iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.61527

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Deepfake Video Detection UsingRes-Next CNN and LSTM

Dr. Suresh M B[1], Likhith P[2], Chethan Gowda R[3]
[1]*Professor and HOD,* [2]*B.E Student,* [3]*B.E Student Department of ISE, East West Institute of Technology, Bengaluru, India*

Abstract: ***Deepfake videos are becoming more and more common in the digital age, which has led to major worries about how they can undermine the trustworthiness of visual media. With deep learning algorithms' rising processing power, producing lifelike human-synthesized films or deepfakes has never been easier. Disinformation and political unrest can be caused by these videos. A novel deep learning-based method has been created to distinguish between actual and AI-generated fraudulent videos to address this problem. The suggested technique uses Attention-based networks (Res-Next CNN), a kind of deep learning architecture that can selectively focus on significant features in a video, to fine-tune the transformer module to search for new sets of feature space to detect false images.***
Index Terms: Fake video detection, Res-Next CNN, LSTM

I. INTRODUCTION

Deepfake films are edited videos that are produced with sophisticated machine learning and artificial intelligence techniques to produce a phony video that appears extremely realistic. We are able to effectively differentiate between pristine and deepfake videos by taking advantage of the limitations of the deepfake generation tools. The existing deep fake creation techniques leave some distinct artifacts in the frames throughout the production process that may not be evident to humans but that trained neural networks can detect. They can be used to hurt people, sway public opinion, and fabricate news. Deepfake video detection is a challenging task that can be accomplished using a variety of methods. Furthermore, the methods for producing deep fakes also progress with technology, making this a sector that is always changing and requiring new research and development. In fake video analysis, a number of machine learning methods are employed. Here are a few instances: Convolutional Neural Networks (CNNs): CNNs have been demonstrated to be successful at identifying deepfake films. CNNs are frequently utilized in image and video analysis. They are able to spot irregularities in the video's visual material, like strange facial expressions and distortions. [4] Recurrent Neural Networks (RNNs): RNNs can be used to assess the audio content of the video and are frequently employed for sequence analysis. They are able to detect irregularities in tone, pitch, and cadence in the audio. Random Forests: Random Forests are an ensemble learning algorithm that combines multiple decision trees to make a prediction. They can be used to analyse the features of the video and determine if they are consistent with a real video or a fake one.

II. BACKGROUND WORK

A. Resnext

A convolutional neural network (CNN) model called ResNeXt has been used to identify fake and morphed videos. The prominent CNN model ResNet, which has attained a contemporary performance in image recognition, is extended by ResNeXt. Utilizing the ImageNet dataset, ResNeXt outperformed ResNet [6]. The introduction of a "cardinality" parameter, which enables the network to be parallelized across many dimensions, is the primary novelty of ResNeXt. This makes it possible for the network to learn a wider range of properties, which is especially helpful for identifying deepfake movies that could differ slightly yet significantly from authentic videos.

The sequential LSTM is based on the 2048-dimensional feature vectors that follow the last pooling layers of ResNeXt. This model has 32 x 4 dimensions and 50 layers. ResNeXt has been utilized as a feature extractor in deepfake video detection in conjunction with other methods including optical flow analysis and attention mechanisms. To identify deepfake videos, ResNeXt was combined with a temporal attention module. All things considered, ResNeXt is an effective CNN model that has been used to detect deepfake videos.

©IJRASET: All Rights are Reserved

B. LSTM

Recurrent neural networks (RNNs) of the LSTM (Long Short-Term Memory) type have been applied to video detection applications, such as deepfake video detection. Because LSTM can retain a memory of previous inputs and use that knowledge to influence future predictions, it is especially well-suited for processing sequential data, such as video frames. Our goal can be accomplished by employing a single LSTM layer with 2048 latent dimensions, 2048 hidden layers, and a 0.4 dropout probability. The sequential processing of the frames using LSTM allows for the comparison of the frame at 't' seconds with the frame at 't-n' seconds, which allows for the temporal analysis of the video, where n is an arbitrary number. One method for detecting deepfake videos using LSTMs is to feed the video frames into the LSTM as a series of inputs. Based on its prior inputs, the LSTM then learns to predict if each frame is authentic or fraudulent. For instance, an LSTM was used to examine the temporal dependencies in a movie and identify deepfake videos in a 2018 article by Afchar et al. [9]. Using multiple benchmark datasets, the authors demonstrated the excellent accuracy of their technique. LSTMs can also be used in conjunction with other methods, such CNNs, to aid in the detection of moving images. To identify deepfake videos, LSTM was combined with an attention mechanism and a CNN. Modern performance was attained by the LSTM approach on multiple benchmark datasets [10]. All things considered, LSTM is an effective method for identifying videos, and its capacity to record temporal dependencies makes it ideal for identifying deepfake videos. Its performance can be further enhancedby combining it with other methods like CNNs and attention processes.

III. LITERATURE SURVEY

DeepFake is a deep learning-driven video generation technology. The technique takes advantage of the spatiotemporal characteristicsof movies by feeding frame sequences into the model. The method makes use of differences between several frames and lower-level properties in areas of interest. Since altered videos have the potential to have detrimental effects on a variety of industries, including politics, journalism, and entertainment, deepfake video detection is a relatively new field of study that has drawn interest.

These are themain conclusions drawn from current research on deepfake video identification. The two main strategies used by deep fake detection algorithms[13] are: 1) looking for artifacts or irregularities in the video, and 2) examining the characteristics of the person or object in the video. To increase the accuracy of deep fake detection, several academics have suggested using extra characteristics, like audio.

This is so that more evidence for detection can be obtained by analyzing the audio, which is used by some deepfake algorithms to further modify the video. The ever-evolving methods employed by those who create deep fakes present a significant obstacle to deep fake detection. To stay up to date with new developments in deep fake technology, researchers need to regularly upgrade theirdetection algorithms. [14] The absence of standardization in assessment metrics and datasets is another difficulty. This can impede advancement in the field and make comparing the efficacy of various detection techniques challenging. In conclusion, the discipline of deep fake video detection is quickly developing. It uses machine learning algorithms to examine video attributes and identify irregularities or artifacts that point to manipulation. Though there are still obstacles to overcome, recent studies have demonstrated encouraging outcomes in the identification of deepfake movies, and efforts are currently underway to create more potent detection techniques.

IV. PROPOSED SYSTEM

To determine the temporal dependencies between each frame and categorize whether the movies are real or fake, the proposed system uses an LSTM-based RNN to process frame-level features extracted from the videos using a Res-Next Convolution neural network. The transformer module was adjusted to use attention-based networks (Res-Next CNN) to look for new feature space sets in order to identify phony photos.

The suggested approach demonstrated good accuracy on films from many sources, demonstrating its effectiveness in real-time manipulation detection, according to the examination of the hybrid dataset. Deep fakes can be used to propagate misinformation and disrupt political processes, hence the system's capacity to identify manipulations in real time is essential.The experiment's outcomes demonstrated the effectiveness of the suggested strategy. We will give the user access to a scaled-down version of a web-based platform where they can post videos, mark them as authentic or fraudulent, and prevent them from beingshared online. Large apps like Facebook, Instagram, and WhatsApp might include this idea into their software to enable simple pre- detection of distorted or fake films.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

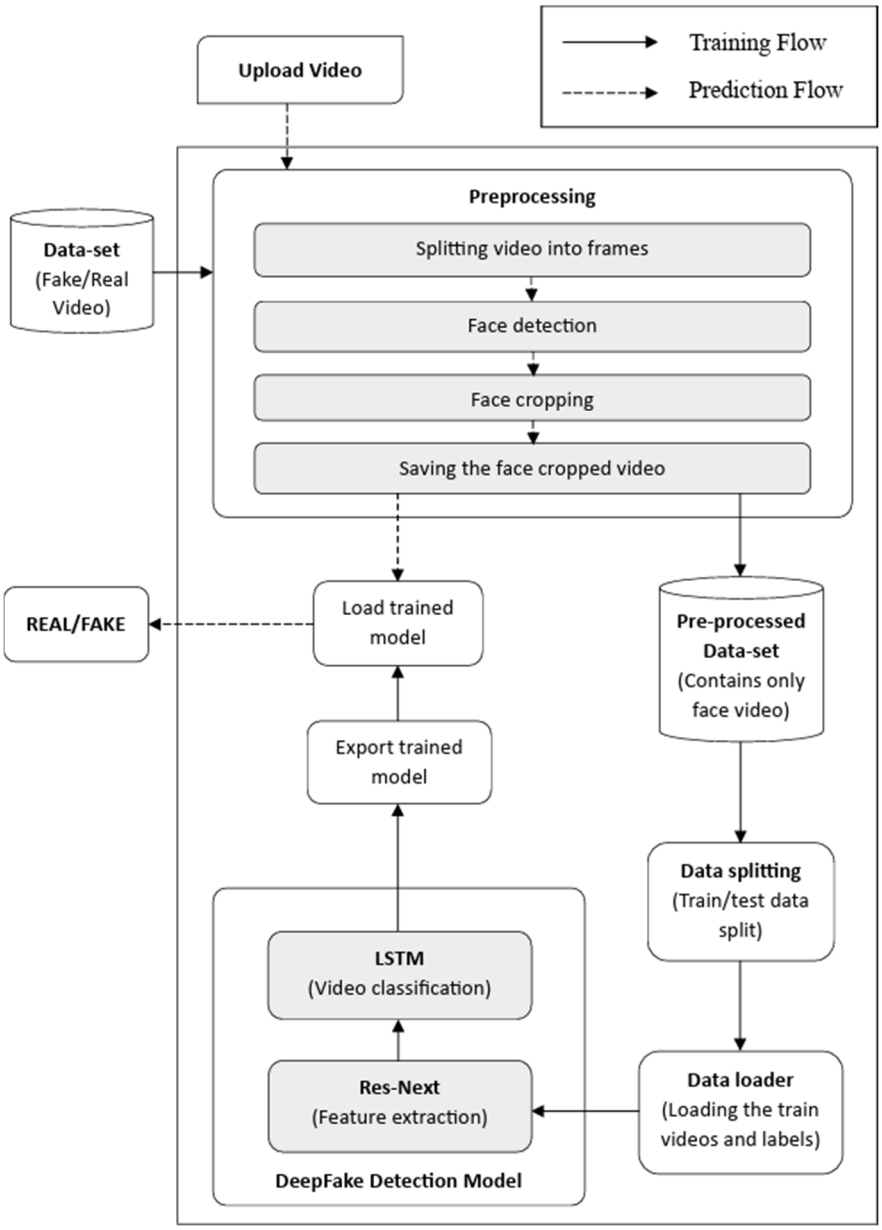


Fig 4.1 Proposed System Architecture

Furthermore, the deployment of such a system onto a web-based platform offers users a practical and accessible means to contribute to the detection and prevention of deepfake dissemination. By allowing users to submit videos, annotate them as authentic or fraudulent, and prevent their spread online, the platform empowers individuals to actively combat the proliferation of misinformation. Integration of this technology into popular social media platforms like Facebook, Instagram, and WhatsApp holds tremendous potential to mitigate the impact of deepfake manipulation on digital ecosystems. This proactive approach not only enhances the resilience of online communities against deceptive content but also fosters a culture of accountability and transparency in digital communication.

A. Dataset

Research on deepfake video detection has made use of a few datasets. The following are a some of the most widely used datasets: Face Forensics ++: Among the biggest and most popular deepfake video datasets is this one. It includes more than 1,000 authentic videos and more than 1,000 deepfake movies produced with a variety of techniques, such as Face2Face, Deep Fake, and Neural Textures.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Another well-liked dataset for identifying deepfake videos is Celeb-DF. It has more than 5,639 deepfake videos made with the Deep Fake technique in addition to over 890 actual footages. The Deep Fake Detection Challenge (DFDC) dataset was produced by Facebook as part of a competition to improve deepfake detection techniques.

Self-generated dataset: We have created this dataset on our own to enhance training and prediction accuracy, as well as to anticipate the detection of fraudulent video in real-time settings and optimize system performance. Usually, these datasets include tagged videos, where each one has a label designating it as authentic or fraudulent. They are frequently employedin the evaluation and training of deepfake video detection models. Nevertheless, there are drawbacks to using these datasetsas well, namely the possibility of bias in the categorization procedure and the narrow range of deepfake films that are included [12]. When utilizing these datasets for deepfake video detection, researchers need to be cautious to take these constraints into account.

V. EXPERIMENT

A hybrid dataset comprising both actual and altered films from many sources was used to train and assess the suggested strategy for identifying fraudulent videos in real-time scenarios. Initially, the films were obtained from a dataset consisting of three sources [1]. DFDC [3] CELEB DF [2] FF++ Lastly, we have our own self-created dataset [4] that we employ to enhance training accuracy and produce real-time video results. Additionally, we combined the several datasets to produce a brand-new dataset. We have taken into consideration 50% real and 50% fake videos in order to prevent the model's trainingbias. The audio-alerted videos in the Deep Fake Detection Challenge (DFDC) dataset [3] are specific to audio alerts; audio deepfakes are not relevant to this study. 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]. [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 6000videos 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. Next, using PyTorch, the model is trained on the videos, and accuracy and loss are verified using a learning rate of 1-e5 and 20 epochs. Finally, the training and testing results are displayed together with a confusion matrix and graph. The trained model has been exported for prediction after training and testing. Ultimately, the user's input is processed using the loaded trained model, and the output is shown together with a confidence level and a prediction as to whether it is true or phony. . This makes their solution a useful tool for identifying and stopping the spread of deepfake content.

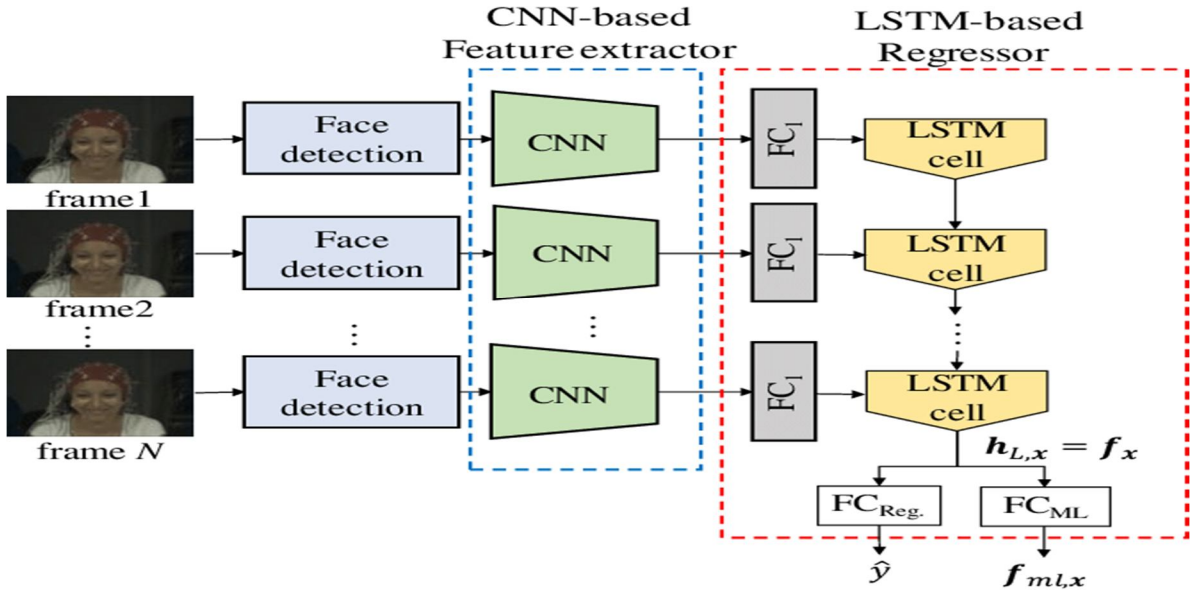


Fig 5.1 System Process Architecture

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

VI. RESULTS AND DISCUSSIONS

The examination of the hybrid dataset demonstrated the effectiveness of the suggested approach in identifying manipulations in real-time circumstances, attaining a high degree of accuracy on movies sourced from many sources. In particular, on the Hybrid dataset evaluation, the suggested strategy yielded an accuracy of 95.83 and a loss value of 0.177. These findings show that, even in practical situations, the suggested method is very successful at identifying video modifications. Because of the system's high accuracy on the hybrid dataset evaluation, deep fakes can be used to counteract the spread of misinformation. Overall, the experiment and analysis of the data indicate that the suggested method is a viable way to identify deep fakes and deal with the problem of misinformation spreading through edited movies. Several factors are involved in the suggested deep learning-based technique for identifying deep fakes; these parameters are utilized to adjust the transformer module and train the LSTM-based RNN and Res-Next CNN models. 1. Learning rate: This parameter controls the amount that the model's parameters are changed while it is being trained. During optimization, it is utilized to regulate the step size that is taken in the gradient's direction. 2. Batch size: During training, the number of samples processed at once is determined by this parameter. More memory is needed for bigger batch sizes, but they can also result in more steady convergence. 3. Number of epochs: The number of times the training data is run through the model is determined by this parameter. 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. While adding more hidden layers can help the model better capture intricate patterns, doing so also raises the possibility of overfitting. 7. Number of neurons: The Res-Next CNN and LSTM-based RNN models' number ofneurons in each hidden layer is determined by this parameter. It is employed to regulate the capacity and sophisticated feature learning capabilities of the model.

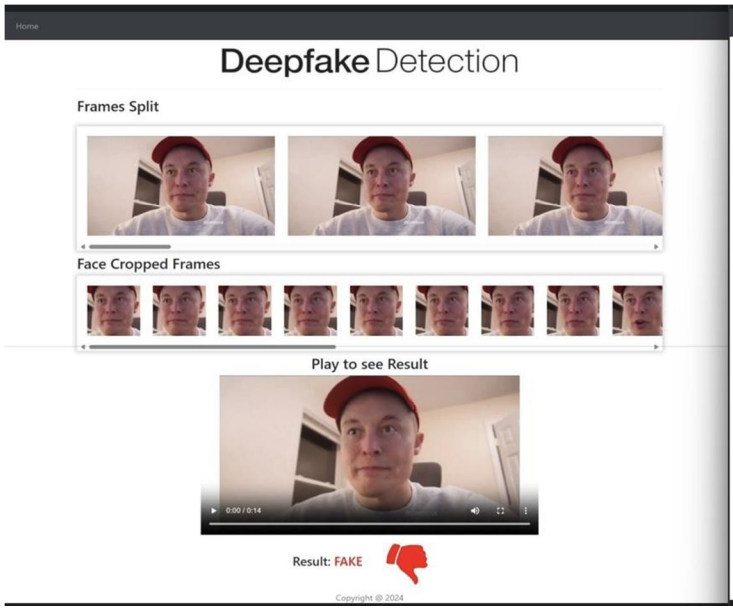
List of Parameters	Face-Forensic++	DeepFake Detection Challenge Dataset (DFDC)	Celeb-DF	Hybrid Dataset (FF +DFDC+Celeb-DF +Self created)
Learning rate	0.001	0.0005	0.0001	0.0001
Batch size	32	64	128	4
Number of epochs	100	50	200	20
Dropout rate	0.5	0.2	0.3	0.4
Weight decay	0.01	0.001	0.0001	0.003
Accuracy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

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

As we can see, there are variations in the dataset values for every parameter when comparing them. For example, the batch size varies from 16 to 128; the number of hidden layers varies from 3 to 6. The learning rate varies from 0.0001 to 0.005. These changes show that the best values for these parameters might vary depending on the particular application and dataset that are utilized, in addition to other considerations like model design and processing resources. All of these example value sets nevertheless show a reasonable range of values that might be employed in a deep learning-based technique for identifying deepfakes. The 92.3% accuracy achieved by the hybrid model is superior to 66.26%, 91.21%, and 79.49% on the DFDC, FF++, and Celeb-DF datasets, in that order. The suggested strategy achieves lower accuracy scores on the DFDC and Celeb-DF datasets, but a higher accuracy score on the FF++ dataset in comparison to the results presented in the prior work we addressed. For training and evaluation, the suggested methods employ a reduced sample size of ≤ 150 samples (frames).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Result predictions of deepfake video detections where authentication is mentioned if video is legal or not



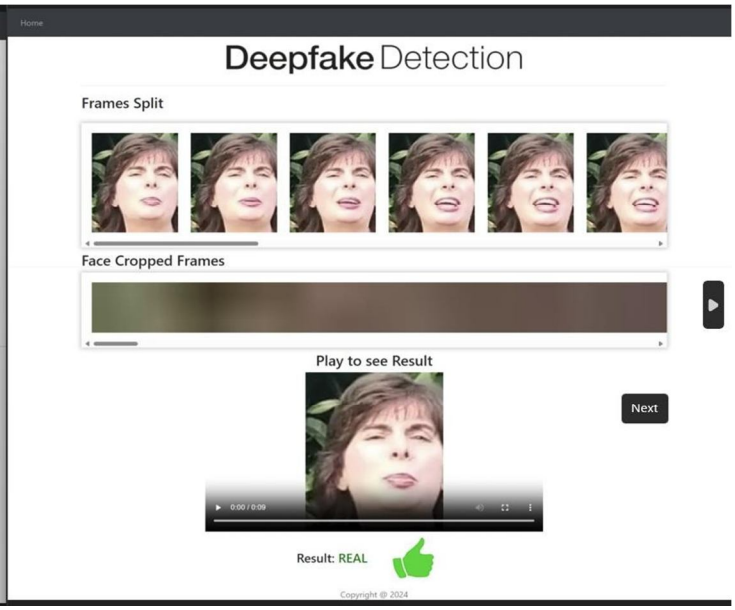
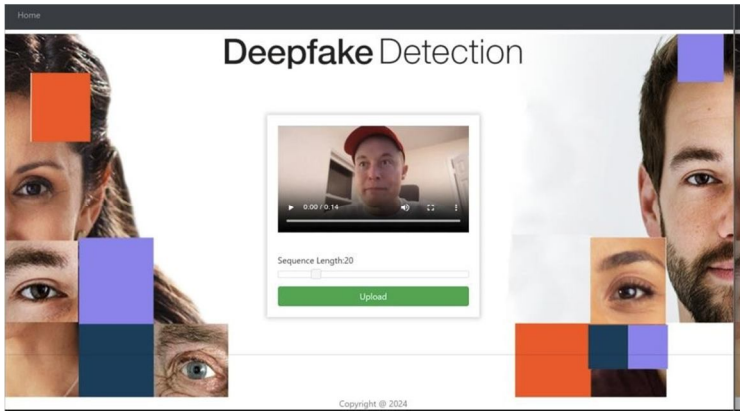


Fig 6.1 Deepfake model depicting the video uploaded is fake. Fig 6.2 Deepfake model depicting the video uploaded is real.

The above figure 6.1 and 6.2 depicts the software being developed for authentication of real video and restriction of deepfake videos. The figure shows that the uploaded video is split into specified frames and the face-cropped frames before producing the result video.



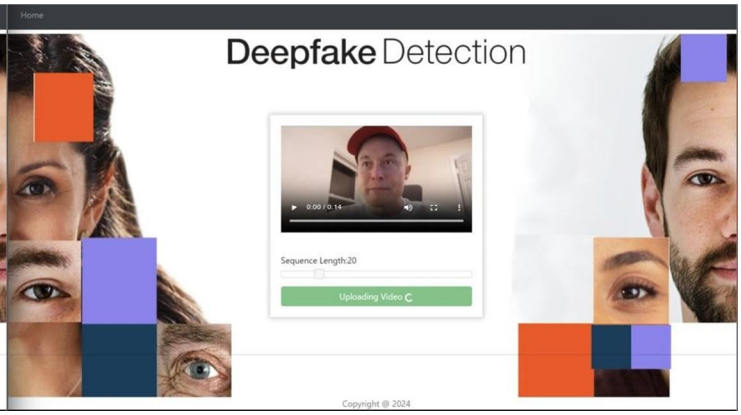


Fig 6.3 Deepfake model depicting the process of specifying number of frames. Fig 6.4 Deepfake model depicting the process of video being uploaded.

The above figure 6.3 and 6.4 depicts the software being developed for authentication of real video and restriction of deepfake videos. The figure shows the process of uploading the video to the model for detecting whether the video is real or fake.
A series of video frames is fed into an LSTM network, which outputs a probability as to whether the video is authentic or fraudulent. The choice to classify something as binary can then be made by comparing this likelihood to a threshold. Numerous versions exist for this fundamental method, contingent upon variables like the LSTM network's size and complexity, the features collected from the video frames, and the screenshots referenced in Figures 6.1, 6.2, 6.3, and 6.4. Each video frame is viewed as a sequence of pixels, and these sequences are fed into the LSTM network in order to detect false videos.



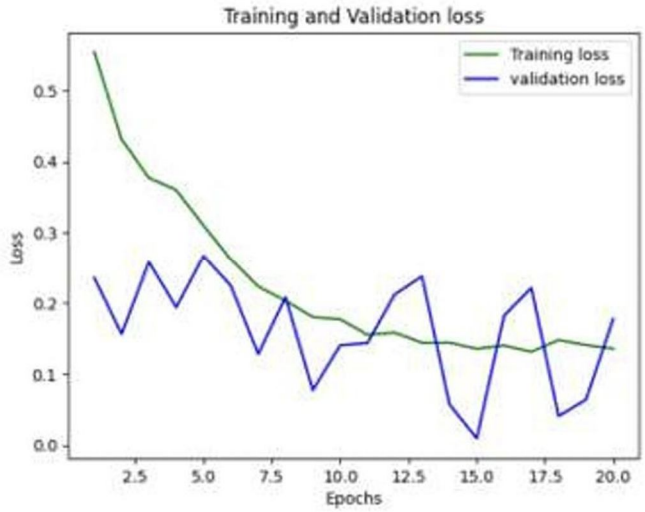


Fig 6.5 Graph depicting the Training and Validation accuracy. Fig 6.6 Graph depicting the Training and Validation loss.

A deep learning model is usually trained on a sizable dataset of both real and false videos in order to detect phony videos, which can be a challenging process. Using a variety of variables, including pixel values, motion, and audio, the model learns during training to differentiate between actual and false videos. Figure 6.5 mentions a parameter called training accuracy and Figure 6.6 mentions a parameter called training loss that quantifies how well the model fits the training set. The loss should normally go down as the model gets stronger at differentiating between actual and fraudulent videos during training. A extremely low training loss, however, may occasionally suggest that the model is overfitting to the training set and may not translate well to new situations, thus it's crucial to closely watch the loss. This measure is significant because it provides an approximation of the model's expected performance on brand-new, untested data. Regarding the identification of fraudulent movies, the validation accuracy denoted in Fig. 6.6 would gauge the model's proficiency in accurately identifying fraudulent videos that it hasn't encountered previously. During training, we should ideally see a gradual decrease in training loss and an improvement in validation accuracy. There is a 95.83% increase in accuracy whenthe dataset is altered. This shows that the model is not overfitting to the training set and is instead learning how to generalize well to new data. But it's crucial to keep a close eye on these measures and modify the model's design or training process as necessary.



Fig 6.7 Confusion matrix for Testing the model performance.

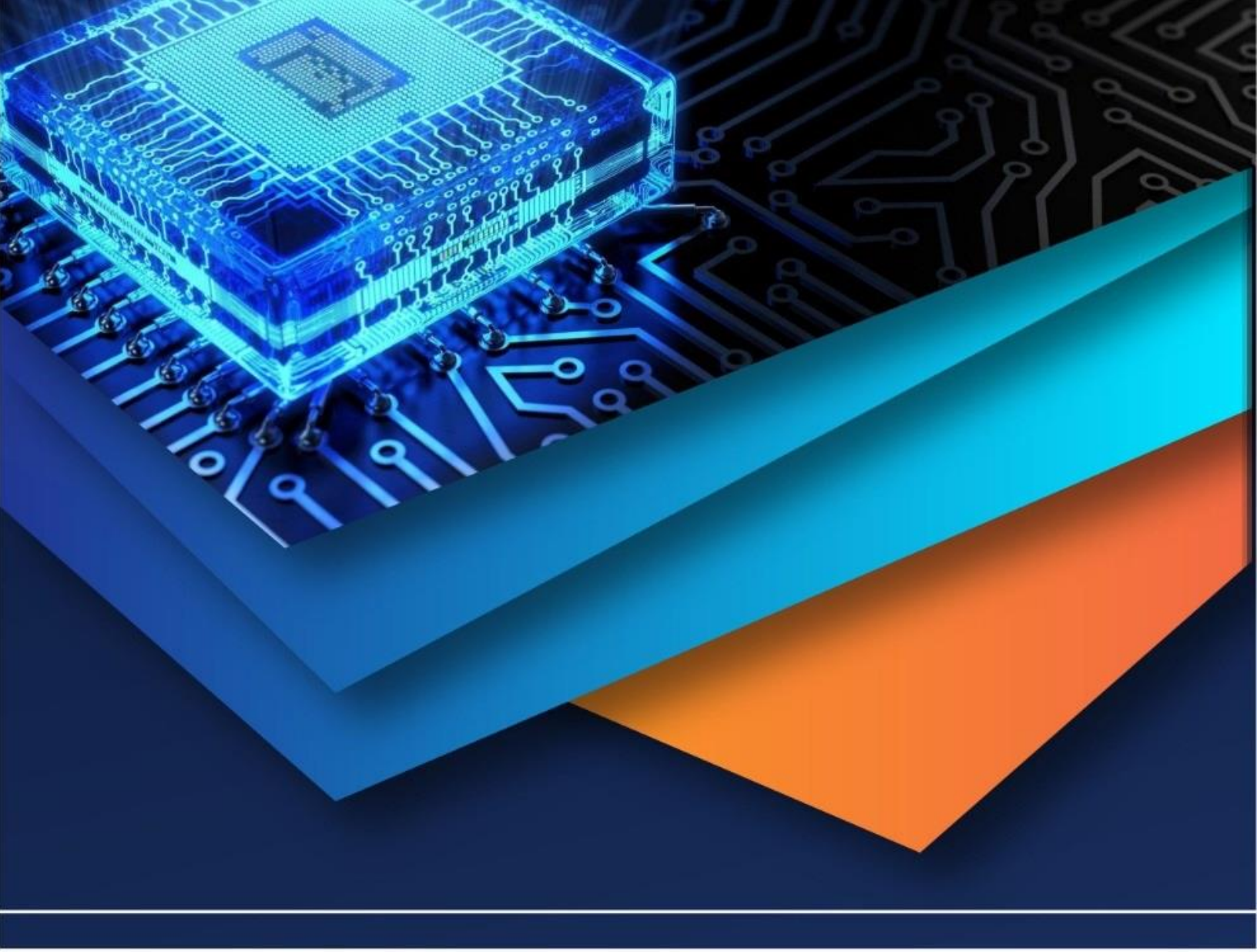
As seen in fig. 6.7, a confusion matrix is a table that's used to assess how well a classification model—like the false video detection model—performs. The number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) that the model produces is shown in the confusion matrix. Within the realm of fake video detection, a genuine video that was accurately identified as fake is called a true positive; a real video that was mistakenly identified as fake is called a false positive; a real video that was correctlyidentified as real is called a true negative; and a fake video that was mistakenly identified as real is called a false negative. 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. The confusion matrix and related performance measures can be examined to learn more about the advantages and disadvantages of the false video detection model. For instance, we might look into ways to strengthen the model's capacity to discern between real and fraudulent movies if it has a high false positive rate—that is, if it mistakenly labels a large number of actual videos as fake. Similarly, we could look into ways to increase the model's sensitivity to phony videos if it has a high false negative rate—that is, if it mistakenly recognizes a large number of bogus videos as real and make the model to perform as required by the users in future models for detection of deepfake videos.

VII. CONCLUSION

In conclusion, the fusion of ResNet CNN and LSTM for deepfake video detection presents a promising approach with severalnotable outcomes. The combination of these two architectures leverages the strengths of both convolutional neural networks (CNNs) in extracting spatial features and long short-term memory (LSTM) networks in capturing temporal dependencies within video sequences. In summary, the integration of ResNet CNN and LSTM represents a significant step forward in the quest to combatdeepfake manipulation. While further research and refinement are warranted, this approach holds great promise in bolstering the security and integrity of multimedia content in an era increasingly defined by digital deception.

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.2023.325141.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)