DEEPFAKE VIDEO DETECTION USING NEURAL NETWORK

Pranav Bire [1], Om Ambalkar [2], Ritesh Bagade [3]
Pradnya Mehta [4], Anuradha Yenkikar [5]
Computer Science Engineering (Artificial Intelligence)
Vishwakarma Institute of Information Technology, Pune
pranav.22210192@viit.ac.in [1], *om.22210200@viit.ac.in* [2], *ritesh.22210962@viit.ac.in* [3], *pradnya.mehta@viit.ac.in* [3], *anuradha.yenkikar@viit.ac.in* [4]

Abstract **-** In recent months, free deep learning-based software tools have facilitated the creation of indistinguishable human synthesized video popularly called as deep fakes. Manipulations of digital videos have been demonstrated for several decades through the good use of visual effects, recent advances in deep learning have led to a drastic increase in the realism of fake content and the accessibility in which it can be Scenarios where this realistic face swapped deep fakes are used to create political distress, fake terrorism events, blackmail peoples are easily envisioned. In the proposed project describes a new deep learning-based method that can effectively distinguish AI-generated fake videos from real videos. The proposed model refers to the use Artificial Intelligence (AI) to fight Artificial Intelligence (AI). Proposed system uses a Res-Next Convolution neural network to extract the frame-level features and these features and further used to train the Long Short-Term Memory (LSTM) based Recurrent Neural Network (RNN) to classify whether the video is subject to any kind of manipulation or not, i.e. whether the video is deep fake or real video. To emulate the real time scenarios and make the model perform better on real time data, The model is evaluated on large amount of balanced and mixed dataset prepared by mixing the various available dataset like Deepfake detection challenge, and Celeb-DF.

Keywords: Deepfake Video Detection, Res-Next Convolution neural network, Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM).

I. INTRODUCTION

The increasing sophistication of smartphone cameras and the availability of good internet connection all over the world has increased the ever-growing reach of social media and media sharing portals have made the creation and transmission of digital videos more easy than ever before. The growing computational power has made deep learning so powerful that would have been thought impossible only a handful of years ago. Like any transformative technology, this has created new challenges. Deep fake is a technique for human image synthesis based on neural network tools like GAN (Generative Adversarial Network) or Auto Encoders etc. These tools super impose target images onto source videos using a deep learning techniques and create a realistic looking deep fake video. These deep-fake video are so real that it becomes impossible to spot difference by the naked eyes. Spreading of the deepfake over the social media platforms have become very common leading to spamming and peculating wrong information over the platform. These types of the deepfake will be terrible, and lead to threating, misleading of common people.

To overcome such a situation, deepfake detection is very important. So, the proposed model describes a new deep learning-based method that can effectively distinguish AI-generated fake videos (DF Videos) from real videos. It's incredibly important to develop technology that can spot fakes, so that the deepfake can be identified and prevented from spreading over the internet.
The GAN split the video into frames and replaces the input image in every frame. Further it reconstructs the video. This process is usually achieved by using autoencoders. During the creation of the deep fake the current deep fake creation tools leaves some distinguishable artifacts in the frames which may not be visible to the human being, but the trained neural networks can spot the changes. The deepfake algorithm can only synthesize face images of a fixed size, and they must undergo an affinal warping to match the configuration of the source's face. This warping leaves some distinguishable artifacts in the output deepfake video due to the resolution inconsistency between warped face area and surrounding context. The model demonstrates that these artifacts can be effectively captured by Res-Next Convolution Neural Networks. Our system uses a Res-Next Convolution Neural Networks to extract frame-level features These features are then used to train a Long Short-Term Memory (LSTM) based Recurrent Neural Network (RNN) to classify whether the video is subject to any kind of manipulation or not, i.e. whether the video is deep fake or real video. To achieve this, the proposed model is trained on combination of available datasets. So that model can learn the features from different kind of images. Model extracted an adequate number of videos from Deepfake detection challenge dataset [1].

II. LITERATURE SURVEY

The explosive growth in deep fake video and its illegal use is a major threat to democracy, justice, and public trust. Due to this there is a increased the demand for fake video analysis, detection and intervention. Some of the related word in deep fake detection are listed below:

Exposing deepfake Videos by Detecting Face Warping Artifacts [2] used an approach to detects artifacts by comparing the generated face areas and their surrounding regions with a dedicated Convolutional Neural Network model. In this work there were two-fold of Face Artifacts. Their method is based on the observations that current DF algorithm can only generate images of limited resolutions, which are then needed to be further transformed to match the faces to be replaced in the source video.

Exposing AI Created Fake Videos by Detecting Eye Blinking [3] describes a new method to expose fake face videos generated with deep neural network models. The method is based on detection of eye blinking in the videos, which is a physiological signal that is not well presented in the synthesized fake videos. The method is evaluated over benchmarks of eye-blinking detection datasets and shows promising performance on detecting videos generated with Deep Neural Network based software deepfake. Their method only uses the lack of blinking as a clue for detection. However certain other parameters must be considered for detection of the deep fake like teeth enchantment, wrinkles on faces etc. Our method is proposed to consider all these parameters.

Using capsule networks to detect forged images and videos [4] uses a method that uses a capsule network to detect forged, manipulated images and videos in different scenarios, like replay attack detection and computer-generated video detection. In their method, they have used random noise in the training phase which is not a good option. Still the model performed beneficial in their dataset but may fail on real time

data due to noise in training. Our method is proposed to be trained on noiseless and real time datasets.

Detection of Synthetic Portrait Videos using Biological Signals [5] technique to extract biological signals from the face regions of real and false portrait video pairs. Utilize transformations to train a CNN and a probabilistic SVM, compute the temporal consistency and spatial coherence, and capture the signal properties in feature sets and PPG maps. Subsequently, the total authenticity probabilities are used to determine the legitimacy of the video.

With the aim of identifying a deepfake video, the proposed model uses a spatially and temporally aware pipeline that performs consistently and has a low false alarm rate at the video level [6]. The main findings state the use of a two-stream convolutional neural network for feature extraction and classification of video manipulation, the introduction of a novel spatial and temporal-aware pipeline for automatic detection of deepfake videos, and the demonstration of promising performance against various benchmarks.

The proposed method of training the SVM with feature-detector-descriptors can be useful in the detection of fake videos [7]. Using a dataset of real and fake videos, the suggested methodology first uses SVM regression trained with feature points from many feature-point detectors to detect deepfake videos. Afterward, various feature point detectors are evaluated.

In order to identify Deepfake films, the study proposed a deep learning-based technique that performed satisfactorily and had strong generalizability [8]. Using data enhancement techniques to provide a significant amount of training samples, the process comprised building a model to recognize Deepfake films based on ResNext.

Deepfake material in photos and videos can be identified using deep learning techniques [9]. The performance of many deepfake video detection models, including CNN models like ResNet, VGG16, and Efficient net and RNN models like Long Short-Term Memory LSTM, is compared in this research. A comparative analysis of several deepfake video detection algorithms in deep learning is conducted in this paper.

Deepfake Video Detection Using Recurrent Neural Networks [10] Using a temporal-aware pipeline, competitor results in the automatic detection of deepfake videos can be obtained. In order to produce competitive results with an easily comprehensible design, the main findings suggest a temporal-aware pipeline for the automatic detection of deepfake films that combines CNN and RNN. The procedure entails using a CNN to obtain frame-level data and training an RNN to identify video as altered or not. The RNN is then evaluated against a large collection of deepfake videos.

The study offers an innovative approach for determining deepfake films with fewer processing cycles and yields more accurate results [11]. To detect deepfake films, the method computes the variance of Laplacian for various face patches, employs a three-layer neural network classifier, and looks for visual artifacts in face areas.

The proposed research explored across multiple methods for detecting deepfake videos, and it achieved 72.5% accuracy on the validation set [12]. The method consisted of examining the temporal structure between frames, concentrating on faces within video frames using facial detection algorithms, and using a CNN and RNN combination to generate embeddings from facial features.

Using Multi-task Cascaded Convolution Neural Network (MTCNN), the study developed a web platform that has the

ability to recognize deepfake videos that have been intentionally manufactured [13]. In the experimental setting, the platform obtained an accuracy of 78.8 percent. The methodology comprised building a web platform with Multi-task Cascaded Convolution Neural Network (MTCNN) for deepfake video detection and noise analysis of facial landmarks.

Hybrid Recurrent Deep Learning Model for DeepFake Video Detection [14] Deepfake videos are created digitally by altering facial features in order to superimpose a particular person's face data on other videos. Using noise-based temporal face convolutional features and hybrid recurrent deep learning models, the methodology proposed multilayer hybrid recurrent deep learning models with proven performance above stacked recurrent deep learning models for the purpose of detecting deepfake videos.

Detecting DeepFakes with Deep Learning [15] A novel architecture called Eff-YNet capable of both segmenting and detecting frames from deepfake videos is proposed. The methodology involves proposing a pipeline with two pathways for examining frames and video clips, developing Eff-YNet for image analysis, implementing ResNet3D for video analysis, and testing the model on the Deepfake Detection Challenge dataset.

Quantum machine learning was able to train the model with 50% less time than classical model [16]. In this project the data from faceforencis was considered and frames are generated from videos followed by application of novel technique in classification by using face detection over videos. These images are transformed using various parameters and models are compared against them. Highest AUC value of 0.95 is obtained using Laplace transformed images. Quantum machine learning is also applied and was able to train the model with 50% less time than classical model.

Deepfake Video Detection Based on MesoNet with Preprocessing Module [17] The preprocessing module enhances discrimination among multi-color channels in face images. The proposed Deepfake detection method combining MesoNet with the preprocessing module maintains good robustness even under heavy compression. The method outperforms on the Celeb-DF dataset. The methodology involves the introduction of a preprocessing module to enhance discrimination between Deepfake and real images, its combination with MesoNet for detection, and verification of its effectiveness through an ablation experiment.

Deepfake Video Detection by Using Convolutional Gated Recurrent Unit [18] since deepfake videos are difficult to recognize by human eyes, it becomes important to automatically detect these forgeries and prevent their abuse. The paper, propose a deep neural network model to detect deepfake videos using a convolutional neural network (CNN) to extract frame-level features. These features are then used to train a convolutional GRU that learns to distinguish between fake and real videos. Evaluation is performed on the recently released Celeb-DF(v2) datasets where 0.83 AUC score was achieved.

III. PROPOSED METHODOLOGY

After examining the problem description, it was possible to determine whether the problem had the potential to be addressed. A lot of different research paper are referred as mentioned above. After checking the feasibility of the problem statement. The next step is the dataset gathering and analysis. We analysed the data set in different approach of training like negatively or positively trained i.e. training the model with only fake or real video’s but found that it may lead to addition of extra bias in the model leading to inaccurate predictions. So, after doing lot of research,

it was concluded that the balanced training of the algorithm is the best way to avoid the bias and variance in the algorithm and get a good accuracy. The proposed paper analysed the solution in terms of cost, speed of processing, requirements, level of expertise, availability of equipment's.

There are many tools available for creating the deepfake, but for deepfake detection there is hardly any tool available. Our approach for detecting the DF will be great contribution in avoiding the percolation of the deepfake over the world wide web. The Proposed paper has provided a machine learning platform for the user to upload the video and classify it as fake or real. This project can be scaled by developing a web-based platform or a browser plugin for automatic deepfake detections. Even big application like WhatsApp, Facebook can integrate this project with their application for easy pre detection of deepfake before sending to another user. One of the important objectives is to evaluate its performance and acceptability in terms of security, accuracy, and reliability. Our method is focusing on detecting all types of deepfakes like replacement deepfake, retrenchment deepfake and interpersonal deepfake.

1. Dataset

For making the model efficient for real time prediction. The data was gathered from different available datasets majorly from Deepfake detection challenge (DFDC) and Celeb-DF [19] dataset. Further, shuffled and mixed-up the collected datasets and created our own new dataset, for more accurate and real time detection on different kind of videos.

Parameter Identified

1. Blinking of eyes
2. Teeth enchantment
3. Bigger distance for eyes
4. Moustaches
5. Double edges, eyes, ears, nose
6. Iris segmentation
7. Wrinkles on face
8. Inconsistent head pose
9. Face angle
10. Skin tone
11. Facial Expressions
12. Lighting
13. Different Pose
14. Double chins
15. Hairstyle
16. Higher cheek bones

To avoid the training biasness of the model it has been developed considering 50% Real and 50% fake videos. Further the pre-processing of DFDC dataset is done and the audio altered videos are removed from the dataset by running a python script. After preprocessing of the DFDC dataset, we have taken 800 Real and 800 Fake videos from the DFDC dataset.

2. Pre-processing

In this step, the videos are pre-processed and all the unrequired and noise is removed from videos. Only the required portion of the video i.e. face is detected and cropped. The first steps in the preprocessing of the video are to split the video into frames. After splitting the video into frames, the face is detected in each of the frame and the frame is cropped along the face. Later the cropped frame is again converted to a new video by combining each frame of the video. The process is followed for each video which leads to creation of processed dataset containing face only videos. The frame that does not contain the face is ignored while preprocessing. To maintain the uniformity of number of frames, the proposed model has selected a unique threshold value based on the mean of total frames count of each video Another reason for selecting a threshold value is limited computation power. As a video of 10 second at 30 frames per second(fps) will have total 300 frames and it is computationally very difficult to process

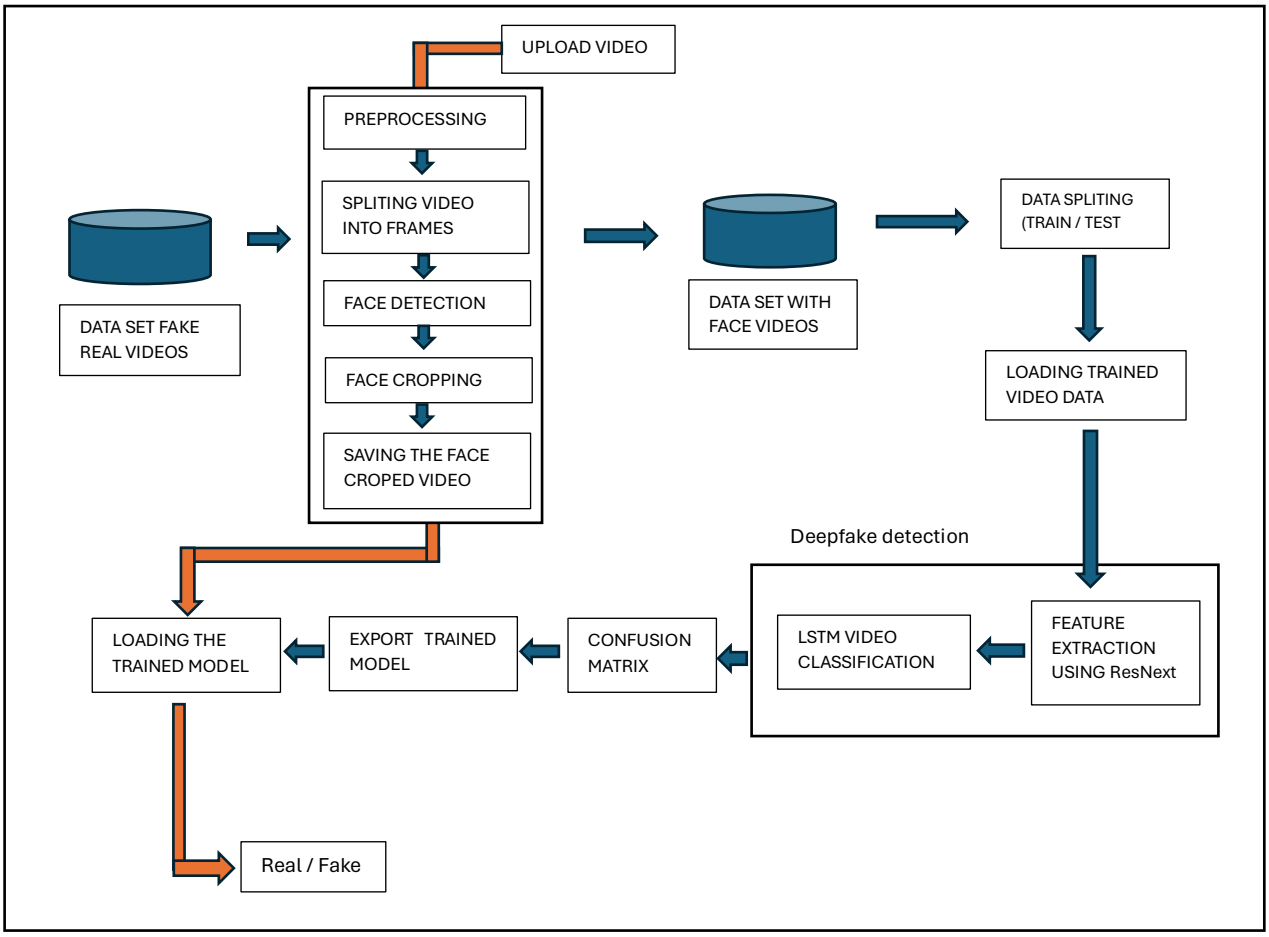

Fig: 1 – Architecture of proposed model

the 300 frames at a single time in the experimental environment. So, based on the available Graphic Processing Unit (GPU) the proposed model has selected 150 frames as the threshold value. While saving the frames to the new dataset model have only saved the first 150 frames of the video to the new video. To demonstrate the proper use of Long Short-Term Memory (LSTM) paper proposed to consider the frames in the sequential manner i.e. first 150 frames and not randomly. The newly created video is saved at frame rate of 30 fps and resolution of 112 x 112. [Fig. 1] demonstrates the complete architecture of the proposed model.

3. Data-set split

The dataset is split into train and test dataset with a ratio of 70% train videos (1120) and 30% (480) test videos. The train and test split is a balanced split i.e. 50% of the real and 50% of fake videos in each split. The training is done for 20 epochs with a learning rate of 1e-5 (0.00001), weight decay of 1e-3 (0.00001) using the Adam optimizer. To enable the adaptive learning rate Adam optimizer with the model parameters is used.

4. Model

Our model is a combination of ResNext 50_32x4d and Long Short-Term Memory (LSTM). ResNext CNN model is used to extract the features at frame level and based on the extracted features a LSTM network is trained to classify the video as deepfake or pristine. Using the Data Loader on training split of videos the labels of the videos are loaded and fitted into the model for training.

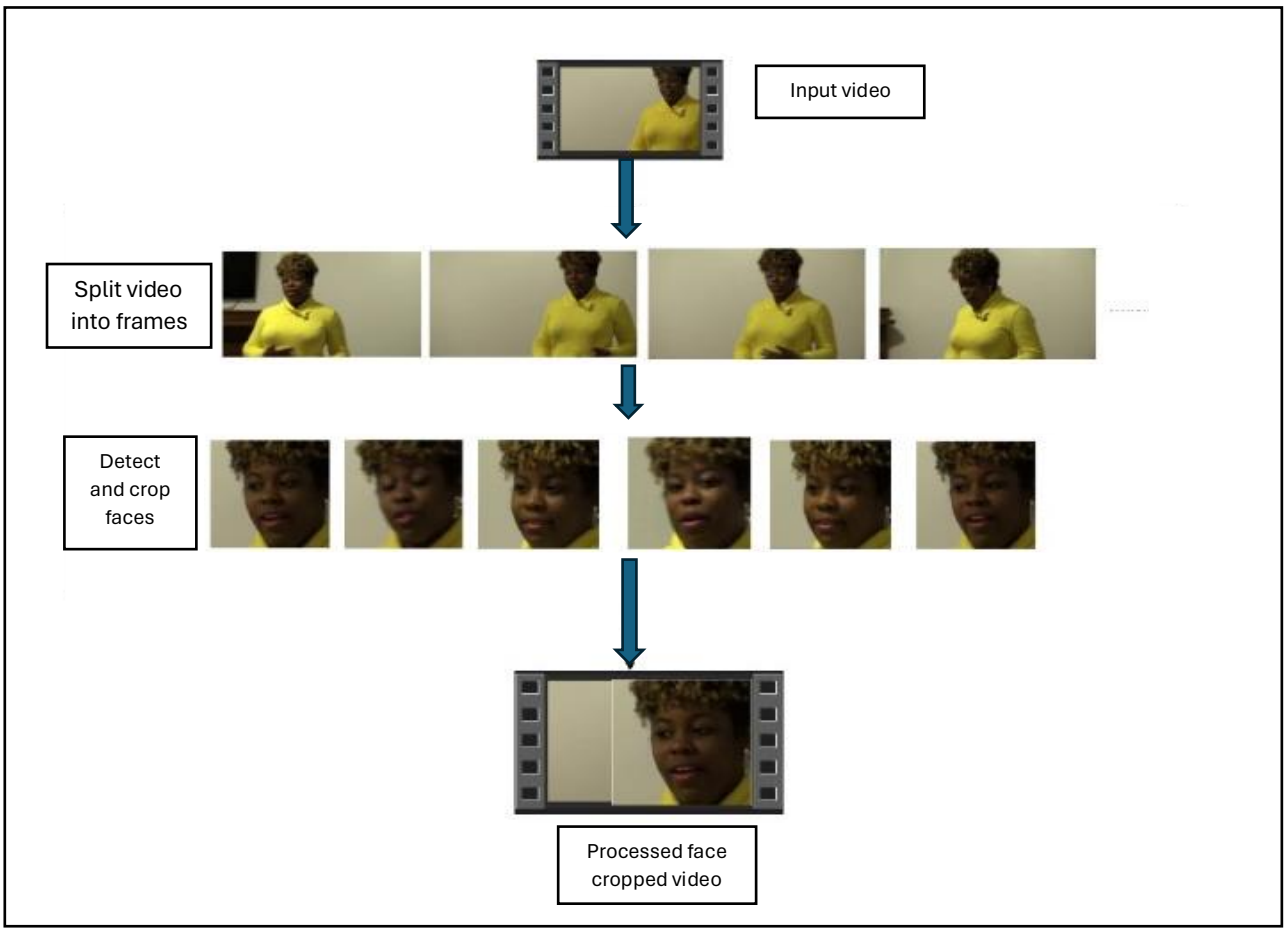

Fig: 2 – Preprocessing of video

ResNext:

ResNext [20] is Residual CNN network optimized for high performance on deeper neural networks. For the experimental purpose proposed paper has used ResNext50_32x4d model. ResNext of 50 layers and 32 x 4 dimensions has been used primarily. Following it, for fine-tuning the network an extra required layers was added, and a proper learning rate is selected to properly converge the gradient descent of the model. The 2048-dimensional feature vectors after the last pooling layers of ResNext is used as the sequential LSTM input.

LSTM for Sequence Processing:

2048-dimensional feature vectors are fitted as the input to the LSTM [21]. Proposed methodology is based on using LSTM layer with 2048 laten dimensions and 2048 hidden layers along with 0.4 chance of dropout, which is capable to do achieve our objective. LSTM is used to process the frames in a sequential manner so that the temporal analysis of the video can be made, by comparing the frame at 't' second with the frame of 't-n' seconds. Where n can be any number of frames before t. The model also consists of Leaky Relu activation function. A linear layer of 2048 input features and 2 output features are used to make the model capable of learning the average rate of correlation between each input and output. An adaptive average polling layer with the output parameter 1 is used in the model. Which gives the target output size of the image of the form H x W.

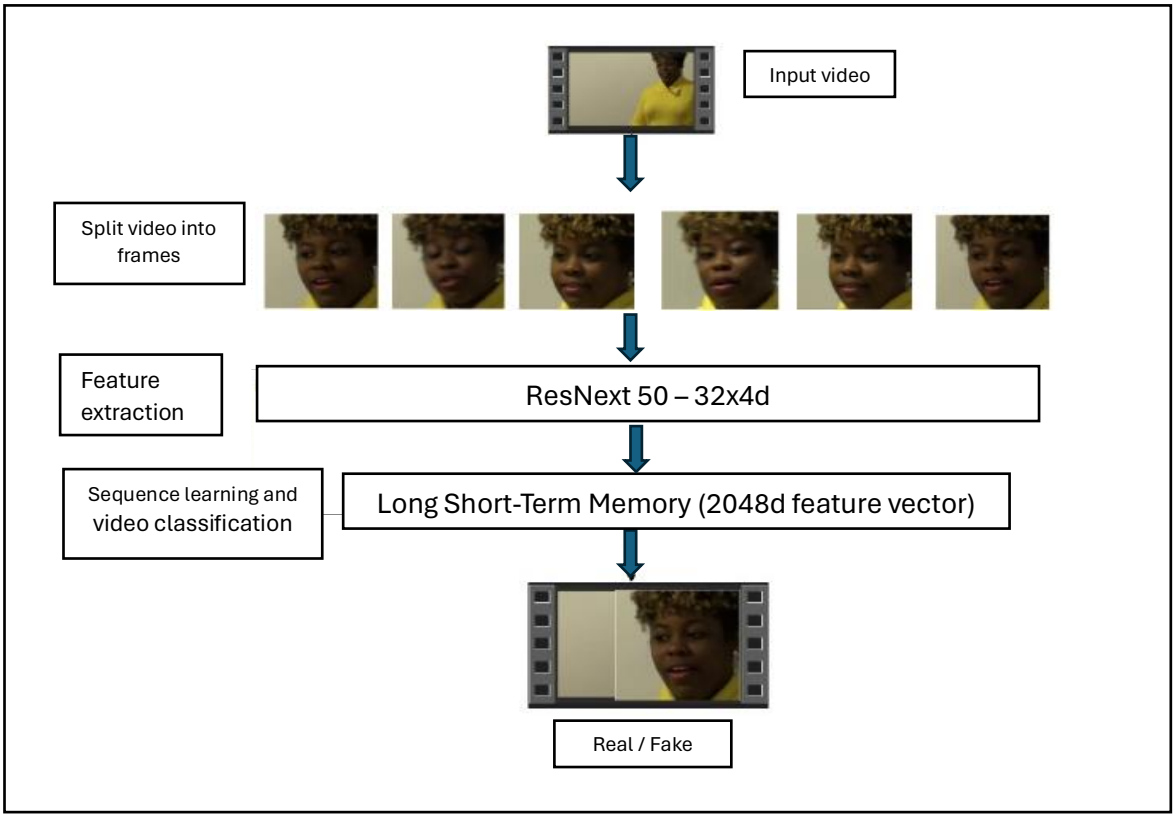

Fig: 3 – Overview of model

[Fig.3] demonstrate the overview of the model architecture. For sequential processing of the frames a Sequential Layer is used. The batch size of 4 is used to perform the batch training. A SoftMax layer is used to get the confidence of the model during predication.

5. Hyper-parameter tuning

In this process a hyper-parameter for achieving the maximum accuracy has been chosen. After reiterating many times on the model. The best hyper-parameters for our dataset are chosen. To enable the adaptive learning rate Adam [21] optimizer with the model parameters is used. The learning rate is tuned to 1e-5 (0.00001) to achieve a better global minimum of gradient descent. The weight decay used is 1e-3. As this is a classification problem so to calculate the loss cross entropy approach is used. To use the available computation power properly the batch training is used. The batch size is taken of 4. Batch size of 4 is tested to be ideal size for training in our development environment. The uploaded video is then passed to the model and prediction is made by the model. The model returns the output whether the video is real or fake along with the confidence of the model. [Fig. 4.1] explains the complete flow of the training flow of the proposed model. [Fig. 4.2] explains the prediction workflow of the model.

6. Prediction

A new video is passed to the trained model for prediction. A new video is also pre-processed to bring in the format of the trained model. The video is split into frames followed by face cropping and instead of storing the video into local storage the cropped frames are directly passed to the trained model for detection.

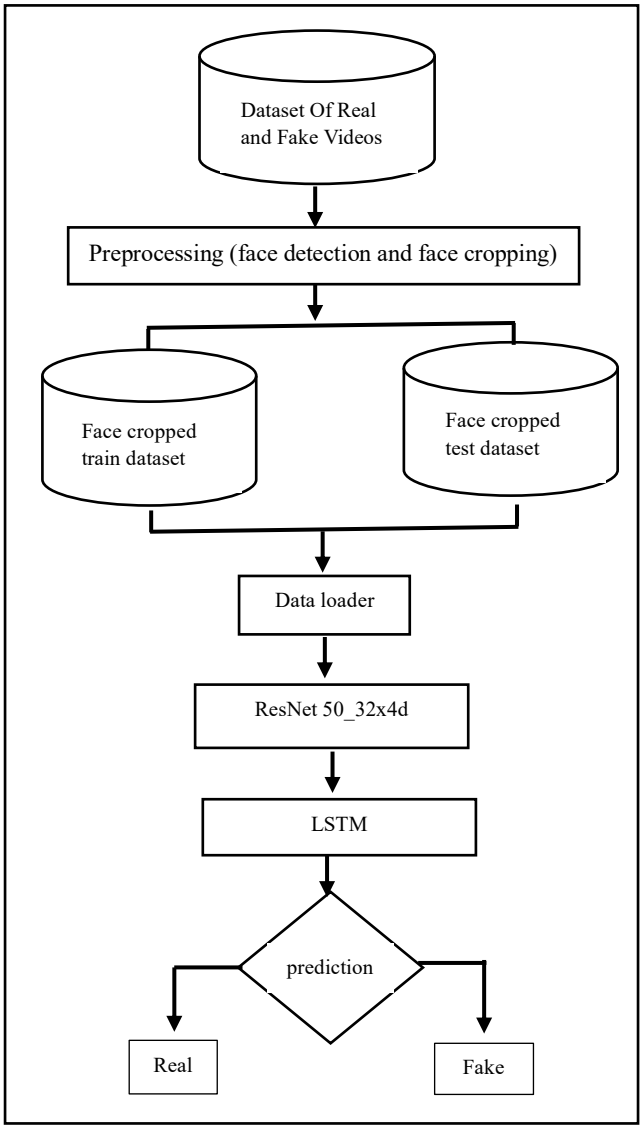

Fig: 4.1 - Training flow

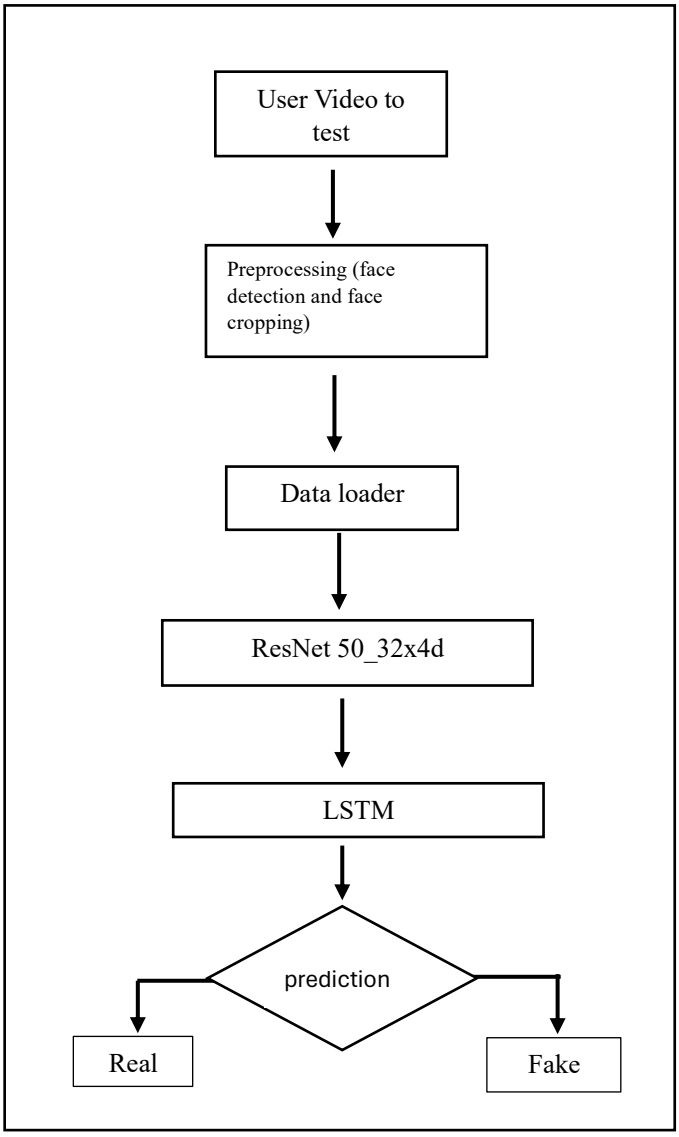

Fig: 4.2 – Prediction flow

IV. Results

Our study introduced a deep learning-based method designed to identify and distinguish between AI generated deepfake videos and genuine videos. By using the Res-Next Convolutional Neural Network (CNN) architecture, the paper defines the development of a system more capable of extracting frame level features crucial for identification of manipulated videos. Along with this Long Short-Term Memory (LSTM) based Recurrent Neural Network (RNN) used successfully for the classification of videos as either genuine or manipulated. To validate proposed method`s effectiveness, an extensive experiment on large datasets is conducted, including the Deepfake Detection Challenge and Celeb-DF dataset and results are as follows. [Fig.5] demonstrates training and validation loss of the model. [Fig. 6] is displaying the training and testing accuracy of model. An increasing slope of accuracy can be observed with decreasing loss.

Sr no.	Model	No. of videos	Sequence length	Accuracy
1	model_80_frames	1600	80	97.48743
2	model_100_frames	1600	100	93.97781
3	model _40_frames	1600	40	89.34681
4	model _20_frames	1600	20	84.21461

Table: 1 – Trained model results

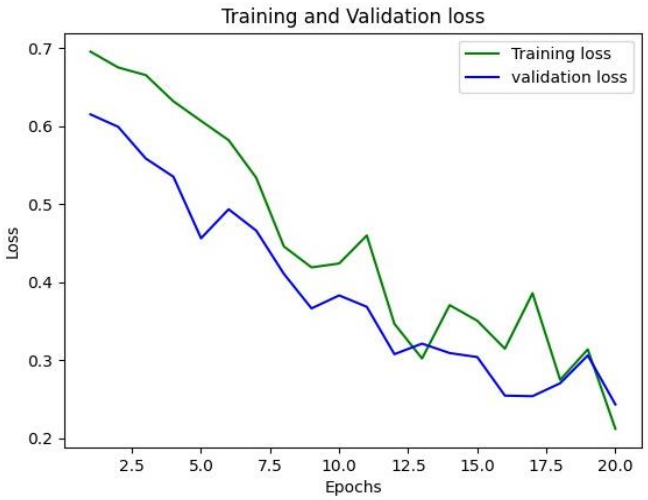

Fig.5 – Training and Validation Loss of model

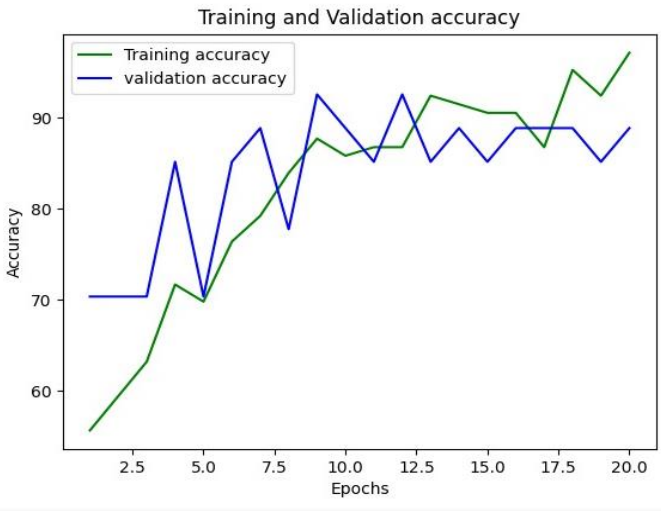

Fig. 6 - Training and Validation Accuracy of model

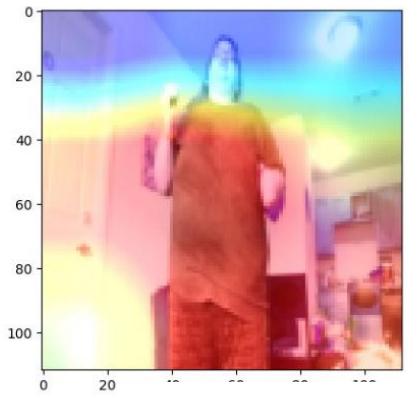

Fig.7 – Output

V. Conclusion

The proposed model presented a neural network-based approach to classify the video as deep fake or real, along with the confidence. [Fig. 7] shows the actual output of the model. Our method is capable of predicting the output by processing 1 second of video (30 frames per second) with a good accuracy. The proposed model is implemented using ResNext CNN model to extract the frame level features and LSTM for temporal sequence processing to spot the changes between the t and t-1 frame. Our model can process the video in the frame sequence of 10,20,40,60,80,100

.

VI. Future Scope

There is always a scope for enhancements in any developed system, especially when the project is built using latest trending technology and has a good scope in future.

- Our method has not considered the audio. That is why there is a scope for integration of audio deepfake detection method in the proposed model.
- Currently only Face Deep Fakes are being detected by the algorithm, but the algorithm can be enhanced in detecting full body deep fakes.

VII. References

[1] Deepfake detection challenge dataset: https://www.kaggle.com/c/deepfake-detectionchallenge

[2] Yuezun Li, Siwei Lyu, "ExposingDF Videos By Detecting Face Warping Artifacts," in arXiv:1811.00656v3.

[3] Yuezun Li, Ming-Ching Chang and Siwei Lyu "Exposing AI Created Fake Videos by Detecting Eye Blinking" in arxiv.

[4] Huy H. Nguyen , Junichi Yamagishi, and Isao Echizen " Using capsule networks to detect forged images and videos ".

[5] Umur Aybars Ciftci, ˙Ilke Demir, Lijun Yin "Detection of Synthetic Portrait Videos using Biological Signals" in arXiv:1901.02212v2.

[6] Waseem, Saima et al. "A Multi-color Spatio-Temporal Approach for Detecting DeepFake." 2022 12th International Conference on Pattern Recognition Systems (ICPRS) (2022): 1-5.

[7] Kharbat, Faten F. et al. "Image Feature Detectors for Deepfake Video Detection." 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA) (2019): 1-4.

[8] Zhou, Xin et al. "Detecting Deepfake Videos via Frame Serialization Learning." 2020 IEEE 3rd International Conference of Safe Production and Informatization (IICSPI) (2020): 391-395.

[9] Das, Athirasree et al. "A Survey on Deepfake Video Detection Techniques Using Deep Learning." 2022 Second International Conference on Next Generation Intelligent Systems (ICNGIS) (2022): 1-4.

[10] Guera, David and Edward J. Delp. "Deepfake Video Detection Using Recurrent Neural Networks." 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (2018)

[11] Habeeba, M. A. Sahla et al. "Detection of Deepfakes Using Visual Artifacts and Neural Network Classifier." (2020).

[12] Dash, Aleksander and Nolan Handali. "CS 230 Final Report: Deepfake Video Detection." (2020).

[13] Shilpah, L and Student. "DETECTION OF THE AI GENERATED DEEPFAKE VIDEO BY USING MULTI-TASK CASCADED CONVOLUTION NEURAL NETWORK." (2021).

[14] Jaiswal, Gaurav. "Hybrid Recurrent Deep Learning Model for DeepFake Video Detection." 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) (2021): 1-5.

[15] Tjon, Eric. "Detecting DeepFakes with Deep Learning." (2021).

[16] Byreddy, Nikhil Reddy and Vladimir Milosavljevic. "DeepFake Videos Detection Using Machine Learning." (2019).

[17] Xia, Zhiming et al. "Deepfake Video Detection Based on MesoNet with Preprocessing Module." Symmetry 14 (2022).

[18] Tu, Yifeng et al. "Deepfake Video Detection by Using Convolutional Gated Recurrent Unit." Proceedings of the 2021 13th International Conference on Machine Learning and Computing (2021): n. pag.

[19] Yuezun Li , Xin Yang , Pu Sun , Honggang Qi and Siwei Lyu "Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics" in arXiv:1909.12962

[20] ResNext Model: https://pytorch.org/hub/pytorch_vision_resnext

[21] Hyeongwoo Kim, Pablo Garrido, Ayush Tewari and Weipeng Xu "Deep Video Portraits" in arXiv:1901.02212v2.