DEEPFAKE DETECTION USING DEEP LEARNING

***Project Reference No.:** 47S_BE_2416*

***College** : BGS Institute of Technology, Mandya*
***Branch** : Department of Electronics and Communication Engineering*
***Guide(s)** : Mrs. Nethravathi H M*
***Student(S)** : Mr. Akhilesh S B*
Ms. Harshitha M N
Mr. Hrishikesh Kumar K N
Ms. Manjushree H B

Keywords:

Deepfake detection, Long Short-Term Memory (LSTM), Deep Learning

Introduction:

Deepfakes, hyper-realistic manipulated videos and images, have become a growing concern. These creations, often powered by deep learning techniques themselves, can be nearly indistinguishable from reality. To combat this, deep learning is being harnessed for deepfake detection.

Deep learning models are trained on massive datasets of real and deepfake content. By analysing subtle inconsistencies in things like facial movements, blinking patterns, and lighting, these models can learn to identify deepfakes with impressive accuracy. This technology is crucial for safeguarding online information and ensuring trust in visual media.

Deepfake detection using deep learning, particularly Long Short-Term Memory (LSTM) networks, represents a cutting-edge approach in combating the proliferation of synthetic media manipulation. LSTM networks, a type of recurrent neural network (RNN), excel at capturing temporal dependencies and long-range dependencies in sequential data, making them well-suited for analysing the temporal dynamics present in deepfake videos. Previous research in this domain has demonstrated the effectiveness of LSTM-based models in discerning subtle inconsistencies and artifacts characteristic of deepfake content, thereby providing a robust defense mechanism against malicious manipulation. By leveraging large-scale datasets of both authentic and manipulated media, researchers have trained and fine-tuned LSTM architectures to achieve remarkable accuracy in detecting deepfakes across various modalities, including video, audio, and images.

Objectives:

- To implement a deep learning model utilizing Convolutional Neural Networks (CNN's), Recurrent Neural Networks (RNN's), Long Short-Term Memory (LSTM) network suited for analysing multimedia data and detecting deep fake content.
- Provide an easy-to-use system to upload the video and distinguish whether the video is real or fake.
- To significantly improve the accuracy and efficiency of deepfake detection in videos, surpassing the capabilities of existing methods.
- Explore the practical applications of deepfake detection: This could involve integrating the detection model into platforms like social media networks, video editing software, or security systems

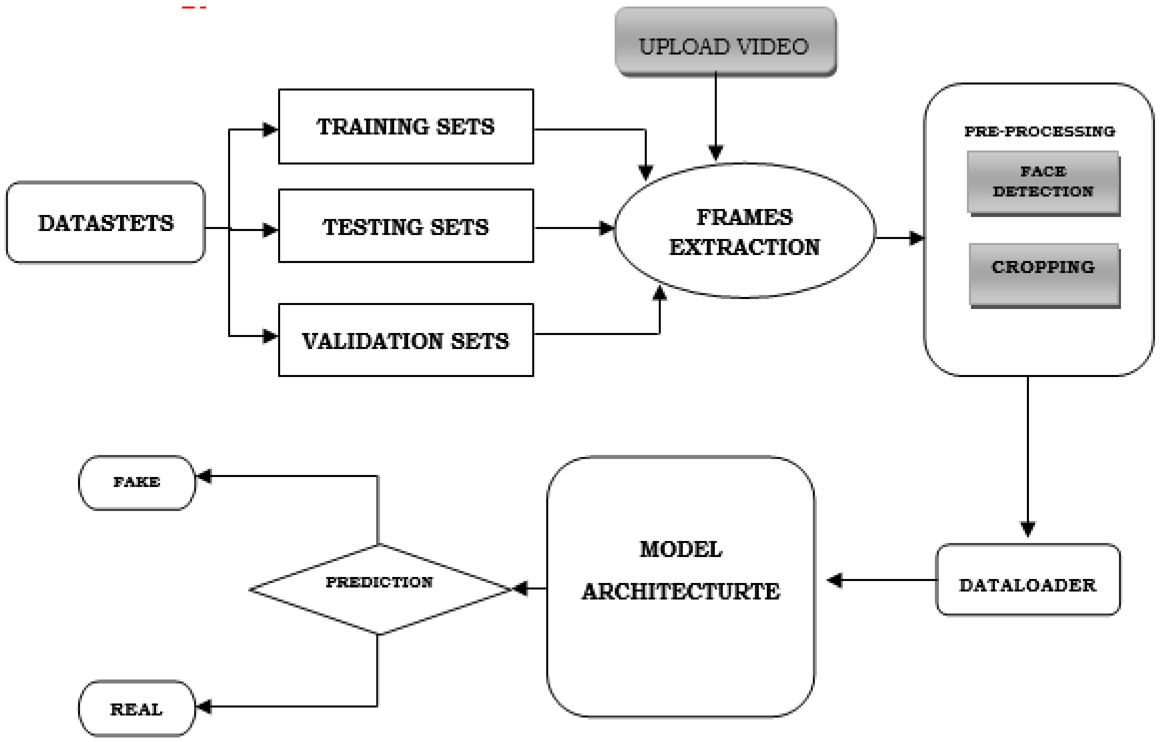
Methodology:

Dataset Collection: The dataset must encompass a substantial number of real and deepfake images and videos, covering various scenarios, lighting conditions, and subjects. For making the model efficient for real time prediction. We have gathered the data from different available data-sets like Face Forensic++(FF), Deepfake detection challenge (DFDC), and Celeb-DF. Further we have mixed the dataset the collected datasets and created our own new dataset, to accurate and real time detection on different kind of videos.

Preprocessing: It involves extracting frames from videos, normalizing pixel values, and augmenting the data to enhance the model's ability to generalize across different conditions. The first steps in the preprocessing of the video are to split the video into frames. Each frame of the video is cropped along the face once the face is identified and separated into individual frames. 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.

Model Training: The features were extracted at the frame level using the pre-trained ResNext CNN model, and an LSTM network was trained to categorize the video as either pristine or deepfake based on the features that were extracted. Insights into the process of splitting data into training and validation sets along with the training process, optimization strategies, and fine tuning specific to LSTM network.

Evaluation Metrics: Introduction to key evaluation metrices, such as accuracy, precision recall, tailored for assessing the efficiency of LSTM-based deepfake detection.



Conclusion:

In conclusion, this work has demonstrated the effectiveness of deep learning for deepfake detection. Our Deep Learning model achieved high accuracy in identifying deepfakes. This project contributes to the ongoing fight against manipulated media by providing a robust and accurate deep learning-based solution. Further exploration is warranted in areas like generalizability across datasets and continuous adaptation to counter deepfake evolution.

Scope for future work:

1. **Advanced Neural Network Architectures:** Continued research into more sophisticated neural network architectures tailored specifically for deepfake detection. This could involve exploring novel network designs, such as attention mechanisms or graph neural networks, to better capture the intricate patterns present in deepfake videos.

2. **Multimodal Fusion:** Integrating multiple modalities of information, such as audio, visual, and temporal cues, to enhance the robustness of deepfake detection systems. Combining information from different sources can provide a more comprehensive understanding of the content and improve the accuracy of detection.

3. **Weakly Supervised and Self-Supervised Learning:** Investigating techniques that require less labelled data for training, such as weakly

supervised or self-supervised learning approaches. This can help alleviate the significant labelling burden associated with deepfake detection datasets and make the models more adaptable to new and emerging types of deepfakes.

4. **Real-Time Detection:** Optimizing deepfake detection algorithms for real-time processing to enable their integration into live video streaming platforms and social media networks. This requires efficient model architectures and inference techniques capable of operating with low latency while maintaining high detection accuracy.