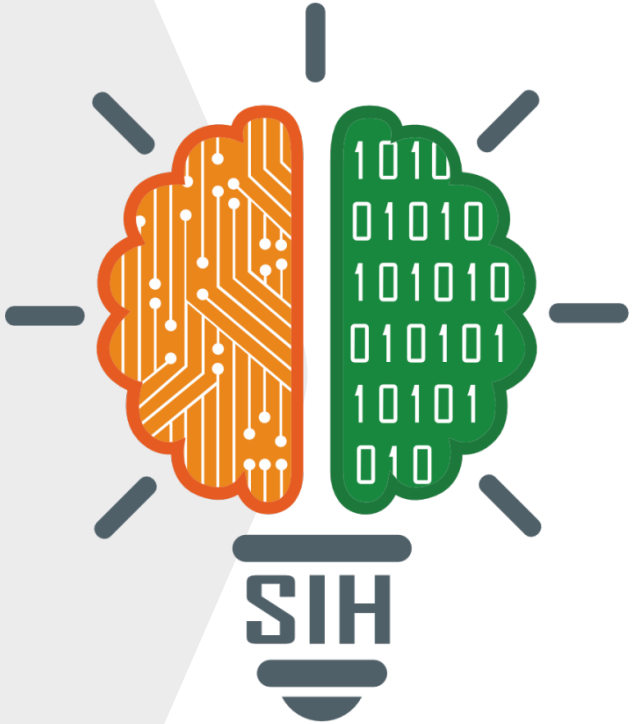
SMART INDIA HACKATHON 202

SMART INDIA HACKATHON 2024

DEEP FAKE VIDEO DETECTION

- **Problem Statement ID – 1683**
- **Problem Statement Title- Detection of face-swap based deep fake videos**
- **Theme- Miscellaneous**
- **PS Category- Software**
- **Team ID- 5358**
- **Team Name- Mafioso**



MAFIOSO

DETECTION OF FACE-SWAP BASED DEEP FAKE VIDEOS

SMART INDIA HACKATHON 2024

Now with the development of generative models making a deepfake videos with fabricated faces becomes a common thing. so to find the fabricated faces we clustered many deep learning model together to create a hybrid model which can identify the fake face from the deepfake videos

Model Architecture: DEEPFACE Generative Adversarial Networks (GANs) + DEEPVOICE Generative Adversarial Networks (GANs) + Long Short-Term Memory (LSTM).

STEP 1: Video will be extracted into frames and audio. Then the frames will be passed to the face extraction unit and the audio will be passed to audio enhancement unit.

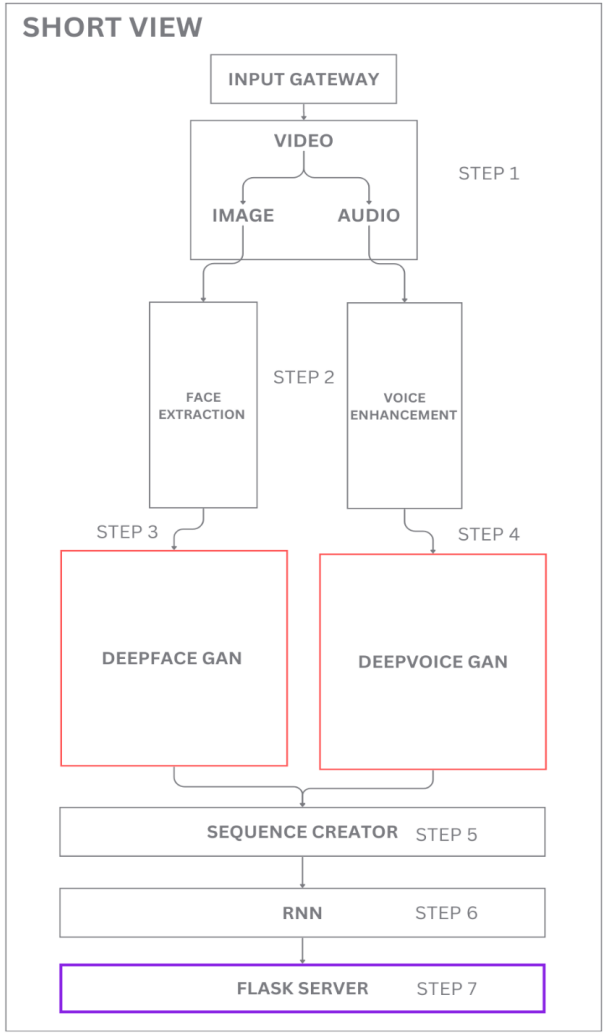
STEP 2: Inside the face extract unit, faces will be recognized and cropped from the frames using the Haar cascade algorithm. Then, the faces will be enhanced, and filters will be applied. In the case of audio, the noises will be filtered to enhance the quality of voice.

STEP 3: Faces that are enhanced will be passed to the discriminator in the deep-face GAN. Then the discriminator will classify whether the faces are fabricated or not. The use of GAN allows the parameters of the discriminators to be trained on new variants of data. So that it is easy to classify new faces.

STEP 4: Audio will be passed to the discriminator in the deep-voice GAN, which then classify whether any fabrications are made in audio to match the video frames.

STEP 5: Then the result from the deep-face GAN and deep-voice GAN will be arranged in the sequence of the frame.
STEP 6: Then the result sequence will be transferred to LSTM(Long Short-Term Memory). Here, LSTM is used to find the hidden patterns in the frame sequence to solve mitigating issues with distorted frames due to face movement or low frame rates.

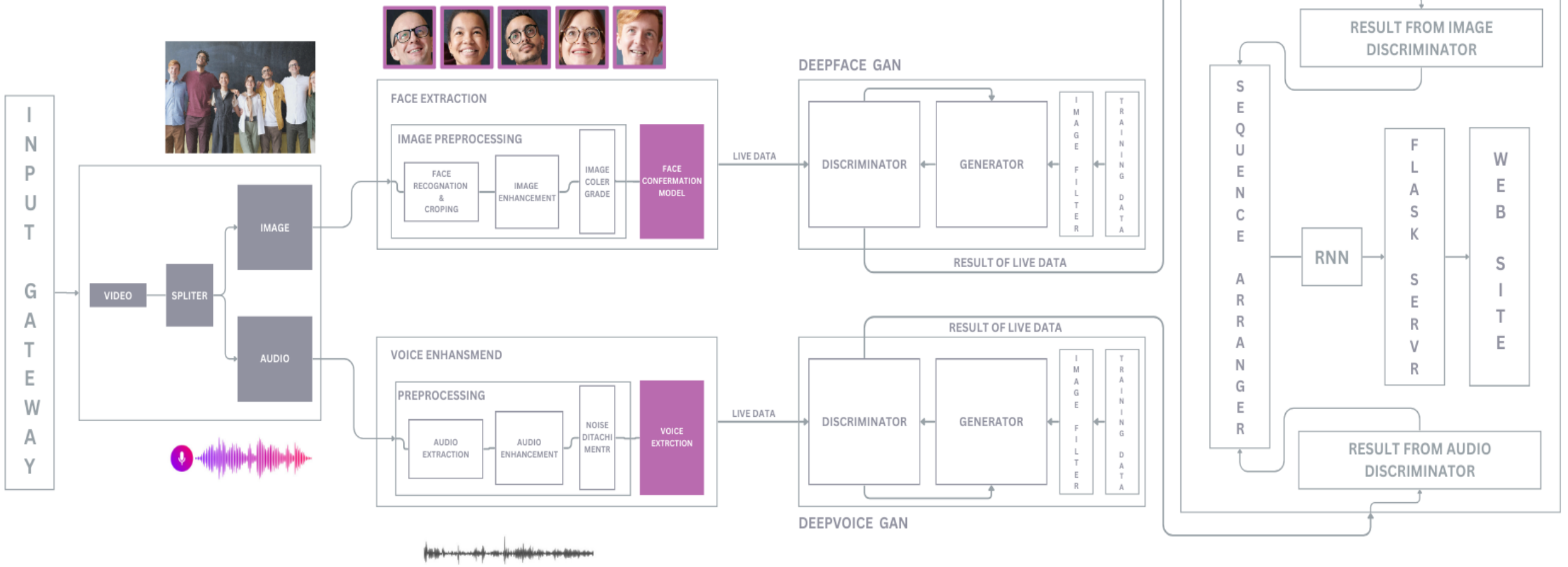
STEP 7: Then in the end the detected fabricated faces with their corresponding LSTM results will be displayed in an analytical dashboard



MAFIOSO

TECHNICAL APPROACH

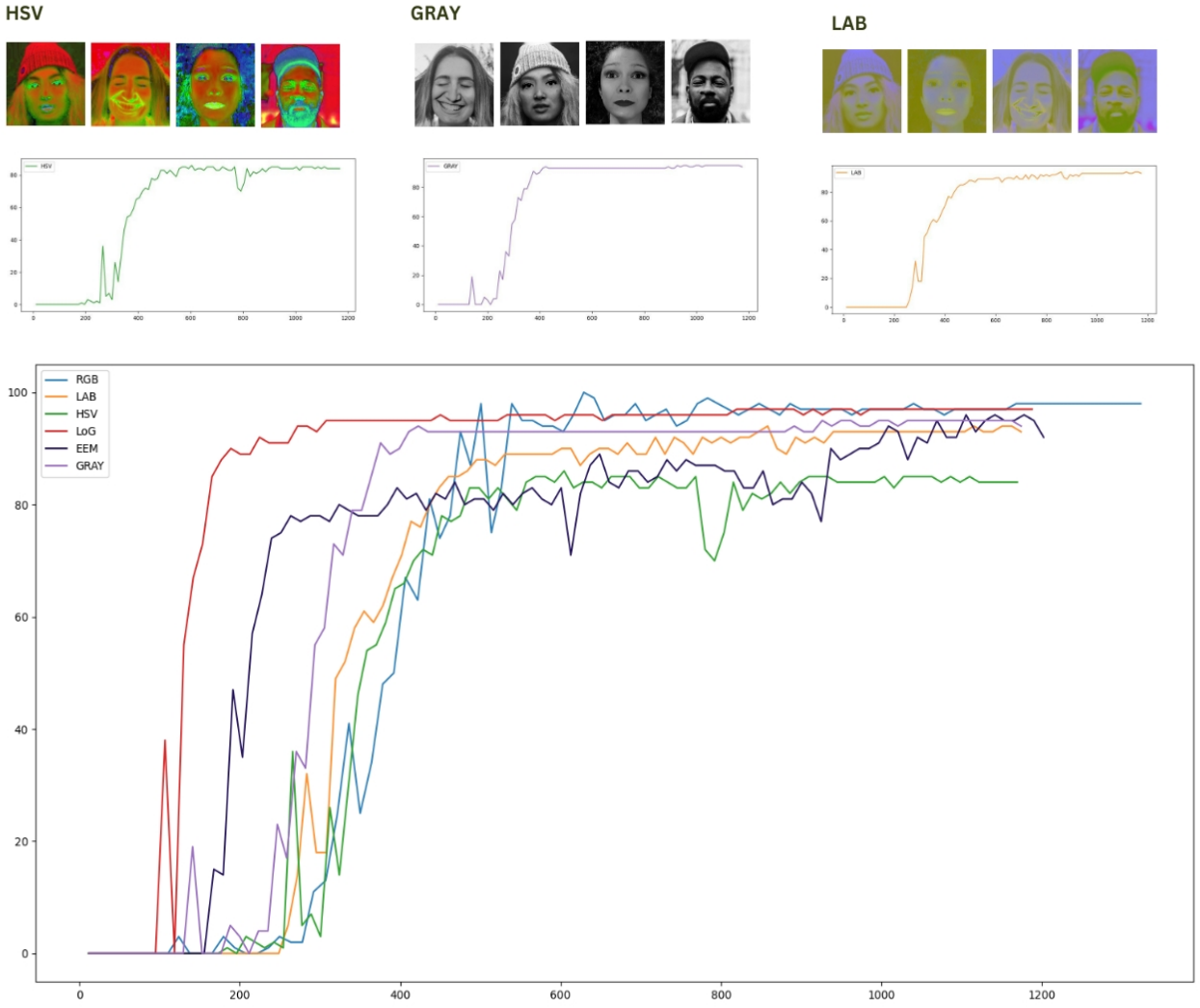
SMART INDIA HACKATHON 2024



MAFIOSO

FEASIBILITY AND VIABILITY

SMART INDIA HACKATHON 2024



We have tested many different Filters to find the best-suited filter to get a high accuracy rate and less training time. So, we take 1000 images in both real and fake as training data and another 100 images as test data then we apply all the filters to both the test and train data. Then we train all the different gradient images in a discriminator model with the same weight and in the same number of epochs to find which filter gives us the ideal output.

The tested filters used are:

*RGB *HSV *GRAYSCALE *LAB *LoG

After the testing, we found out that the RGB and the LoG formats give the highest accuracy compared to all the other image formats.

When we take a closer look at the data of both RGB and LoG which is collected in training we can see that the RGB format is a little more accurate than the LoG format but the LoG format is a little faster than the RGB format while training. But the difference is very minute so we can choose whatever format suites our requirement more effectively. For our training, we chose the LoG format for the training data due to the reason that is faster, takes less computing power, and easy to identify hidden patterns due to the outline mapping of images.

MAFIOSO

IMPACT AND BENEFITS

SMART INDIA HACKATHON 2024

*Unlike regular CNN training we use the generator to train the discriminator in a GAN. so that the parameters of the discriminator are trained in a different variate of image every single time rather than training in a single set of pictures like in a regular CNN

*In this method we use both face and voice to determine deepfake rather than only relaying or depending on a single entity.

*the result of both the deep-face and deep-voice models are directly sent into an RNN (LSTM) model in the order of the frame sequence. So that we can able to find the hidden patterns in the frame sequence to solve a mitigating issues with distorted frames due to face movement or low frame rates

*Due to the use of the GAN in the training of discriminators we don't need to collect fake images from the fabricated videos for the training which is a highly time-consuming job.

*Using this type of model to test fabricated videos will be less costly than the traditional ways

*We can also add several other models into the existing model in the feature for better predictability of the model and it will not affect the existing model.

*Due to the face extraction and labeling in the image proceeding we can accurately identify which faces are fabricated and which faces are not by the result from the deep-face models.

RESEARCH AND REFERENCES

https://arxiv.org/pdf/1909.12962v4
https://www.kaggle.com/datasets/dagnelies/deepfake-faces?select=faces_224
https://www.kaggle.com/datasets/greatgamedota/faceforensics
https://arxiv.org/pdf/1901.08971v3
https://github.com/Daisy-Zhang/Awesome-Deepfakes-Detection
https://arxiv.org/pdf/1809.00888v1
https://arxiv.org/pdf/2406.04932v1
https://www.kaggle.com/datasets/birdy654/deep-voice-deepfake-voice-recognition

RESEARCH:

*Development of new hybrid model using different deep learning models

*Finding of a best-suited filter to get a high accuracy rate and less training time in deep face GAN