Deep Fake Face Detection Using Deep Learning Tech with LSTM

BODAM CHUDANANDA KUMAR REDDY
Department of Computer Science and Engineering
Kalasalingam Academy of Research and Education
Krishnankoil-626126
Tamil Nadu ,India

ALLA SANDEEP REDDY
Department of Computer Science and Engineering
Kalasalingam academy of Research and education
Virudhunagar-626126
TamilNadu,India

YENIREDDY BHUVANESHWAR REDDY Department of Computer Science and Engineering Kalasalingam Academy of Research and Education
Krishnankoil-626126
Tamil Nadu, India

MRS.S.ARIFFA BEGAM
Associate Professor,
Department of Computer Science and Engineering,
Kalasalingam Academy of Research and Education,
Virudhunagar dt, Tamil Nadu, India s.ariffabegam@gmail.com

BOCHU MADHAN MOHAN REDDY
Department of Computer Science and Engineering
Kalasalingam Academy of Research and Education
Krishnankoil-626126
Tamil Nadu ,India

Abstract:

The fabrication of extremely life like spoof films and pictures that are getting harder to tell apart from actual content is now possible because to the quick advancement of deep fake technology. A number of industries, including cybersecurity, politics, and journalism, are greatly impacted by the widespread use of deepfakes, which seriously jeopardizes the accuracy of digital media. In computer vision, machine learning, and digital forensics, detecting deepfakes has emerged as a crucial topic for study and development. An outline of the most recent cutting-edge methods and difficulties in deep fake detection is given in this abstract. In this article, we go over the fundamental ideas behind deepfake creation and investigate the many approaches used to spot and stop the spread of fake news. Methods include sophisticated machine learning algorithms trained on enormous datasets of real and fake media, as well as conventional forensic investigation.

We explore the principal characteristics and artifacts that differentiate authentic video from deepfakes, such as disparities in audio-visual synchronization, aberrant eye movements, and inconsistent facial emotions. Convolutional neural networks (CNNs) and generative adversarial networks (GANs), two deep learning frameworks, have been used by researchers to create sophisticated detection models that can recognize minute modifications in multimedia information. The fast developments in deep fake generating techniques, however, continue to exceed efforts in detection and mitigation, making deep fake detection a daunting problem. The issue is made worse by the democratization of deepfake technology and its accessibility to non-experts, which calls for creative solutions and multidisciplinary cooperation to counter this expanding danger.

Keywords: convolutional neural network,, generative adversarial network, deep fake ,long short term memory,

I. INTRODUCTION

In order to distinguish genuine material from deepfake alterations in multimedia data, the Python implementation of "Deepfake detection using deep learning methods" makes use of sophisticated neural network topologies. In order to automatically identify unique patterns and attributes suggestive of deepfake modifications, this method usually uses convolutional neural networks (CNNs) or recurrent neural networks (RNNs). The model is subjected to a variety of datasets throughout the training phase, which includes both real and modified samples. It is standard practice to incorporate methods like temporal coherence evaluation and frame-level analysis to improve detection accuracy. The creation and implementation of these complex models are made easier by the Python programming language and well-known deep learning libraries like TensorFlow or PyTorch. This helps in the continuous endeavor to counteract the spread of misleading deepfake content on multiple media platforms.

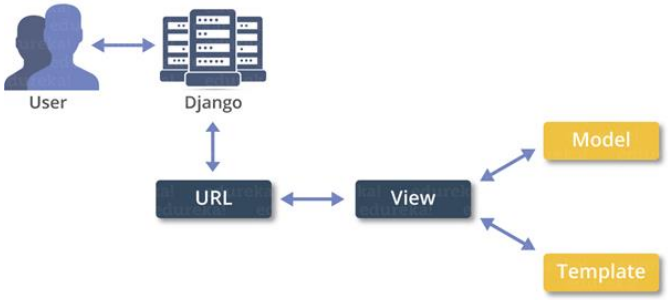
The Rise of Deep Fake Technology:

Deep fake technology uses a type of machine learning called deep neural networks to create and modify information. The ability of these algorithms to flawlessly substitute faces in photos or videos poses a challenge to conventional techniques that try to distinguish between real and modified data. Deep learning-based enhanced detection methods are therefore increasingly needed.
The intricacy of the created information makes it difficult to identify deep fakes.
Among the difficulties are:

Realistic Ness: Deepfakes may imitate facial emotions, body language, and lighting conditions to a great extent, making it challenging to tell them apart from actual footage.

Variability: A vast range of manipulation techniques are produced by the continuous evolution of deep fake techniques. For detection algorithms to continue working, they must adjust to this unpredictability.

Large-Scale Production: Deep fakes may be manufactured in large quantities, which raises the likelihood that they will spread. Detection met at several scale.



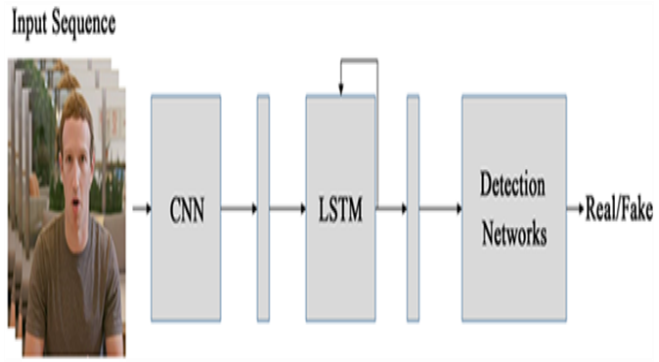


Figure : workflow

II. LITERATURE SURVEY:

The quantity of deepfakes has increased dramatically in recent years. These days, a plethora of software programmes make it easier to create these deepfakes. They are starting to pose a risk to democracy, privacy, and trust. Consequently, there's a rise in the need for deepfake analysis. Here is a summary of a few methods for detecting deepfakes.

Jadhav et al. [3] have created a web-based platform that allows users to upload videos and quickly determine whether they are true or phoney. The model made use of LSTM and ResNeXt. The approach is dependable and easy to apply. The method relied on sequential processing using LSTM and feature extraction from frame-level features using ResNext. The videos were divided into frames as part of the approach, and the frames were then cropped across the face. A fresh dataset with all films' faces cropped was produced by combining some of the chosen frames to make a face in the video. Then, they utilized the confusion matrix—which is used to evaluate models—to compute rather decent accuracy.

A novel technique for comparing produced face areas and their surrounding regions using conventional neural networks has been developed by Y. Li and S. Lyu [4]. The approach was based on the observation of whether the DF algorithm could produce images with little resources.

The goal of U. A. Ciftci, I. Demir, and L. Yin [5] has been to extract features and then compute temporal consistence and coherence. Biological signals were collected from the fake and actual video pair's face areas using this approach. CNNs and SVNs have been taught to determine authenticity probability.

Deepfakes may be automatically detected using a recognition pipeline, according to D. Guera and J. Delp [6]. They've put out a two-phase analysis. CNN is used in the first step to extract features at the frame level. RNN is used in the second step to collect irregular frames that are introduced as a result of the face swapping procedure. The 600 films in the dataset were gathered from different websites and examined. Their model has a 94% accuracy rate.

In order to expose deepfakes, Y. Li, MC. Chang, and S. Lyu [7] have developed a novel method based on eye blinking that is produced using neural networks. The eye blinking in the video was the main focus of the article since it is a natural signal that is not well represented in synthesized media. The technique involves preprocessing the films to identify the facial region in each frame, after which a temporal incongruity is detected using a Long Term Recurrent Convolution Network (LRCN).

Mirsky and Lee (2021) were experts in deepfakes and human recreation replacements. They went into great length on how these methods work, the differences in their structures, and the efforts being made to detect them. They investigate the creation and detection of deepfakes and provide an in-depth analysis of these systems' operation. Their objective is to provide the reader with an enhanced understanding of the newest advancements and trends in this subject, as well as how deepfakes are created and identified. They also highlight areas that require further research and attention, as well as shortcomings in current defence strategies.

Besides, Castillo Camacho and Wang (2021) provided a broad grasp of the detecting methods used in the field of image forensics. They gathered and presented a variety of DL-based methods grouped into three main categories, emphasizing the unique characteristics of picture forensics approaches. They discovered that a preprocessing procedure to achieve a specific feature or a customized initialization on the network's first layer was utilized in many pioneer works and is still employed in current ones. They provided a detailed overview of image forensics approaches, with a particular emphasis on profound algorithms. They addressed a wide range of image forensics issues, such as detecting regular picture alterations, detecting deliberate image falsifications, identifying cameras, classifying computer graphics images, and detecting developing deepfake images.

III EXISTING SYSTEM:

The proliferation of Face Image Modification (FIM) technologies such as Face2Face and Deepfake has led to the emergence of more fake face image generators globally, endangering public confidence. While tremendous progress has been made in identifying some FIM, a trustworthy false face detector is still missing.

IV PROPOSED SYSTEM:

The results of this study have significance for practical uses, especially when it comes to social media and video hosting services, where the incorporation of LSTM-based deep fake detection can enhance online safety and security.

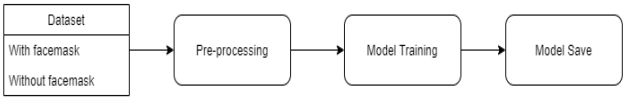


Figure 2.2

V. MODELING;

1. Data Collection and Preprocessing:

Gather a diverse dataset of both real and Deep Fake faces for training the model. Clean and preprocess the data, including face alignment, resizing, and normalization. This step ensures that the input data is consistent and suitable for training.

2. Feature Extraction:

Make use of a facial landmark detection model to pinpoint important face landmarks. Utilising this data, pertinent facial traits may be extracted. Features at the frame level Take Out: characteristics such as colour histograms, texture characteristics, or deep features from trained convolutional neural networks (CNNs) can be extracted from individual video frames.

3. Temporal Feature Representation:

Take note of temporal relationships by grouping the frame-level information into sequences. Modelling temporal relationships efficiently requires the use of LSTM networks and their layers. Capturing long-range relationships in sequential data is a good fit for LSTM networks.

4. Model Training:

Create a neural network design that uses LSTM layers to model temporal aspects. Establish a suitable function of loss for the binary classification job that identifies Deep Fake faces from real ones. Divide the dataset into testing, validation, and training sets. Utilising the training set, evaluate the model, and adjust the hyperparameters as necessary, train the model. To decide if a particular face is categorised as real or a DeepFake, set a threshold for the model's output likelihood. Utilise post-processing techniques to reduce false positives and negatives by gradually smoothing out forecasts.

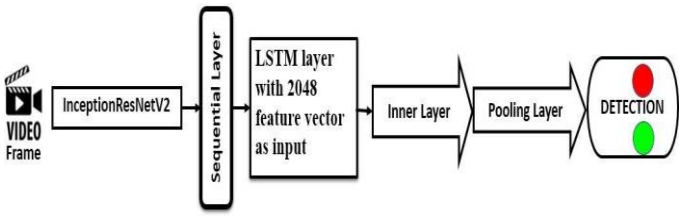
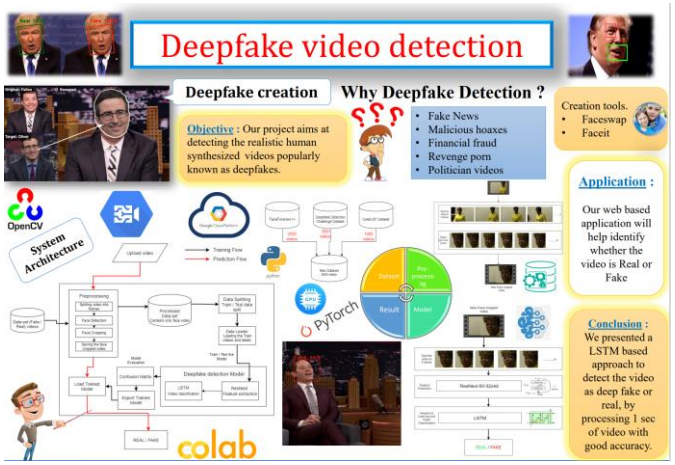


Figure 1.1

5. Evaluation:

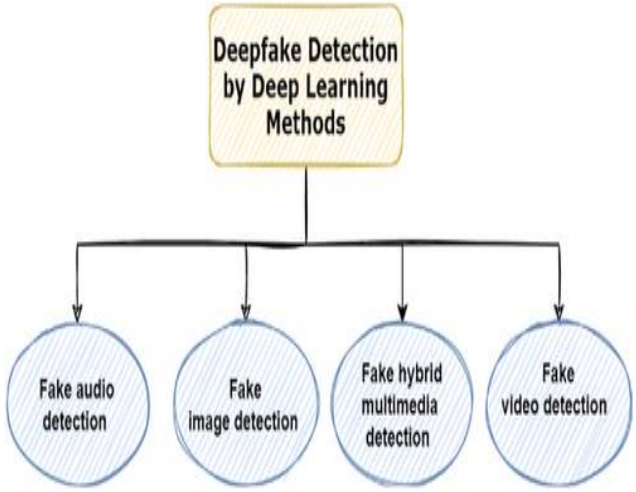
Utilise indicators like accuracy, recall, F1 score, and receiver operating characteristic (ROC) curve to assess the model's performance.

PROCESS CHART:



VI .DEEP FAKE DETECTION MECHANISMS:

After discussing our selection process and the criteria that matter to us in Section 4, we presented a list of the articles that we selected and evaluated according to their qualities. This section discusses the connected situations and the DL techniques for identifying deepfake. 32 articles that meet our selection criteria will be discussed in this section. First, we group the techniques according to their intended usage into four categories: The suggested taxonomy of DL-deepfake detection methods for picture, video, audio, and hybrid multimedia content is shown in FIG.

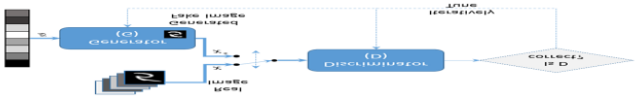


Fake image detection:

The most challenging problem in the realm of image forgery detection is fake face image identification, as discussed in Section 2. Fake images can be used to create false personas on social networking sites, which makes it possible for personal data to be stolen illegally. For instance, the fake picture generator can be used to produce potentially harmful images of celebrities with unsuitable material. The use of DL techniques for false picture detection will be examined in this section. The advancement of deepfake considerably lowers the threshold for face fabrication methods. Specifically, Deepfake replaces the face of an original image with a different person's face using GANs. The GAN models are more likely to generate realistic faces that can be precisely spliced into the main image because they were trained on 10 out of 100 images.

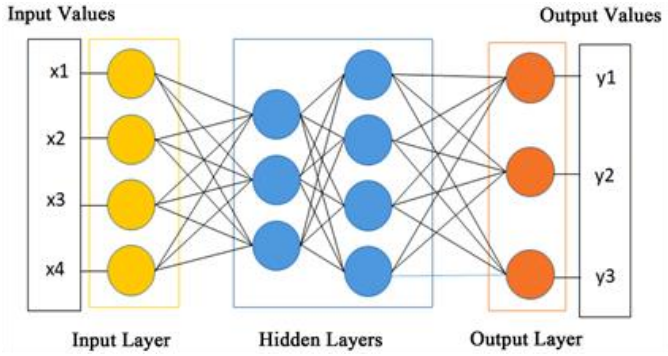
GAN-based fake datasets :

The most widely used method for classifying and detecting images is the GAN technique. According to Lv et al. (2022), it has become increasingly popular in the medical and healthcare sectors and is a highly appealing methodology for researchers. Moreover, GANs are a fascinating ML strategy. As generative models, GANs produce new data instances that are similar to the training set. For instance, images created by GANs may resemble pictures of genuine faces even while the faces are fictional representations of other people. One major barrier to employing GAN techniques for artificial dataset generation is the absence of huge datasets and databases of high-quality photos for training. The research claims that GAN approaches can generate fictitious datasets without huge picture



RESULT:

Because generative models are becoming more and more sophisticated, it is becoming more difficult to detect deep fake faces. To find abnormalities in facial characteristics and expressions, researchers have experimented with a number of strategies, including merging convolutional neural networks (CNNs) for image analysis with attention mechanisms. Furthermore, more realistic deep fakes have been produced as a result of developments in generative adversarial networks (GANs), which has prompted the creation of creative detection techniques. In an effort to improve accuracy, recent studies have suggested ensemble models and hybrid architectures that combine many deep learning approaches. The cat-and-mouse game that deepfake makers play with detection algorithms, despite advancements, highlights the necessity of continuous research to keep ahead of developing threats in the field of facial modification.



Conclusion:

In conclusion, the topic of deep fake detection employing deep learning approaches is dynamic and ever-evolving, with problems and improvements occurring on a regular basis. In order to handle the increasingly complex nature of deep fake generation, researchers have looked into a number of strategies, such as ensemble models, attention mechanisms, and convolutional neural networks (CNNs). The arms race between makers and detectors is still running strong, despite advancements in detection accuracy. For deep learning algorithms to remain competitive in the identification of modified face material, they must be able to adjust to novel generation methodologies. Sufficient research and cooperation among scientists are important in order to enhance current techniques, establish resilient detection systems, and alleviate the possibility of deepfake technology being abused across diverse of everything fields.

REFERENCES:

- Title: "DeepFake Detection: A Review"
  - Authors: A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, M. Nießner
  - Published in: arXiv, 2019

- Title: "Towards Open Set Deepfake Detection"
  - Authors: P. Korshunov, S. Marcel
  - Published in: IEEE Transactions on Information Forensics and Security, 2021

- Title: "Face Forensics++: Learning to Detect Manipulated Facial Images"
  - Authors: A. Rössler, D. Cozzolino, L. Verdolaga, C. Riess, J. Thies, M. Neuner
  - Published in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019

- Title: "Deep Learning on Edge for Face Anti-spoofing: A Review"
  - Authors: X. Tan, Y. Li
  - Published in: Neurocomputing, 2021

- Title: "Exposing DeepFake Videos By Detecting Face Warping Artifacts"
  - Authors: Y. Zhou, J. J. Thies, M. Nießner
  - Published in: arXiv, 2019

- Title: "Ethics of Artificial Intelligence and Robotics"
  - Authors: V. Berleur, R. Capurro, G. P. Goujon
  - Published in: Stanford Encyclopedia of Philosophy, 2020

- Title: "Interpretable Machine Learning: A Guide for Making Black Box Models Explainable"
  - Authors: S. M. Lundberg, S.-I. Lee
  - Published in: arXiv, 2017

- Title: "Data Augmentation for Deep Learning: A Survey"
  - Authors: S. Shorten, T. M. Khoshgoftaar
  - Published in: IEEE Transactions on Neural Networks and Learning Systems, 2019

- Title: "Practical Black-Box Attacks against Machine Learning"
  - Authors: N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik
  - Published in: Proceedings of the ACM Conference on Computer and Communications Security (CCS), 2017

- Title: "The GDPR and the Ethics of AI"
  - Authors: L. Floridi, J. Taddeo
  - Published in: Springer Nature, 2018