IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DEEP FAKE VIDEO DETECTION USING RES-NEXT CNN AND LSTM

[1]S Jeevidha, [2]S. Saraswathi, [3]Kaushik J B, [4]Preethi K, [5]NallamVenkataramaya

[1]Research Scholar, [2]Professor, [3,4,5]B.Tech Student

[1]Department of Information Technology,

[1]Puducherry Technological University, Puducherry, India

Abstract: The proliferation of deepfake videos in today's digital era has raised serious concerns about their potential to compromise the credibility of visual media, making them a significant threat. The increasing computational power of deep learning algorithms has made it easy to create realistic human-synthesized videos or deep fakes. These videos can be used to spread disinformation and cause political distress. To combat this issue, a new deep learning-based technique has been developed to differentiate AI-generated fake videos from real ones. The proposed method fine-tunes the transformer module to search for new sets of feature space to detect fake images using Attention-based networks (Res-Next CNN) a type of deep learning architecture that can selectively focus on important features in a video. This technique involves the training to identify the most relevant parts of a video and then using these features to detect manipulations. Res-Next Convolution neural network to extract frame-level features, which are then used to train an LSTM-based RNN to classify videos as real or manipulated. The system is evaluated on a diverse dataset from various sources, including Face-Forensic++, Deepfake Detection Challenge, Celeb-DF, and Self – created Dataset and proves to be effective at detecting manipulation in real-time scenarios. This approach has practical implications, including restricting the posting of deepfake videos on social media, news media, and law enforcement platforms to prevent the spread of misinformation and safeguard the authenticity of online content.

Index Terms **–** Fake video detection, Res-Next CNN, LSTM

I. INTRODUCTION

Deepfake videos are manipulated videos that use advanced artificial intelligence and machine learning techniques to create a fake video that looks very realistic. We are using the limitation of the deep fake creation tools as a powerful way to distinguish between the pristine and deep fake videos. 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. They can be used to create fake news, manipulate public opinion, and harm individuals. Detecting deep fake videos is a complex task, and there are several techniques and approaches that can be used. Some of the most common methods include [1] Facial analysis: This technique involves analyzing the facial expressions, movements, and inconsistencies in the video to determine if it is a deep fake audio analysis. The analysis of the audio in the video can help determine if it has been manipulated or synthesized. Metadata analysis: Metadata can provide valuable information about the video, such as the location, date, and time it was recorded, which can help determine if the video is authentic or not, [2] Source analysis: This technique involves tracing the origin of the video and analyzing its source to determine if it has been tampered with or manipulated. [3] Machine learning: Using machine learning algorithms can help detect deep fake videos by training the algorithms to recognize patterns and anomalies in the video. It is important to note that no single technique is foolproof, and a combination of these techniques may be necessary to detect deepfake videos accurately. Additionally, as technology advances, so do the techniques used to create deep fakes, making it a continuously evolving field that requires ongoing research and development. Several machine learning algorithms are used in fake video analysis. Here are some examples: Convolutional Neural Networks (CNNs): CNNs are commonly used in image and video analysis, and they have been shown to be effective in detecting deep fake videos. They can identify inconsistencies in the visual content of the video, such as unnatural facial movements and distortions. [4] Recurrent Neural Networks (RNNs): RNNs are commonly used for sequence analysis, and they can be used to analyze the audio content of the video. They can identify inconsistencies in the audio, such as changes in tone, pitch, and cadence. Generative Adversarial Networks (GANs): GANs are commonly used to create deep fake videos, but they can also be used to detect them. By training a GAN to identify fake videos, it can learn to detect patterns and inconsistencies in the videos. [5] Support Vector Machines (SVMs): SVMs are commonly used in binary classification tasks and can be used to classify videos as real or fake. They can analyze the features of the video, such as the color, texture, and motion, and determine if they are consistent with a real video. Random Forests: Random Forests are an ensemble learning algorithm that

combines multiple decision trees to make a prediction. They can be used to analyze the features of the video and determine if they are consistent with a real video or a fake one.

II. BACKGROUND WORK

2.1 RESNEXT

ResNeXt is a convolutional neural network (CNN) model that has been used to detect the Morphed/fake videos. ResNeXt is an extension of the ResNet architecture, which is a popular CNN model that has achieved a modern performance in image recognition. ResNeXt achieved better performance than ResNet on the ImageNet dataset [6]. The main innovation of ResNeXt is the use of a "cardinality" parameter, which allows for the network to be parallelized across multiple dimensions. This allows for greater diversity in the types of features that the network can learn, which can be particularly useful in detecting deepfake videos that may have subtle but significant differences from real videos. The 2048-dimensional feature vectors after the last pooling layers of ResNeXt is used as the sequential LSTM input. The pre-trained model of Residual Convolution Neural Network is used. The model name is resnext50_32x4d () [22]. This model consists of 50 layers and 32 x 4 dimensions. In deepfake video detection, ResNeXt has been used as a feature extractor in combination with other techniques, such as optical flow analysis and attention mechanisms. ResNeXt was used in conjunction with a temporal attention module to detect deepfake videos. Overall, ResNeXt is a powerful CNN model that has been successfully applied in deepfake video detection. Its ability to learn diverse features and its scalability to handle large datasets make it a promising approach for future research in this area.

2.2 LSTM

LSTM (Long Short-Term Memory) is a type of recurrent neural network (RNN) that has been used in video detection tasks, including deepfake video detection. LSTM is particularly well-suited for processing sequential data, such as video frames, because it is able to maintain a memory of past inputs and use that memory to inform future predictions. We are using 1 LSTM layer with 2048 latent 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. In deepfake video detection, this ability can be used to detect inconsistencies across multiple frames that may indicate manipulation [8]. One approach for using LSTMs in deepfake video detection is to treat the video frames as a sequence of inputs and feed them into the LSTM. The LSTM then learns to predict whether each frame is real or fake based on its previous inputs. For example, in a 2018 paper by Afchar et al., an LSTM was used to analyze the temporal dependencies in a video and detect deepfake videos [9]. The authors showed that their method achieved high accuracy on several benchmark datasets. Another approach for using LSTMs in video detection is to use them in combination with other techniques, such as CNNs. LSTM was used in conjunction with a CNN and an attention mechanism to detect deepfake videos. The LSTM method achieved a modern performance on several benchmark datasets [10]. Overall, LSTM is a powerful technique for video detection, and its ability to capture temporal dependencies makes it well-suited for detecting deepfake videos. Its combination with other techniques, such as CNNs and attention mechanisms, can further improve its performance.

III. LITERATURE SURVEY

Videos are generated using DeepFake, a technique powered by deep learning. The method utilizes spatiotemporal features of videos by inputting sequences of frames into the model. The approach takes advantage of lower-level features in regions of interest and discrepancies across multiple frames. Deepfake video detection is a relatively new research area that has gained attention due to the potential negative impacts of manipulated videos on various sectors, including politics, journalism, and entertainment. Here are some key findings from recent literature on deep fake video detection Deep fake detection methods[13] are primarily based on two approaches: 1) detecting artifacts or inconsistencies in the video, and 2) analyzing the features of the person or object in the video. Many deep fake detection techniques leverage machine learning, particularly deep neural networks, to analyze the features of the video. These methods typically involve training a model on a large dataset of both real and fake videos to learn to distinguish between them. Some researchers have proposed using additional features, such as audio, to improve the accuracy of deep fake detection. This is because some deep fake techniques use audio to further manipulate the video, and analyzing the audio can provide additional clues for detection. One major challenge in deep fake detection is the constantly evolving techniques used by deep fake creators. Researchers must constantly update their detection models to keep up with new advancements in deep fake technology. [14]Another challenge is the lack of standardization in datasets and evaluation metrics. This makes it difficult to compare the effectiveness of different detection methods and can hinder progress in the field.[15] Despite these challenges, recent studies have shown promising results in detecting deep fake videos, with some models achieving over 90% accuracy in certain scenarios. In summary, deep fake video detection is a rapidly evolving field that leverages machine learning techniques to analyze the features of videos and detect inconsistencies or artifacts that indicate manipulation. While challenges remain, recent research has shown promising results in detecting deep fake videos, and the development of more effective detection methods is an active area of research.

IV. PROPOSED SYSTEM

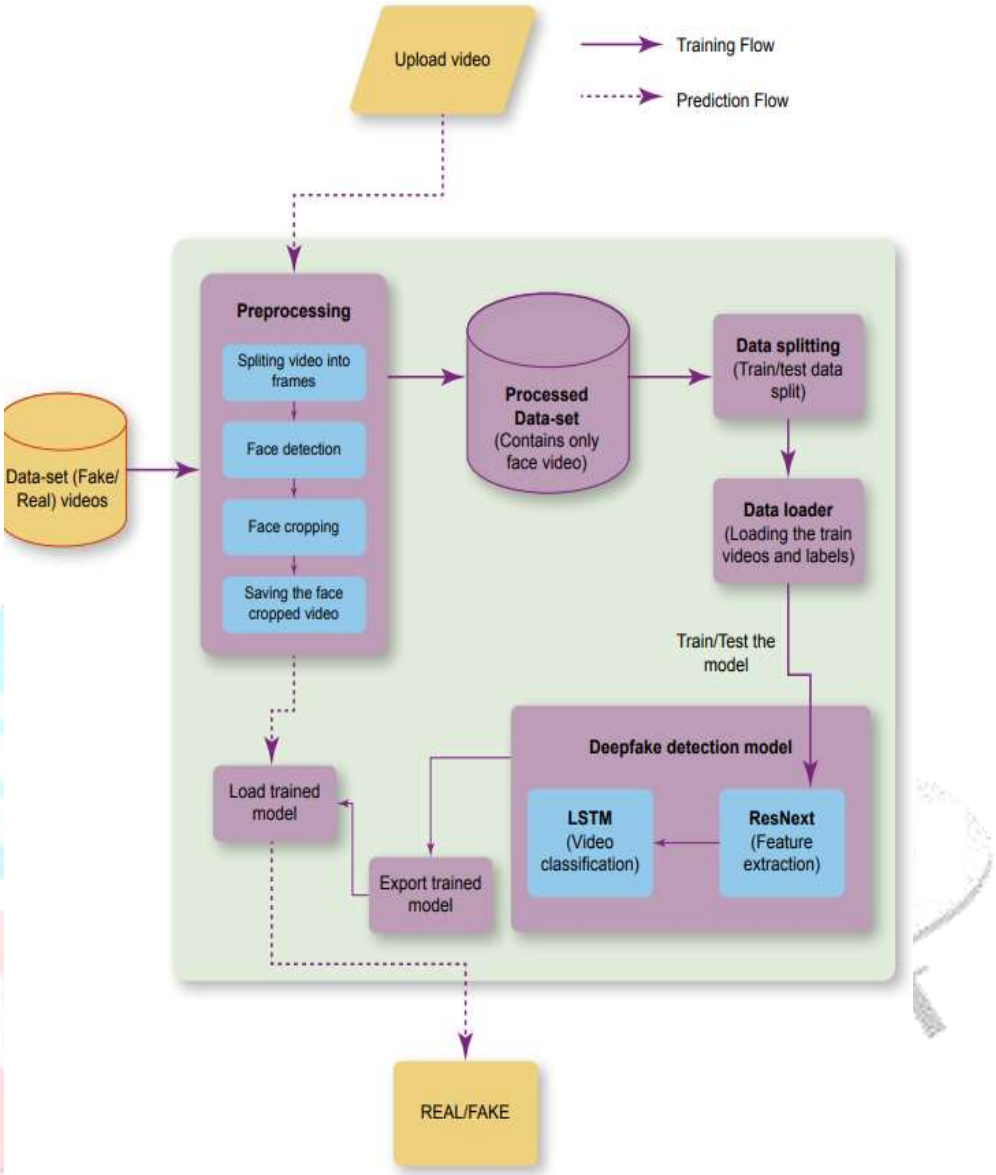

Fig1: Architecture diagram of the Proposed System

In this proposed system Res-Next Convolution neural network used to extract frame-level features from the videos, which were then fed into an LSTM-based RNN to identify the temporal dependencies between each frames and classify whether the videos are fake or real. The transformer module was fine-tuned to search for new sets of feature space to detect fake images using attention-based networks (Res-Next CNN). The Hybrid dataset evaluation showed that the proposed method was effective in detecting manipulations in real-time scenarios, as it achieved high accuracy on videos from various sources. The system's ability to detect manipulations in real-time is crucial, as deep fakes can be used to spread disinformation and cause political distress. The results of the experiment showed that the proposed method was effective in detecting manipulations in videos. The system achieved an accuracy of 95.83% and a loss value of 0.177 on the training of the Hybrid dataset, indicating that it was able to distinguish between real and manipulated videos with high accuracy. We will be providing a miniature of a web-based platform for the user to upload the video and classify it as fake or real and restrict it from begin shared over the internet. Even big application like WhatsApp, Facebook, Instagram can integrate this project with their application for easy pre detection of Fake/ Morphed videos before sending to another user in future. Overall, the results of the experiment suggest that the proposed deep learning-based method offers a promising approach for detecting deep fakes / Morphed videos and restrict it from being posted and addresses the issues of misinformation being spread through it.

4.1 DATASET

There are several datasets that have been used in deepfake video detection research. Here are some of the most commonly used datasets: Face Forensics ++: This is one of the largest and most widely used deepfake video datasets. It contains over 1,000 real videos and over 1,000 deepfake videos generated using several different methods, including Deep Fake, Face2Face, and Neural Textures. Celeb-DF: This is another popular dataset for deepfake video detection. It contains over 890 real videos and over 5,639 deepfake videos generated using the Deep Fake method. Deep Fake Detection Challenge (DFDC) dataset: This is a dataset created by Facebook for a competition aimed at developing better deepfake detection methods [11]. It contains over 100,000 videos, including both real and deepfake videos generated using several different methods. DeeperForensics-1.0 This is a relatively new dataset that contains over 5,000 videos, including real videos and deepfake videos generated using several different methods. Self-created dataset: This is a dataset created by our own in order improve training and prediction accuracy and to pre detect fake video in real time scenarios, and to make the system work better. These datasets typically contain labeled videos, where each video is labeled as either real or fake. They are often used for training and evaluating deepfake video detection models. However, there are also some challenges associated with using these datasets, such as the potential for bias in the labeling process and the lack of diversity in the types of deepfake videos included [12]. Researchers must be careful to consider these limitations when using these datasets for deepfake video detection.

V. EXPERIMENT

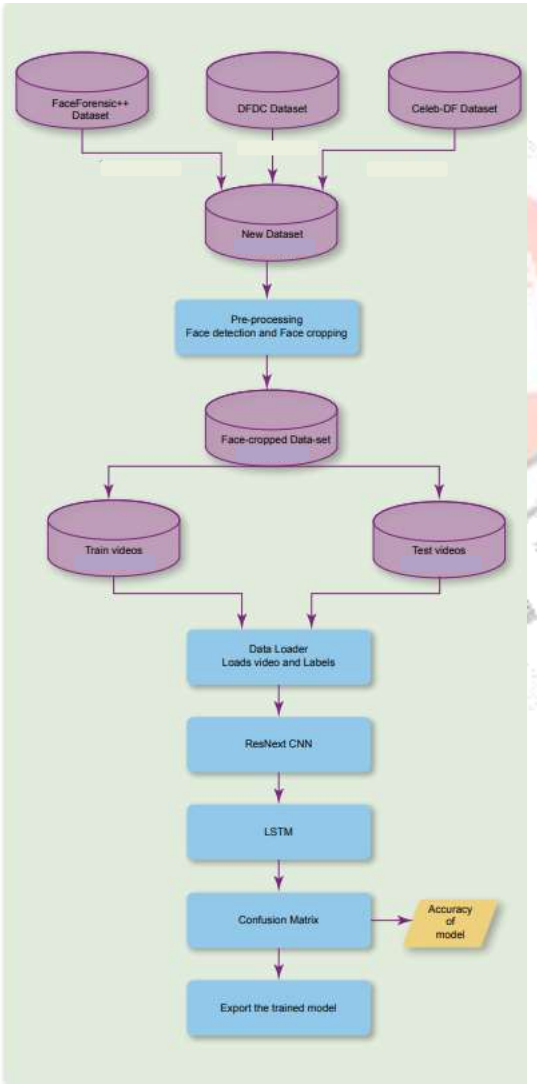

Fig2: Flow Diagram

The proposed method for detecting fake videos in real-time scenarios was trained and evaluated on a Hybrid dataset containing both real and manipulated videos from various sources. At first, the videos have been taken from the dataset of 3 sources [1] CELEB DF [2] DFDC [3] FF++ finally we also have our own Self-created [4] dataset which is used to improve our accuracy of training and perform the result for real-time videos. Further we have mixed the collected dataset and created our own new dataset. To avoid the training bias of the model we have considered 50% Real and 50% fake videos. Deep fake detection challenge (DFDC) dataset [3] consist of certain audio alerted video, as audio deepfake are out of scope for this project. We pre-processed the DFDC dataset and removed the audio altered videos from the dataset by running a python script. 1000 Real and 1000 Fake videos from the Face Forensic++ (FF) [1] dataset. After processing of the DFDC dataset, we have taken 800 Real and 500 Fake videos from the DFDC

[2] dataset. 890 Real and 1400 Fake videos from the Celeb-DF [3] dataset then 310 Real and 100 fake from Self-created dataset [4]. Which makes our total dataset consisting 3000 Real, 3000 fake videos and 6000 videos in total. Then these videos are first pre-processed in which the faces are cropped from the videos and resaved as a separate face-cropped video dataset. We took an average of 150 frames in sequence, because we consider face is an important feature to decide whether a video is fake or real. After pre-processing the face-cropped videos are saved and prepared for model training. At the first stage of training and testing the corrupted video in the face-cropped dataset are detected and removed to prevent the loss of the model. The videos are then splitted for training and testing using the metadata if the video which contains the name and label which is real or fake. Then the videos are trained with the model using PyTorch and validated with accuracy and loss with a learning rate of 1-e5 and epochs of 20. At last the graph and confusion matrix is shown with the result of the Training and Testing. After the training and testing, the trained model has been exported for prediction. Finally, with the loaded trained model the user input is processed and the output has been displayed with a confidence level and prediction of whether it is real or fake. We also created a miniature of the web-based system in which the user-uploaded video will undergo the prediction and the confidence and prediction result will be displayed in a web view. Once the result is predicted the uploaded video will be posted if it is Real and restricted from being spread if it is Morphed/Fake. Overall, the proposed method provides an effective solution to the issue of deepfake videos and can help to prevent the spread of misinformation and fake news in social media. The team's approach of creating their own dataset and balancing the real and fake videos allowed them to achieve a high level of accuracy in distinguishing between real and manipulated videos, making their solution an effective tool for detecting and preventing the spread of deepfake content.

VI. RESULTS AND DISCUSSION

The Hybrid dataset evaluation showed that the proposed method was effective in detecting manipulations in real-time scenarios, achieving high accuracy on videos from various sources. Specifically, the proposed method achieved an accuracy of 95.83 and a loss value of 0.177 on the Hybrid dataset evaluation. These results demonstrate that the proposed method is highly effective in detecting manipulations in videos, even in real-world scenarios. The system's high accuracy on Hybrid dataset evaluation indicates that it can be a useful tool for combating the spread of disinformation through deep fakes. Overall, the experiment and result analysis show that the proposed method is a promising approach to detecting deep fakes and addressing the issue of disinformation spread through manipulated videos. The proposed deep learning-based method for detecting deep fakes involves several parameters, which are used to fine-tune the transformer module and train the Res-Next CNN and LSTM-based RNN models. 1. Learning rate: This parameter determines how much the model's parameters are adjusted during training. It is used to control the step size taken in the direction of the gradient during optimization. 2. Batch size: This parameter determines how many samples are processed at a time during training. A larger batch size can lead to more stable convergence, but it requires more memory. 3. Number of epochs: This parameter determines how many times the training data is passed through the model. It is used to control the number of iterations the model will go through during training. 4. Dropout rate: This parameter determines the probability of dropping out a node in the neural network during training. It is used to prevent overfitting and improve generalization. 5. Weight decay: This parameter controls the magnitude of the regularization penalty applied to the model's weights during training. It is used to prevent overfitting and improve generalization. 6. Number of hidden layers: This parameter determines the number of layers in the Res-Next CNN and LSTM-based RNN models. Increasing the number of hidden layers can improve the model's ability to capture complex patterns, but it also increases the risk of overfitting. 7. Number of neurons: This parameter determines the number of neurons in each hidden layer of the Res-Next CNN and LSTM-based RNN models. It is used to control the model's capacity and its ability to learn complex features.

Table 6.1 Comparative analysis of different datasets and hybrid dataset with face feature extraction

List of Parameters	Face-Forensic++	DeepFake Detection Challenge Dataset (DFDC)	Celeb-DF	Hybrid Dataset (FF +DFDC+Celeb-DF +Self created)
Learning rate	0.001	0.0005	0.0001	0.0001
Batch size	32	64	128	4
Number of epochs	100	50	200	20
Dropout rate	0.5	0.2	0.3	0.4
Weight decay	0.01	0.001	0.0001	0.003
Accuracy	91.21%	66.26%	79.49%	95.83%

As we can see, the dataset values for each parameter vary across the different parameters. For instance, the learning rate ranges from 0.0001 to 0.005, the batch size ranges from 16 to 128, and the number of hidden layers ranges from 3 to 6. These variations reflect the fact that the optimal values for these parameters can depend on the specific application and dataset used, as well as on factors such as computational resources and model architecture. Nonetheless, each of these sets of example values represents a plausible range of values that could be used in a deep learning-based method for detecting deep fakes. The proposed method in this

leverages a feature extraction approach to extract temporal features, which are then fed to a hybrid model based on the combination of CNN and RNN architectures. The hybrid model achieves accuracies of 92.3% which is higher than 66.26%, 91.21%, and 79.49% on the DFDC, FF++, and Celeb-DF datasets respectively. Compared to the results reported in the previous work we discussed, the proposed method achieves lower accuracy scores on the DFDC and Celeb-DF datasets, but a higher accuracy score on the FF++ dataset. The proposed methods use a smaller sample size of ≤ 150 samples (frames) for training and evaluation.

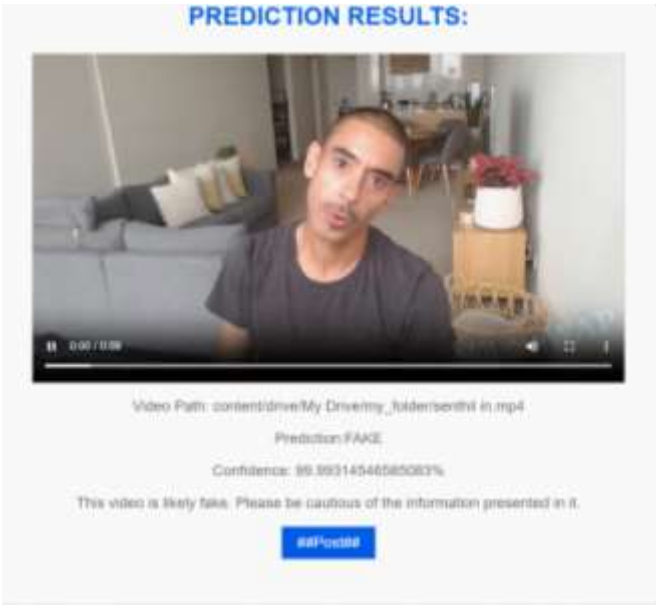

Fig.6.1 Prediction of Fake video

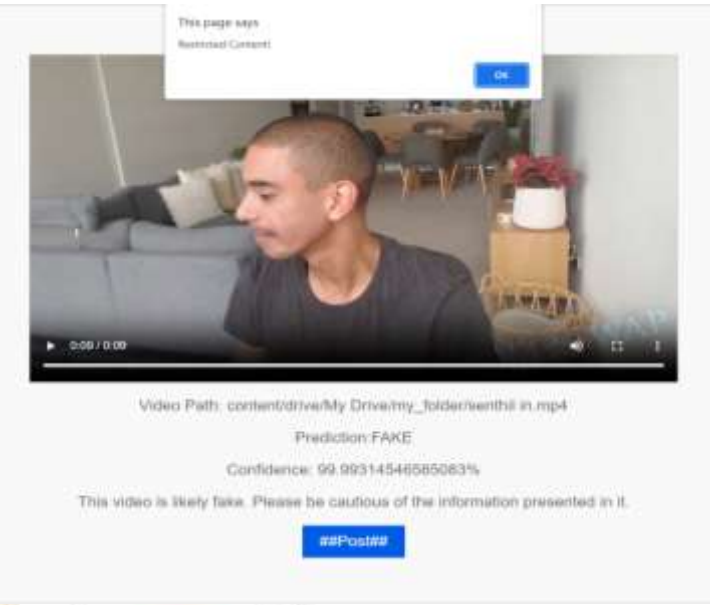

Fig.6.2 Restriction on uploading Fake video

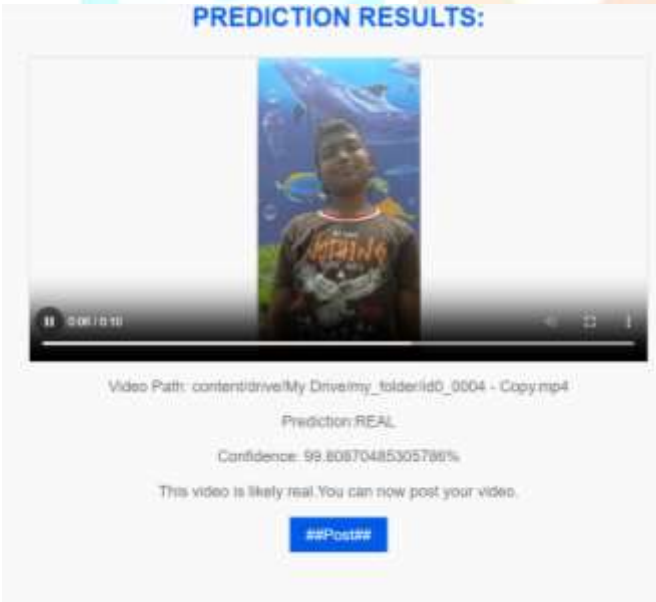

Fig.6.3 Prediction of Real video

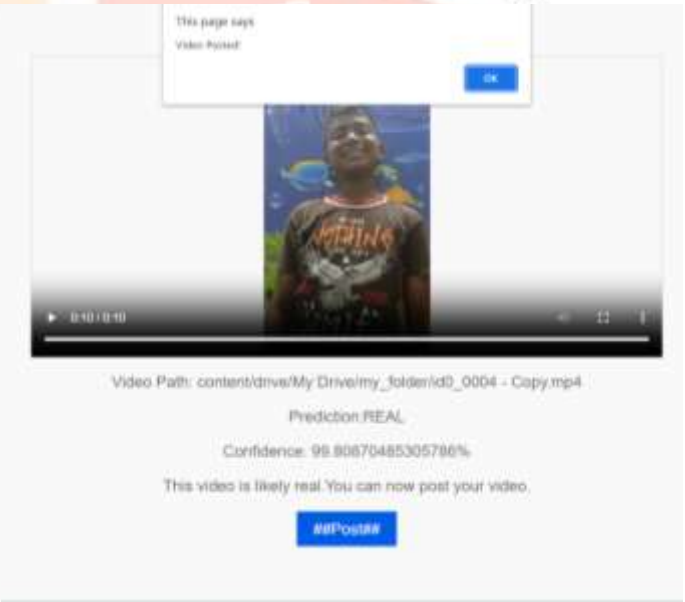

Fig.6.4 Successful Authentication of Real video

LSTM network takes as input a sequence of video frames and produces a probability that the video is real or fake. This probability can then be compared to a threshold to make a binary classification decision. There are many variations on this basic approach, depending on factors such as the size and complexity of the LSTM network, the features that are extracted from the video frames and the screenshots mentioned in Fig 6.1, 6.2, 6.3, 6.4, The idea behind using LSTMs for fake video detection is to treat each video frame as a sequence of pixels, and to feed these sequences into the LSTM network. The LSTM network can then learn to identify patterns in the pixel sequences that are indicative of fake videos.

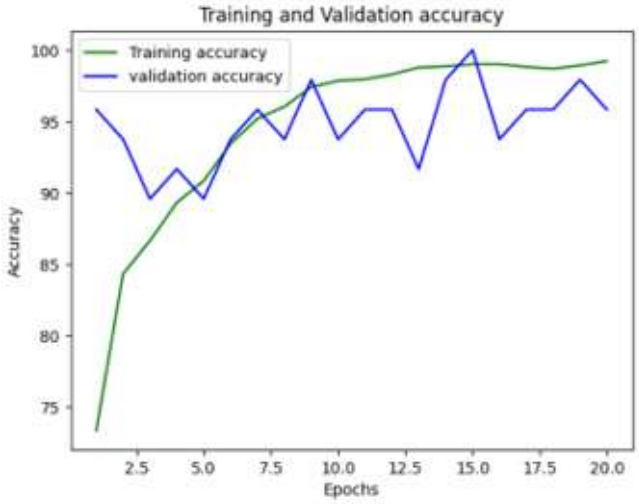

Fig6.5 Accuracy rate of the Trained Model

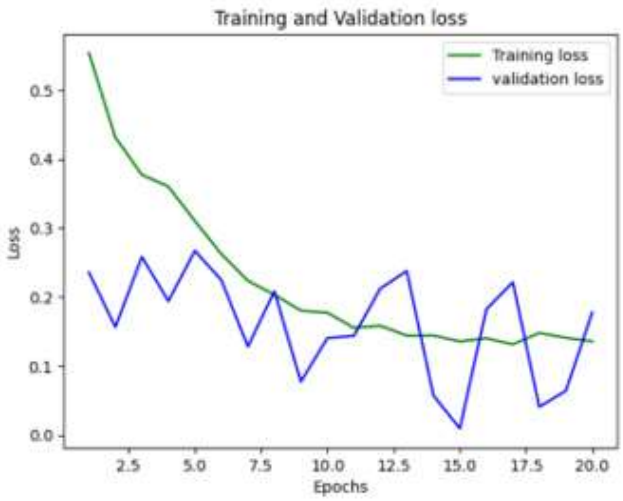

Fig6.6 Loss rate of the Trained Model

Detecting fake videos can be a complex task that typically involves training a deep learning model on a large dataset of both real and fake videos. During training, the model tries to learn to distinguish between real and fake videos based on various features such as pixel values, motion, and audio. The training loss is a metric which is mentioned in fig 6.5 that measures how well the model is fitting the training data. Typically, during training, the loss should decrease as the model learns to better distinguish between real and fake videos. However, it's important to monitor the loss carefully, as a very low training loss can sometimes indicate that the model is overfitting to the training data, and may not generalize well to new, unseen data. The validation accuracy is a metric that measures how well the model performs on a separate set of validation data, which is usually held out from the training data. This metric is important because it gives an estimate of how well the model is likely to perform on new, unseen data. In the case of detecting fake videos, the validation accuracy mentioned in Fig 6.6 would measure how well the model is able to correctly identify fake videos that it has not seen before. Ideally, during training, we would like to see the training loss decrease over time, while the validation accuracy increases. When the dataset changes the accuracy increases up to 95.83%. This indicates that the model is learning to generalize well to new data, and is not overfitting to the training data. However, it's important to carefully monitor both metrics, and make adjustments to the model architecture or training procedure as needed to achieve the best performance.

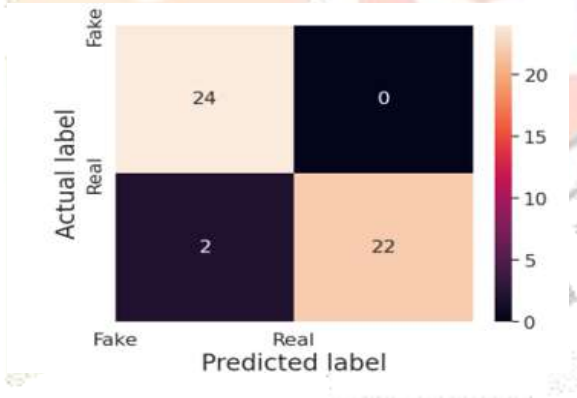

Fig 6.7 Confusion Matrix for Testing

A confusion matrix is a table used to evaluate the performance of a classification model, mentioned in the fig 6.7 such as a fake video detection model. The confusion matrix displays the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) produced by the model. In the context of fake video detection, a true positive represents a fake video that was correctly identified as fake, a false positive represents a real video that was incorrectly identified as fake, a true negative represents a real video that was correctly identified as real, and a false negative represents a fake video that was incorrectly identified as real. The confusion matrix can be used to calculate various performance metrics such as accuracy, precision, recall, and F1 score. For example, accuracy is calculated as (TP+TN)/(TP+FP+TN+FN), precision is calculated as TP/(TP+FP), recall (also known as sensitivity) is calculated as TP/(TP+FN), and F1 score is a weighted average of precision and recall. By examining the confusion matrix and the associated performance metrics, we can gain insights into the strengths and weaknesses of the fake video detection model. For example, if the model has a high false positive rate (i.e., it incorrectly identifies many real videos as fake), we might investigate ways to improve its ability to distinguish between real and fake videos. Similarly, if the model has a high false negative rate (i.e., it incorrectly identifies many fake videos as real), we might investigate ways to improve its sensitivity to fake videos.

REFERENCES

[1] Agarwal, Shruti et al. "Watch Those Words: Video Falsification Detection Using Word-Conditioned Facial Motion." *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (2021): 4699-4708.

[2] N. Khatri, V. Borar, and R. Garg, "A Comparative Study: Deepfake Detection Using Deep-learning," 2023, doi: 10.1109/Confluence56041.2023.10048888.

[3] V. N. Tran, S. H. Lee, H. S. Le, B. S. Kim, and K. R. Kwon, "Learning Face Forgery Detection in Unseen Domain with Generalization Deepfake Detector," in *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*, 2023, vol. 2023-January, doi: 10.1109/ICCE56470.2023.10043436.

[4] V. H and T. G, "Antispoofing in face biometrics: a comprehensive study on software-based techniques," *Comput. Sci. Inf. Technol.*, vol. 4, no. 1, 2023, doi: 10.11591/csit.v4i1.p1-13.

[5] P. Gupta, C. Singh Rajpoot, and A. Professor, "A Deep Learning Technique based on Generative Adversarial Network for Heart Disease Prediction," doi: 10.4186/ej.20xx.xx.x.xx.

[6] J. Peng, M. Sun, Z. Zhang, T. Tan, and J. Yan, "Efficient neural architecture transformation search in channel-level for object detection," in *Advances in Neural Information Processing Systems*, 2019, vol. 32.

[7] Z. Shang, H. Xie, L. Yu, Z. Zha, and Y. Zhang, "Constructing Spatio-Temporal Graphs for Face Forgery Detection," *ACM Trans. Web*, 2023, doi: 10.1145/3580512.

[8] A. Maclaughlin, J. Dhamala, A. Kumar, S. Venkatapathy, R. Venkatesan, and R. Gupta, *Evaluating the Effectiveness of Efficient Neural Architecture Search for Sentence-Pair Tasks*. .

[9] A. Hesham, Y. Omar, E. El-fakharany, and R. Fatahillah, "A Proposed Model for Fake Media Detection Using Deep Learning Techniques," in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 152, 2023.

[10] R. M. Jasim and T. S. Atia, "An evolutionary-convolutional neural network for fake image detection," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 29, no. 3, 2023, doi: 10.11591/ijeecs.v29.i3.pp1657-1667.

[11] P. Pei, X. Zhao, Y. Cao, and C. Hu, "Visual Explanations for Exposing Potential Inconsistency of Deepfakes," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2023, vol. 13825 LNCS, doi: 10.1007/978-3-031-25115-3_5.

[12] C. B. Miller, "Technology and the Virtue of Honesty," in *Technology Ethics: A Philosophical Introduction and Readings*, 2023.

[13] W. Lu *et al.*, "Detection of Deepfake Videos Using Long-Distance Attention," *IEEE Trans. Neural Networks Learn. Syst.*, 2023, doi: 10.1109/tnnls.2022.3233063.

[14] Q. Xu, H. Qiao, S. Liu, and S. Liu, "Deepfake detection based on remote photoplethysmography," *Multimed. Tools Appl.*, 2023, doi: 10.1007/s11042-023-14744-z.

[15] Y. Patel *et al.*, "An Improved Dense CNN Architecture for Deepfake Image Detection," *IEEE Access*, 2023, doi: 10.1109/ACCESS.2023.3251417.