Development of AI/ML based solution for detection of face-swap based deep fake videos Project Id:-21

Vaishnavi Patil
Akansha Patil
Sakshi Marathe
Prajwal Chaudhari
Guided by Prof. Dr.A.D.Waghmare

SSBT's College of Engineering & Technology,
Bambhori, Jalgaon - 425 001, Maharashtra, India

Outline

Introduction

- Deepfakes represent one of the most advanced applications of artificial intelligence
- Where technology is used to create highly convincing yet fabricated video and audio content. This innovation has the power to entertain and educate, but it also carries serious risks that we must address.
- This innovation has the power to entertain and educate, but it also carries serious risks that we must address.

Motivation

- **Security and Privacy:-** Deepfakes can lead to privacy violations and security threats, such as impersonation.
- **Misinformation and Disinformation:-** Deepfakes can be used to create misleading content, potentially impacting public opinion, political events, and social trust. Detecting them helps maintain the integrity of information.
- **Public Awareness and Education:-** Developing detection tools can help raise awareness about deepfakes, fostering a more informed public that can critically evaluate media content.

Problem Statement

- The problem at hand is the proliferation of deepfake content, which poses a substantial threat to the authenticity and trustworthiness of multimedia in various domains.
- The ultimate goal is to create a robust, reliable, and scalable solution that can assist platforms, organizations, and users in verifying the authenticity of video content and preventing the spread of false information.

Scope

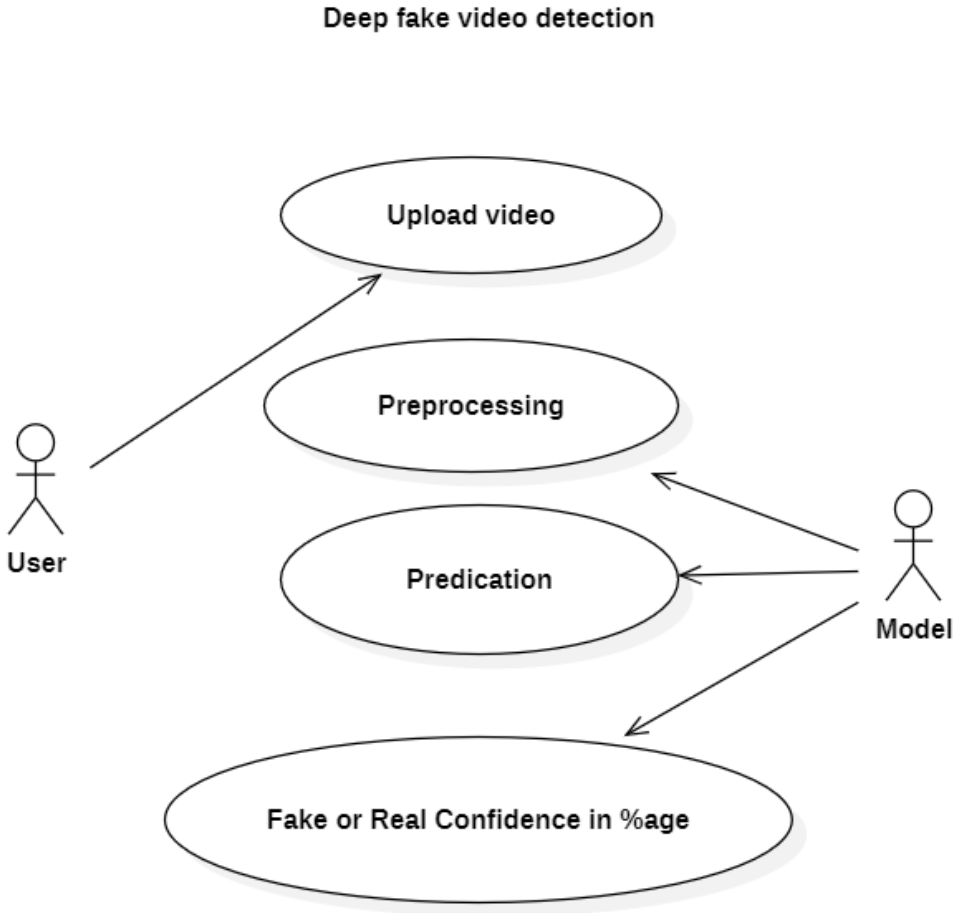
Scope for Identification of Deepfake Videos Using AI/ML-Based Solutions

- **Dataset Development**
- **Research and Development**
- **Model Development**
- **User Interface Development**
- **Real-Time Detection**

Objectives

- **Real-Time Detection:-** Create algorithms capable of detecting deepfakes in real-time, enabling immediate verification of video authenticity for applications in media, security, and social platforms.
- **User-Friendly Interface:-** Develop an intuitive interface for users to upload and analyze videos for deepfake detection, making the technology accessible to non-experts.
- **Public Awareness and Education:-** Create educational resources that inform the public about deepfake and the importance of verification, fostering a more informed digital community.

Use Case Diagram



Sequence Diagram

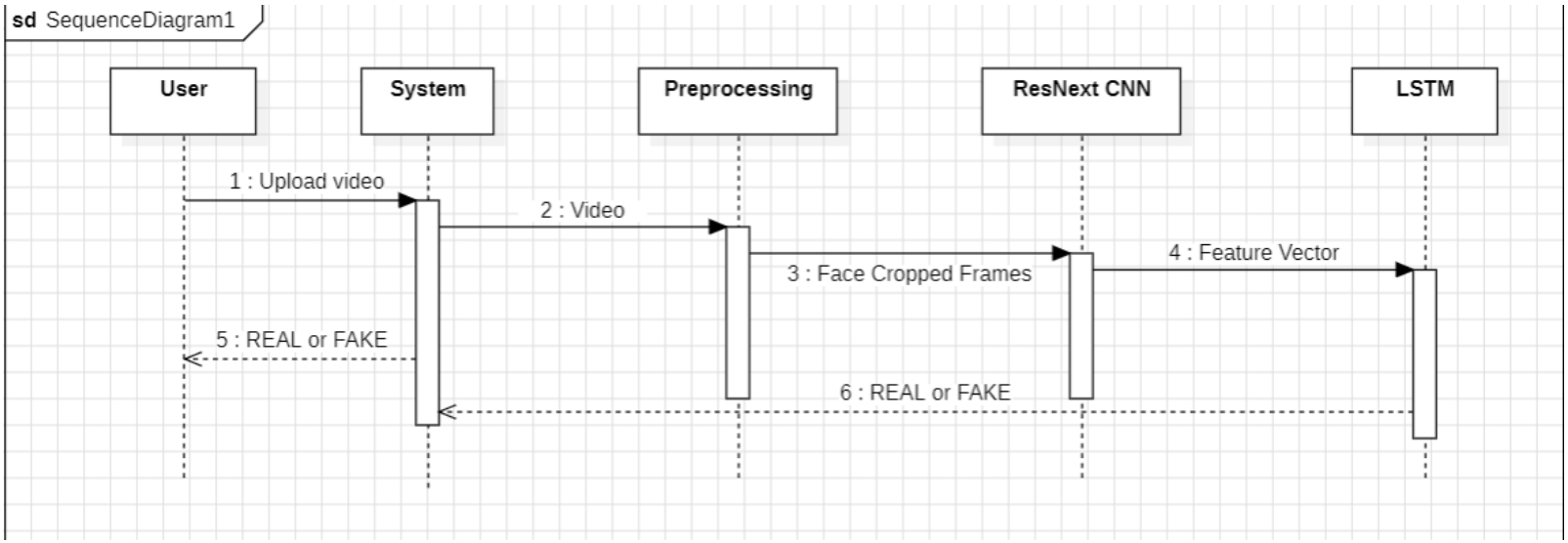


Figure: Sequence Diagram

Class Diagram

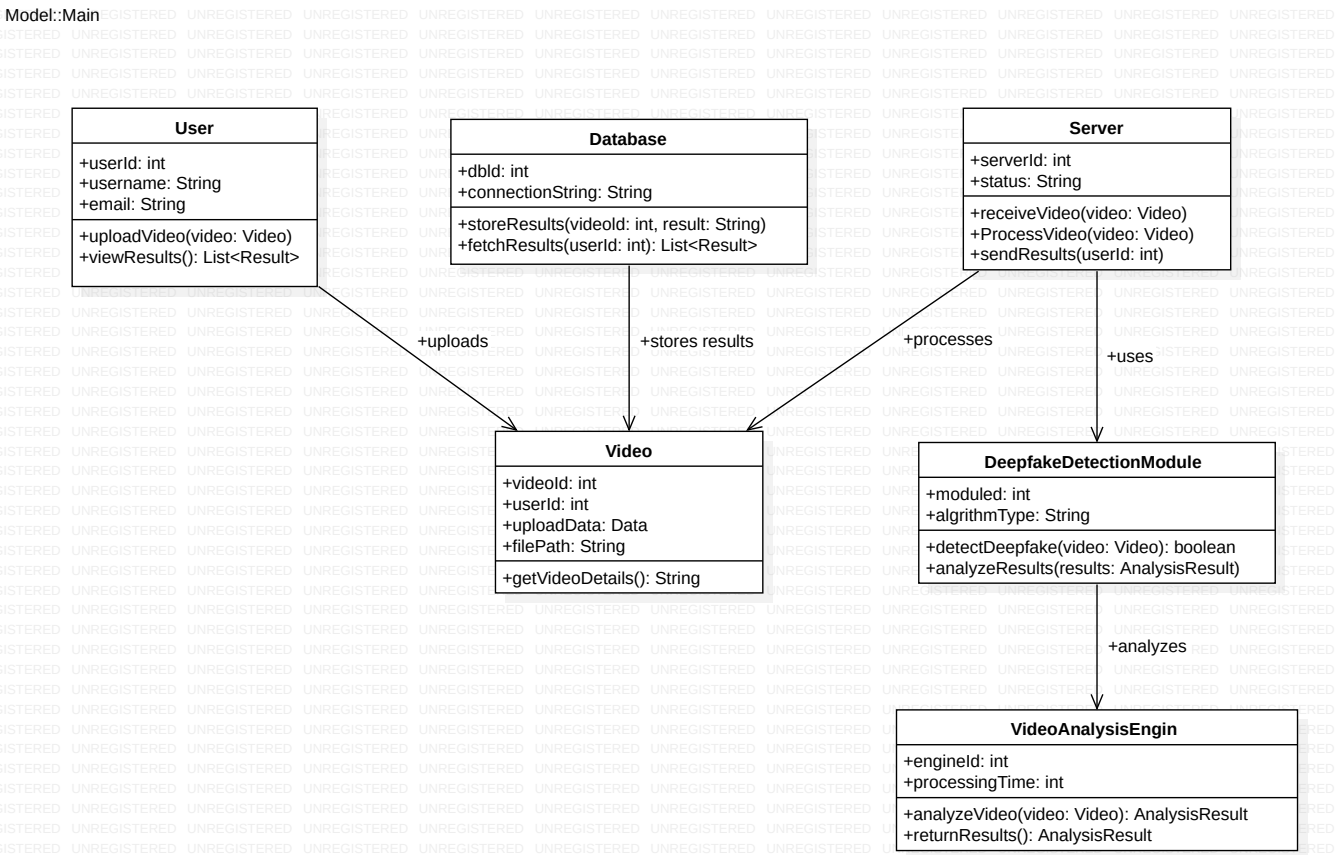
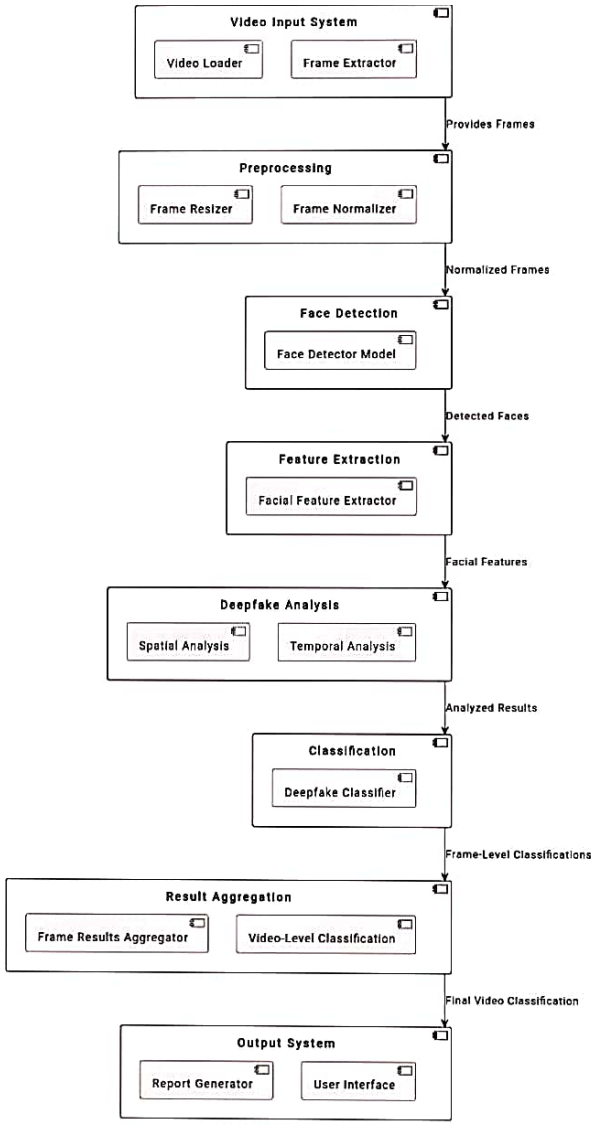


Figure:Class Diagram

Component Diagram



System Architecture

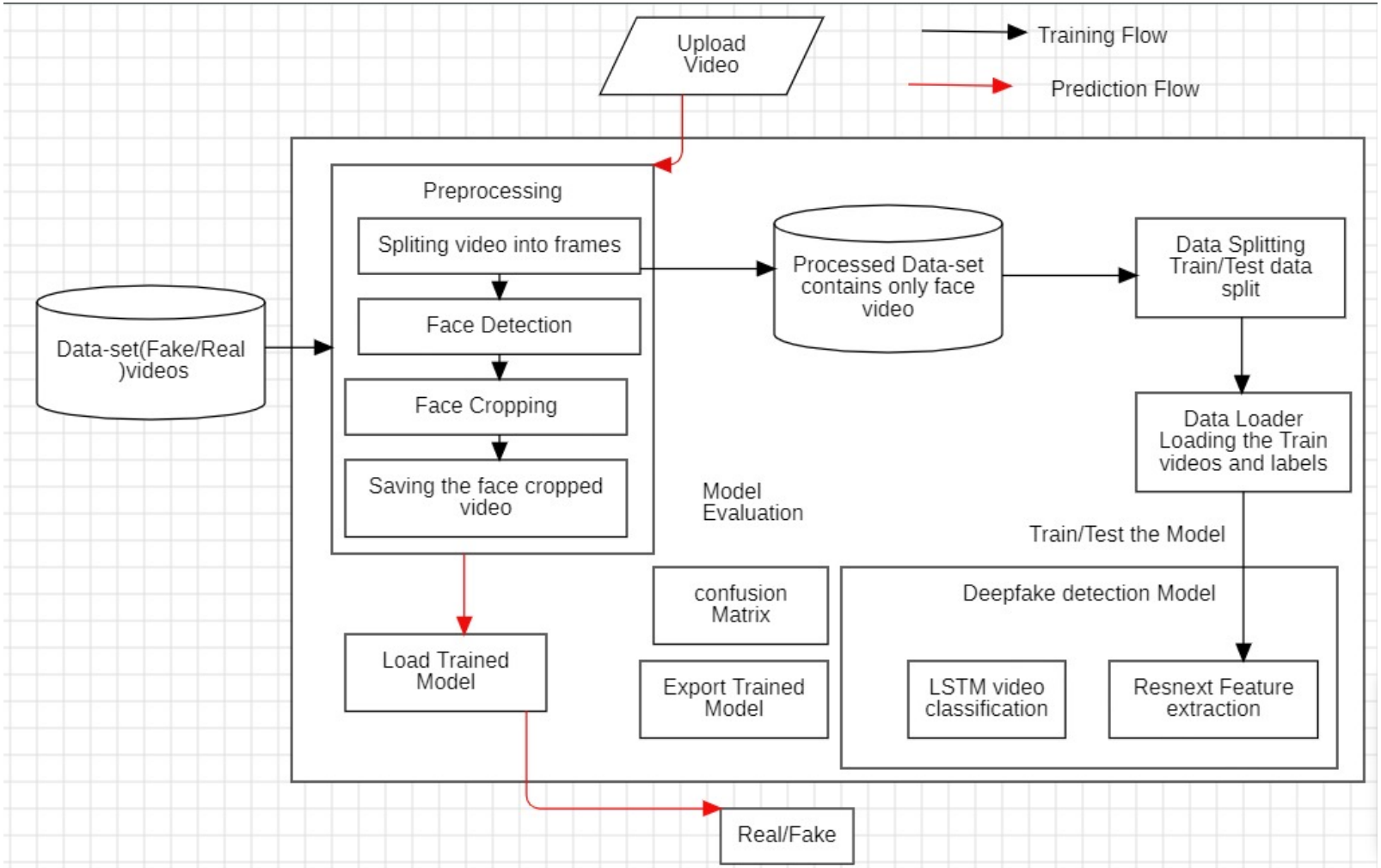
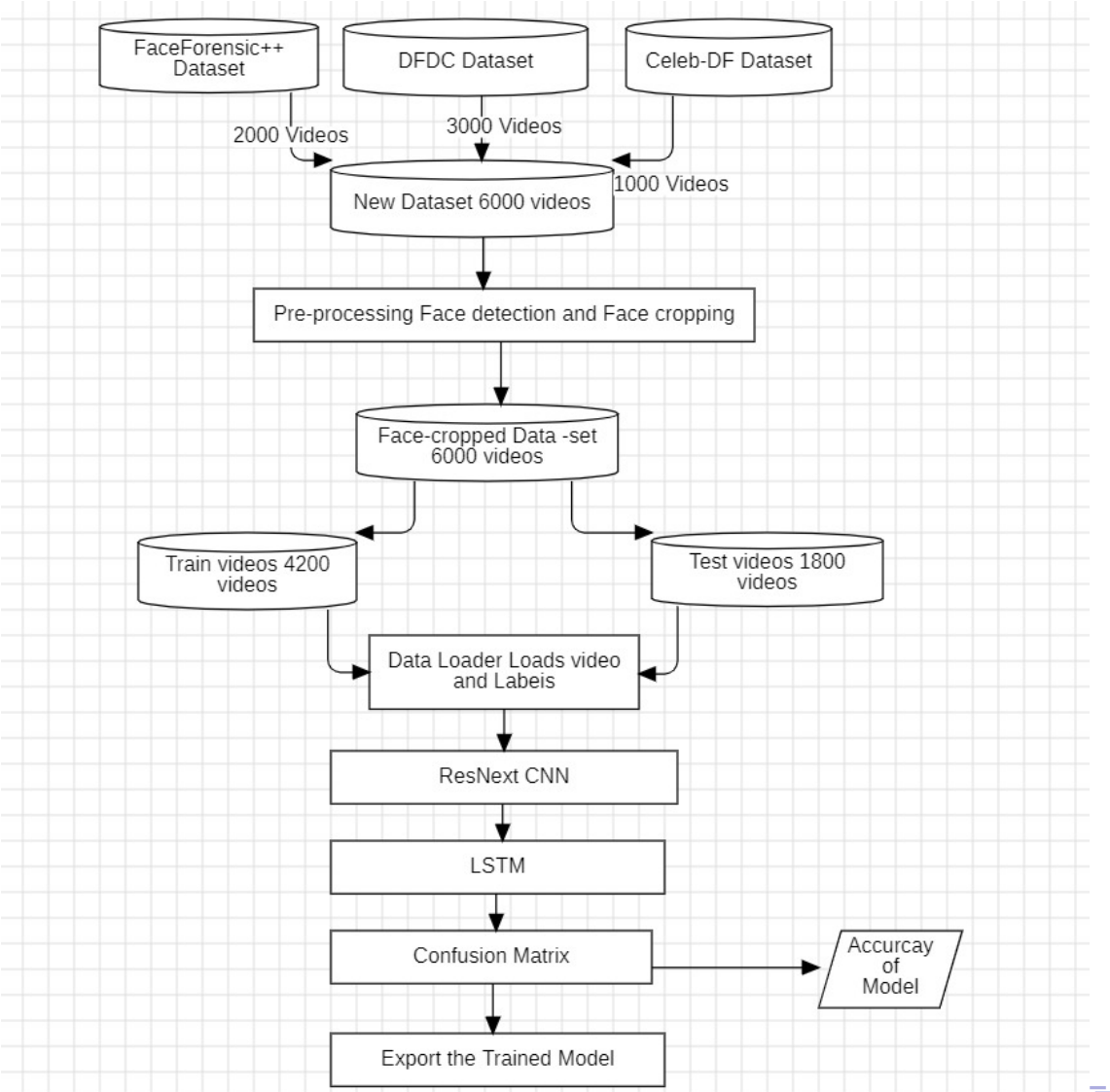
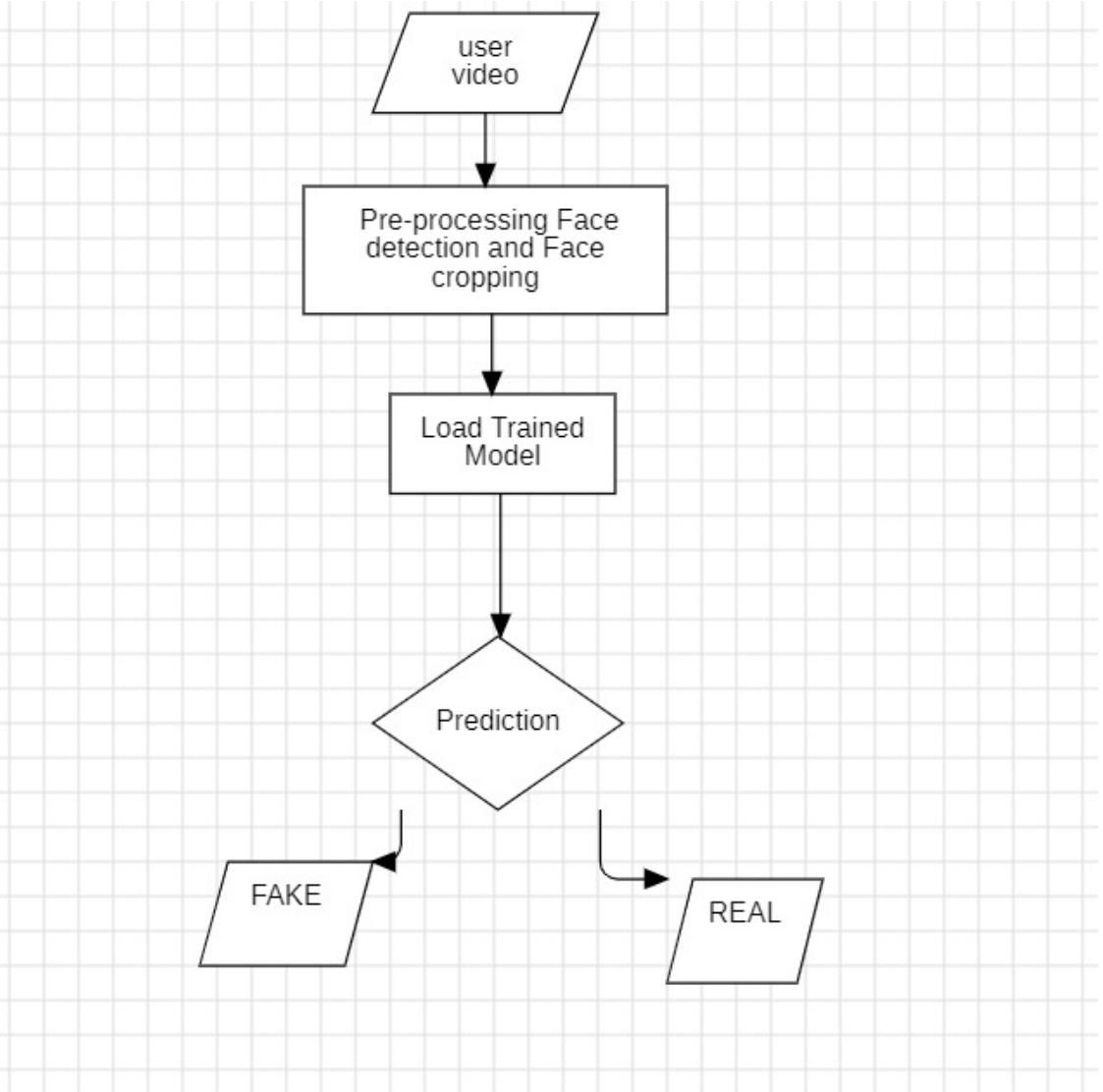


Figure: System Architecture

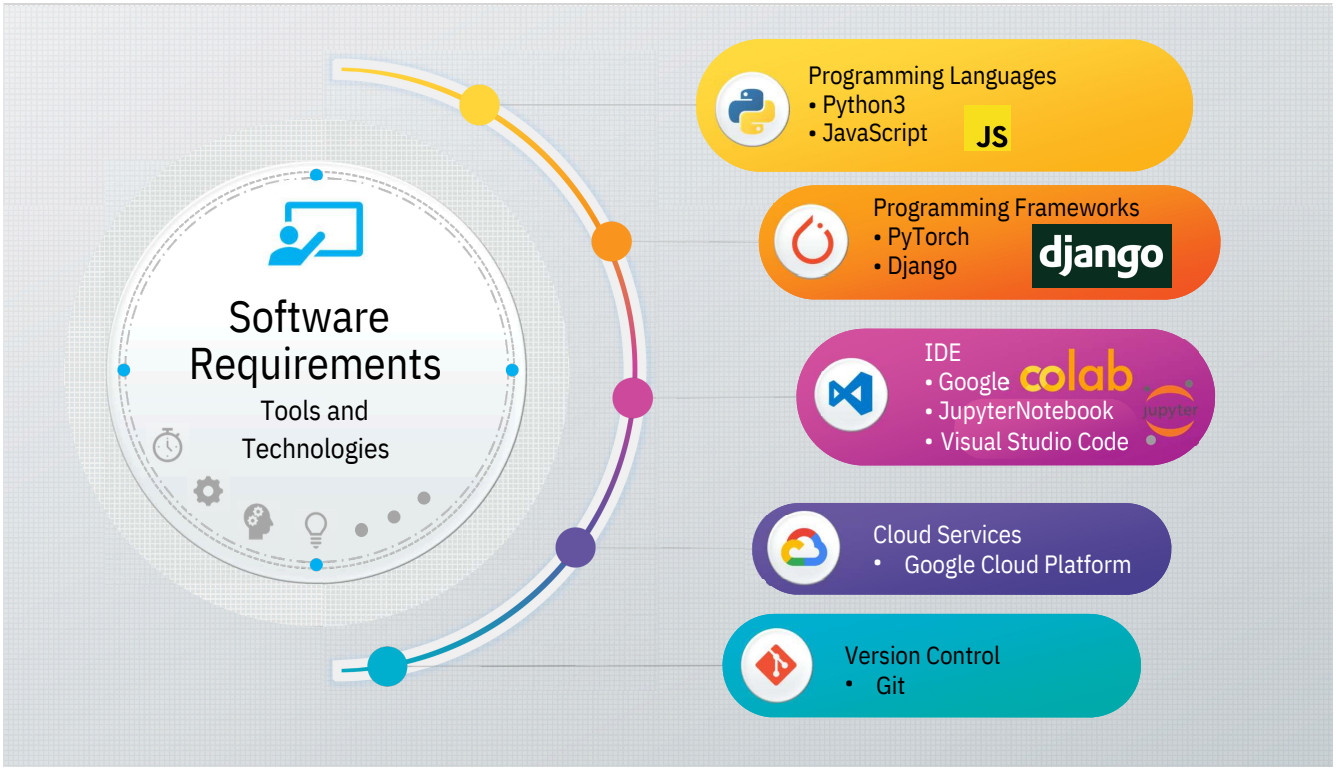
Training Workflow



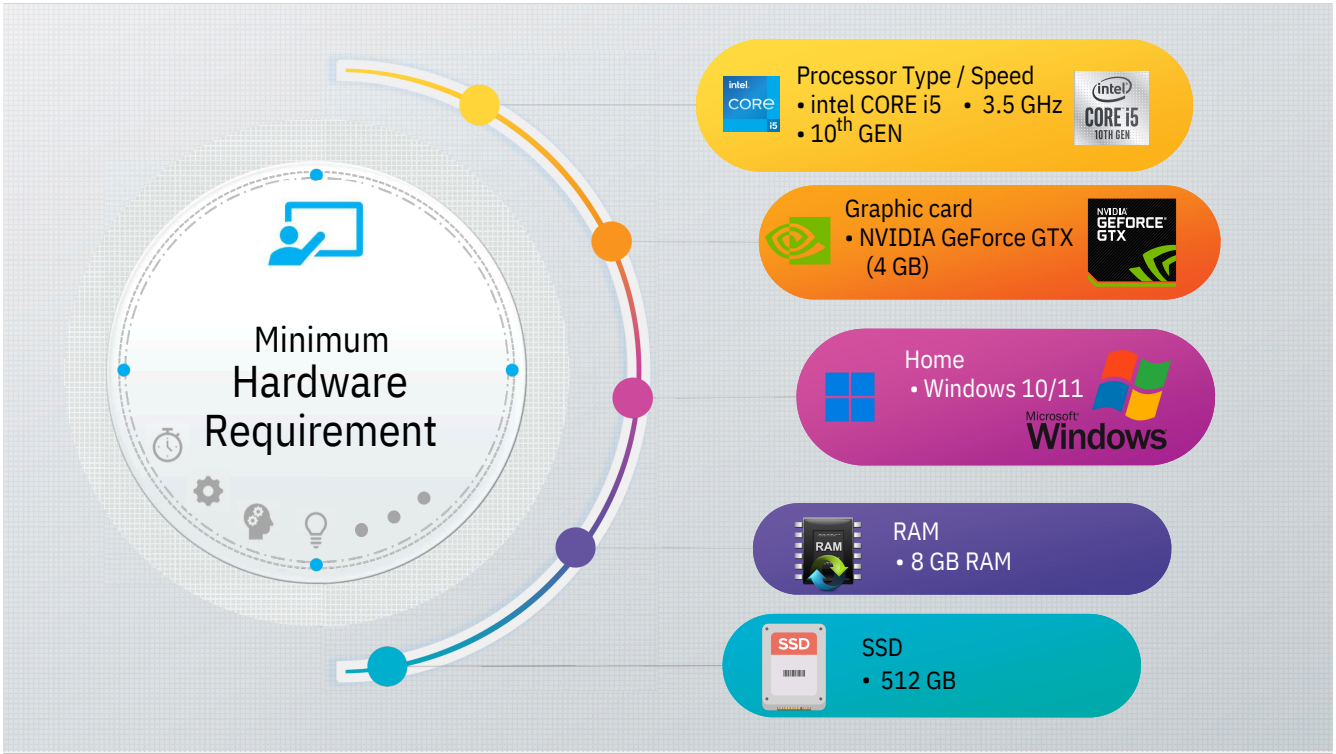
Testing Workflow



Software Requirements



Hardware Requirements



Literature Survey

Works	DataSet	Model Features	Remarks
Sabir et al. [46]	FaceForensics++	Use spatio-temporal features of video streams. Bidirectional Recurrent Neural Network (RNN) + DenseNet/ResNet50.	Not applicable to long video clips. Not trained on a large dataset.
Güera and Delp [27]	Hollywood-2 Human Actions (HOHA)	Temporal inconsistencies of deepfake video is taken into account. Inception-V3 + LSTM.	Didn't take into account of compressed videos.
Li et al. [37]	Closed Eyes in the Wild (CEW)	Used Long Term Recurrent Convolutional Networks. Measured the eye blinking rate. VGG16 + LSTM + FC	Applied to uncompressed videos.
Afchur et al. [6]	Downloaded from internet and processed.	Mesonet structures - Meso-4 and MesoInception-4 used. 2 inception modules + 2 classic convolution layers + 2 FC layers.	Accuracy is lower for highly compressed video.
Li et al. [38]	UADFV and DeepfakeTIMIT	Face warping artifacts. Used 4 CNN models. Measured resolution inconsistency between the warped face area and face.	Compression has not been considered.
Matern et al. [39]	A combination of various sources.	Facial texture difference, and missing details in eye and teeth. Logistic regression model and neural network.	Not for compressed video.
Nguyen et al. [42]	Four major datasets.	VGG-19 + Capsule Network.	Accuracy is low for highly compressed data.
Hashmi et al. [29]	DFDC whole dataset	CNN+LSTM Used facial landmarks and convolutional features	Computation complexity is high. Minimum video length is 10 seconds. Works well for long videos.
Kumar et al. [35]	FaceForensics++ Celeb-DF	Triplet Architecture. Metric learning approach.	For highly compressed video.
Previous work by the authors [40]	FaceForensics++	Face Artifacts Analysis XceptionNet + Classifier Network	For compressed video. High Accuracy.

Conclusion

- In conclusion, the identification of deepfake videos through AI and machine learning solutions represents a critical advancement in safeguarding the integrity of digital content.
- These technologies leverage sophisticated algorithms to detect inconsistencies and anomalies that are often imperceptible to the human eye.
- As deepfake technology continues to evolve, ongoing research and development in detection methods will be essential to stay ahead of potential threats.

Result

Table 10.1: Trained Model Results

Model Name	Dataset	No. of videos	Sequence length	Accuracy
model_90_acc _20_frames_ FF_data	FaceForensic++	2000	20	90.95477
model_95_acc _40_frames_ FF_data	FaceForensic++	2000	40	95.22613
model_97_acc _60_frames_ FF_data	FaceForensic++	2000	60	97.48743
model_97_acc _80_frames_ FF_data	FaceForensic++	2000	80	97.73366
model_97_acc _100_frames_ FF_data	FaceForensic++	2000	100	97.76180
model_93_acc _100_frames_ celeb_FF_data	Celeb-DF + FaceForensic++	3000	100	93.97781
model_87_acc _20_frames_ final_data	Our Dataset	6000	20	87.79160
model_84_acc _10_frames_ final_data	Our Dataset	6000	10	84.21461
model_89_acc _40_frames_ final_data	Our Dataset	6000	40	89.34681

References

- Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies,Matthias Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images" in arXiv:1901.08971.
- Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., Ortega-Garcia, J. (2020). Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection. Information Fusion, 64, 131-148.
- Verdoliva, L. (2020). Media Forensics and Deepfakes: An Overview. IEEE Journal of Selected Topics in Signal Processing, 14(5), 910-932.
- https://link.springer.com/article/10.1007/s10462-024-10810-6