting, face detection, and cropping, resulting in a processed dataset containing only the facial regions of interest. To maintain uniformity, a threshold value was determined based on the mean frame count, and only the first 150 frames were retained for each video. This enabled sequential processing of frames, crucial for temporal analysis. The model architecture consisted of a combination of a pre-trained ResNext CNN model for feature extraction and an LSTM network for classification. The 2048-dimensional feature vectors extracted from the ResNext model served as input for the LSTM, which processed the frames sequentially, enabling temporal analysis by comparing frames at different time intervals. Additional layers, including Leaky ReLU activation, linear layers, and adaptive average pooling, were incorporated to enhance the model's learning capacity and correlation between input and output.

In order to attain the best possible performance, a comprehensive process of hyperparameter tuning was undertaken. The Adam optimizer was utilized with a learning rate of 1e-5 and weight decay of 1e-3, ensuring improved convergence of the gradient descent and enhancing overall optimization efficacy.Cross-entropy loss calculation facilitated accurate classification, and batch training with a batch size of 4

maximized computational efficiency. The developed model was integrated into a user-friendly interface utilizing the Django framework. Users could upload videos for deepfake detection, and the model provided real-time predictions along with confidence scores. The output was rendered on the face of the playing video, enhancing interpretability and user engagement.

In conclusion, this research endeavors to mitigate the widespread dissemination of deepfake videos by proposing an effective and accurate detection methodology. The combination of LSTM and CNN, along with the novel dataset created from FaceForensics++, Celeb-DF, and DFDC datasets, showcases promising results in identifying deepfakes. Future work may involve further improving the model's performance, expanding the dataset, and addressing emerging challenges in the realm of deepfake detection.

FUTURE SCOPE

The research conducted in this thesis opens up several avenues for future exploration and advancements in the field of deepfake detection. Here are some potential areas of focus for further research:

- **Dataset Expansion:** While the novel dataset created by combining FaceForensics++, Celeb-DF, and DFDC datasets has shown promising results, there is room for expanding the dataset further. Including a more diverse range of videos, covering different contexts and scenarios, can enhance the model's robustness and generalizability.
- **Addressing New Deepfake Generation Techniques:** As technology evolves, so do the techniques used to create deepfake videos. Future research should stay vigilant and adapt the detection models to effectively identify deepfakes generated using emerging methods. Continuous monitoring and incorporation of new deepfake generation techniques into the training pipeline will be essential to stay ahead of potential threats.
- **Improving Model Performance:** Enhancing the accuracy and efficiency of the LSTM-CNN model should remain a priority. Exploring different model architectures, such as incorporating attention mechanisms or utilizing transformers, may lead to improved performance in deepfake detection. Additionally, exploring the effectiveness of transfer learning from other domains or leveraging advanced pre-trained models can also be fruitful.
- **Addressing Adversarial Attacks:** Deepfake creators may attempt to deceive detection models by incorporating adversarial attacks. Researching and developing robust defenses against such attacks is crucial to ensure the reliability and effectiveness of deepfake detection systems.
- **Real-Time Detection:** Enhancing the model's real-time prediction capabilities can significantly contribute to combating the rapid dissemination of deepfake videos on social media platforms. Exploring techniques for optimizing the model's efficiency without compromising accuracy will be valuable in practical deployment scenarios.

- **Collaboration and Benchmarking:** Collaboration among researchers, organizations, and institutions working in the field of deepfake detection is vital for progress. Establishing standardized benchmarks and evaluation metrics can facilitate fair comparison and drive innovation. Collaborative efforts can also foster the sharing of datasets, models, and techniques, ultimately advancing the field as a whole.

- Lastly a model can be designed which can able to detect the videos having the audio deepfakes and frames that does not contains the facial part of the body.

In conclusion, the thesis lays a solid foundation for deepfake detection using LSTM combined with CNN, utilizing a novel dataset created from existing datasets. The future scope lies in expanding the dataset, addressing new deepfake generation techniques, improving model performance, addressing adversarial attacks, enabling real-time detection, and fostering collaboration and benchmarking. These endeavors will contribute to the development of more robust and reliable deepfake detection systems to counter the growing threat of deceptive videos in various domains.

REFERENCES

[1] Li, Yuezun, and Siwei Lyu. "Exposing deepfake videos by detecting face warping artifacts." arXiv preprint arXiv:1811.00656 (2018).

[2] Li, Yuezun, Ming-Ching Chang, and Siwei Lyu. "In ictu culi: Exposing ai created fake videos by detecting eye blinking." In 2018 IEEE international workshop on information forensics and security (WIFS), pp. 1-7. IEEE, 2018.

[3] Nguyen, Huy H., Junichi Yamagishi, and Isao Echizen. "Capsule-forensics: Using capsule networks to detect forged images and videos." In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2307-2311. IEEE, 2019.

[4] Ciftci, Umur Aybars, Ilke Demir, and Lijun Yin. "Fakecatcher: Detection of synthetic portrait videos using biological signals." IEEE transactions on pattern analysis and machine intelligence (2020).

[5] Ciftci, Umur Aybars, Ilke Demir, and Lijun Yin. "Fakecatcher: Detection of synthetic portrait videos using biological signals." IEEE transactions on pattern analysis and machine intelligence (2020).

[6] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial networks." Communications of the ACM 63, no. 11 (2020): 139-144.

[7] Güera, David, and Edward J. Delp. "Deepfake video detection using recurrent neural networks." In 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS), pp. 1-6. IEEE, 2018.

[8] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition. CVPR. 2016." arXiv preprint arXiv:1512.03385 (2016).

[9] An Overview of ResNet and its Variants : https://towardsdatascience.com/an-overview-of-resnet andits-variants-5281e2f56035

[10]Long Short-Term Memory: From Zero to Hero with Pytorch: https://blog.floydhub.com/long-short-term memoryfrom-zero-to-hero-with-pytorch/

[11]Sequence Models And LSTM Networks https://pytorch.org/tutorials/beginner/nlp/sequence_models_tutorial.html

[12]https://discuss.pytorch.org/t/confusedabout-the-image preprocessing-in-classification/3965

[13]https://www.kaggle.com/c/deepfakedetection challenge/data

[14]https://github.com/ondyari/FaceForensics

[15] Güera, David, and Edward J. Delp. "Deepfake video detection using recurrent neural networks." In 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS), pp. 1-6. IEEE, 2018.

[16] Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A.Efros. "Image-to-image translation with conditional adversarial networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125-1134. 2017.

[17] Raja, Kiran, Sushma Venkatesh, and R. B. Christoph Busch. "Transferable deep-cnn features for detecting digital and print-scanned morphed face images." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 10-18. 2017.

[18] de Freitas Pereira, Tiago, André Anjos, José Mario De Martino, and Sébastien Marcel. "Can face anti-spoofing countermeasures work in a real world scenario?." In 2013 international conference on biometrics (ICB), pp. 1-8. IEEE, 2013.

[19] Rahmouni, Nicolas, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. "Distinguishing computer graphics from natural images using convolution neural networks." In 2017 IEEE workshop on information forensics and security (WIFS), pp. 1-6. IEEE, 2017.

[20] Song, Fengyi, Xiaoyang Tan, Xue Liu, and Songcan Chen. "Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients." Pattern Recognition 47, no. 9 (2014): 2825-2838.

[21] King, Davis E. "Dlib-ml: A machine learning toolkit." The

Journal of Machine Learning Research 10 (2009): 1755-1758

[22] Verdoliva, Luisa. "Media forensics and deepfakes: an overview." *IEEE Journal of Selected Topics in Signal Processing* 14, no. 5 (2020): 910-932.

[23] Masood, Momina, Mariam Nawaz, Khalid Mahmood Malik, Ali Javed, Aun Irtaza, and Hafiz Malik. "Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward." *Applied Intelligence* 53, no. 4 (2023): 3974-4026.

[24] Li, Yuezun, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu. "Celeb-df: A large-scale challenging dataset for deepfake forensics." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3207-3216. 2020.

[25] Agarwal, Shruti, Hany Farid, Yuming Gu, Mingming He, Koki Nagano, and Hao Li. "Protecting World Leaders Against Deep Fakes." In *CVPR workshops*, vol. 1, p. 38. 2019.

[26] Matern, Falko, Christian Riess, and Marc Stamminger. "Exploiting visual artifacts to expose deepfakes and face manipulations." In *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, pp. 83-92. IEEE, 2019.

[27] Zhang, Xu, Svebor Karaman, and Shih-Fu Chang. "Detecting and simulating artifacts in gan fake images." In *2019 IEEE international workshop on information forensics and security (WIFS)*, pp. 1-6. IEEE, 2019.

[28] Hasan, Haya R., and Khaled Salah. "Combating deepfake videos using blockchain and smart contracts." *Ieee Access* 7 (2019): 41596-41606.

[29] Chai, Lucy, David Bau, Ser-Nam Lim, and Phillip Isola. "What makes fake images detectable? understanding properties that generalize." In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVI 16*, pp. 103-120. Springer International Publishing, 2020.

[30] Verdoliva, Luisa. "Media forensics and deepfakes: an overview." *IEEE Journal of Selected Topics in Signal Processing* 14, no. 5 (2020): 910-932.

[31] Karniadakis, George Em, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. "Physics-informed machine learning." Nature Reviews Physics 3, no. 6 (2021): 422-440.

[32] Khan, Salman, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, and Mubarak Shah. "Transformers in vision: A survey." *ACM computing surveys (CSUR)* 54, no. 10s (2022): 1-41.

[33] Ramesh, Aditya, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. "Zero-shot text-to-image generation." In *International Conference on Machine Learning*, pp. 8821-8831. PMLR, 2021.

[34] Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. "High-resolution image synthesis with latent diffusion models." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684-10695. 2022.

[35] Grill, Jean-Bastien, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch et al. "Bootstrap your own latent-a new approach to self-supervised learning." *Advances in neural information processing systems* 33 (2020): 21271-21284.

[36] Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein et al. "On the opportunities and risks of foundation models." *arXiv preprint arXiv:2108.07258* (2021).

[37] Chen, Ting, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E. Hinton. "Big self-supervised models are strong semi-supervised learners." *Advances in neural information processing systems* 33 (2020): 22243-22255.

[38] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in Neural Information Processing Systems* 33 (2020): 6840-6851.

[39] Sarker, Iqbal H. "Machine learning: Algorithms, real-world applications and research directions." *SN computer science* 2, no. 3 (2021): 160.

[40] Zhuang, Fuzhen, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. "A comprehensive survey on transfer learning." *Proceedings of the IEEE* 109, no. 1 (2020): 43-76.

[41] Shorten, Connor, and Taghi M. Khoshgoftaar. "A survey on image data augmentation for deep learning." *Journal of big data* 6, no. 1 (2019): 1-48.

[42] Clark, Kevin, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. "Electra: Pre-training text encoders as discriminators rather than generators." *arXiv preprint arXiv:2003.10555* (2020).

[43] Song, Yang, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. "Score-based generative modeling through stochastic differential equations." *arXiv preprint arXiv:2011.13456* (2020).

[44] Karras, Tero, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401-4410. 2019.

LIST OF PUBLICATION

1. DeepFake Detection: Unmasking the Real from the Fake in Videos with LSTM-CNN Fusion and a Novel Dataset accepted in 5th IEEE International Conference on Advances in Computing, Communication Control and Networking- ICAC3N

2. Unveiling Deepfake: A Novel LSTM-CNN Approach for Video Authenticity Assessment with Multi-Dataset Integration accepted in 5th IEEE International Conference on Advances in Computing, Communication Control and Networking- ICAC3N

8% Overall Similarity

Top sources found in the following databases:

- 3% Internet database
- Crossref database
- 7% Submitted Works database
- 1% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	**OP Jindal University, Raigarh on 2022-12-28** Submitted works	2%
2	**Bournemouth University on 2022-05-12** Submitted works	2%
3	**ijsrd.com** Internet	<1%
4	**coursehero.com** Internet	<1%
5	**University of Portsmouth on 2019-09-10** Submitted works	<1%
6	**au-e.com** Internet	<1%
7	**biorxiv.org** Internet	<1%
8	**jmest.org** Internet	<1%

LIST OF PUBLICATION

1. DeepFake Detection: Unmasking the Real from the Fake in Videos with LSTM-CNN Fusion and a Novel Dataset accepted in 5th IEEE International Conference on Advances in Computing, Communication Control and Networking- ICAC3N

2. Unveiling Deepfake: A Novel LSTM-CNN Approach for Video Authenticity Assessment with Multi-Dataset Integration accepted in 5th IEEE International Conference on Advances in Computing, Communication Control and Networking- ICAC3N

Gmail RAUNAK PODDAR <raunakpoddar1998@gmail.com>

Acceptance Notification 5th IEEE ICAC3N-23 & Registration: Paper ID 839

Microsoft CMT <email@msr-cmt.org> 29 May 2023 at 14:40
Reply-To: Vishnu Sharma <vishnu.sharma@galgotiacollege.edu>
To: Raunak Poddar <raunakpoddar1998@gmail.com>

Dear Raunak Poddar,
Delhi Technological University

Greetings from ICAC3N-23 ...!!!

Congratulations....!!!!!

On behalf of the 5th ICAC3N-23 Program Committee, we are delighted to inform you that the submission of “Paper ID- 839 “ titled “ DeepFake Detection: Unmasking the Real from the Fake in Videos with LSTM-CNN Fusion and a Novel Dataset “ has been accepted for presentation and further publication with IEEE at the ICAC3N- 23 subject to incorporate the reviewers and editors comments in your final paper. All accepted papers will be submitted to IEEE for inclusion into conference proceedings to be published on IEEE Xplore Digital Library.

For early registration benefit please complete your registration by clicking on the following Link:
https://forms.gle/8e6RzNbho7CphnYN7 on or before 05 June 2023.

Registration fee details are available @ https://icac3n.in/register.
https://drive.google.com/file/d/1RGu6i4eGUI5B07zOfgRmDRfyjOhOhiC6/view?usp=share_link

You can also pay the registration fee by the UPI. (UPI id - icac3n@ybl) or follow the link below for QR code:
https://drive.google.com/file/d/1Ry-sF0apvy_0zUjM8lNW02lZgWAsE0YD/view?usp=sharing

You must incorporate following comments in your final paper submitted at the time of registration for consideration of publication with IEEE:

Reviewers Comments:
The topic chosen "DeepFake Detection: Unmasking the Real from the Fake in Videos with LSTM-CNN Fusion and a Novel Dataset" is interesting and relevant.
Paper is well written and technically sound.
Results are well explained.

Editor Note:
1. All figures and equations in the paper must be clear. Equation and tables must be typed and should not be images.
2. Final camera ready copy must be strictly in IEEE format available on conference website www.icac3n.in.
3. Transfer of E-copyright to IEEE and Presenting paper in conference is compulsory for publication of paper in IEEE.
4. If plagiarism is found at any stage in your accepted paper, the registration will be cancelled and paper will be rejected and the authors will be responsible for any consequences. Plagiarism must be less then 20% (checked through Turnitin). However the author will be given fair and sufficient chance to reduce plagiarism.
5. Change in paper title, name of authors or affiliation of authors will not be allowed after registration of papers.
6. Violation of any of the above point may lead to rejection of your paper at any stage of publication.
7. Registration fee once paid will be non refundable.

If you have any query regarding registration process or face any problem in making online payment, you can Contact @ 8168268768 (Call) / 9467482983 (Whatsapp/UPI) or write us at icac3n23@gmail.com.

Regards:
Organizing committee
ICAC3N – 2023

Gmail RAUNAK PODDAR <raunakpoddar1998@gmail.com>

Acceptance Notification 5th IEEE ICAC3N-23 & Registration: Paper ID 959

Microsoft CMT <email@msr-cmt.org> 30 May 2023 at 11:31
Reply-To: Vishnu Sharma <vishnu.sharma@galgotiacollege.edu>
To: Raunak Poddar <raunakpoddar1998@gmail.com>

Dear Raunak Poddar,
Delhi Technological University

Greetings from ICAC3N-23 ...!!!

Congratulations....!!!!!

On behalf of the 5th ICAC3N-23 Program Committee, we are delighted to inform you that the submission of “Paper ID- 959 “ titled “ Unveiling Deepfake: A Novel LSTM-CNN Approach for Video Authenticity Assessment with Multi-Dataset Integration “ has been accepted for presentation and further publication with IEEE at the ICAC3N- 23 subject to incorporate the reviewers and editors comments in your final paper. All accepted papers will be submitted to IEEE for inclusion into conference proceedings to be published on IEEE Xplore Digital Library.

For early registration benefit please complete your registration by clicking on the following Link: https://forms.gle/8e6RzNbho7CphnYN7 on or before 05 June 2023.

Registration fee details are available @ https://icac3n.in/register.
https://drive.google.com/file/d/1RGu6i4eGUI5B07zOfgRmDRfyjOhOhiC6/view?usp=share_link

You can also pay the registration fee by the UPI. (UPI id - icac3n@ybl) or follow the link below for QR code:
https://drive.google.com/file/d/1Ry-sF0apvy_0zUjM8lNW02lZgWAsE0YD/view?usp=sharing

You must incorporate following comments in your final paper submitted at the time of registration for consideration of publication with IEEE:

Reviewers Comments:
The topic chosen "Unveiling Deepfake: A Novel LSTM-CNN Approach for Video Authenticity Assessment with Multi-Dataset Integration" is interesting and relevant.
Formatting must be strictly as per template.
References are not in proper format. Format and assign number to the references properly.
An overview of paper is desired to eradicate typo and grammatical error.
Formatting and Quality of figures is not good. Use High quality images. (fig. 1 specifically)
Table must be properly captioned and numbered as per IEEE conference template.

Editor Note:
1. All figures and equations in the paper must be clear. Equation and tables must be typed and should not be images.
2. Final camera ready copy must be strictly in IEEE format available on conference website www.icac3n.in.
3. Transfer of E-copyright to IEEE and Presenting paper in conference is compulsory for publication of paper in IEEE.
4. If plagiarism is found at any stage in your accepted paper, the registration will be cancelled and paper will be rejected and the authors will be responsible for any consequences. Plagiarism must be less then 20% (checked through Turnitin). However the author will be given fair and sufficient chance to reduce plagiarism.
5. Change in paper title, name of authors or affiliation of authors will not be allowed after registration of papers.
6. Violation of any of the above point may lead to rejection of your paper at any stage of publication.
7. Registration fee once paid will be non refundable.

If you have any query regarding registration process or face any problem in making online payment, you can Contact @ 8168268768 (Call) / 9467482983 (Whatsapp/UPI) or write us at icac3n23@gmail.com.

Regards:
Organizing committee
ICAC3N – 2023