A Deep Learning-Based Approach for Inappropriate Content Detection and Classification of YouTube Videos

Now-a-days YouTube content are access by all age group of peoples as this provide digital entertainment content on various topics such as Sports, Religious, Movies and many more. This channel provide vast amount cartoon entertainment for KIDS. Some malicious users are taking advantage of this digital content to spread inappropriate content for kids the in the form of cartoons. Such inappropriate content may put bad influence on growing kids and need a technique to prevent such content before showing to kids.

To prevent such content many machine learning and deep learning algorithms are introduced but their inappropriate content detection and classification accuracy is not up to the mark. To overcome from this issue author of this paper employing combination of CNN and BI-LSTM algorithm to detect inappropriate content.

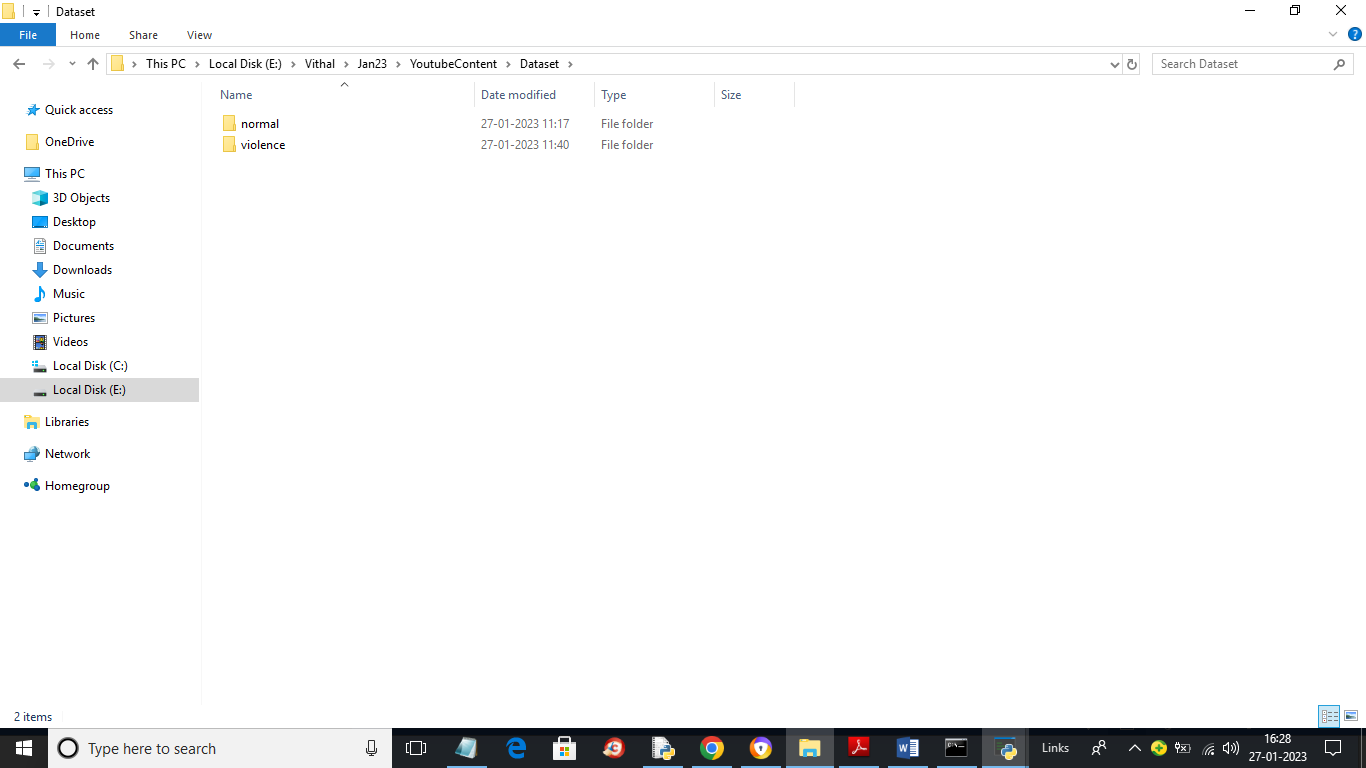
Pre-trained EfficientNetB7 CNN algorithm is employed to extract features from the YouTube video images and then retrained with BI-LSTM algorithm to enhance prediction accuracy.

In propose work a novel deep learning-based architecture is proposed for the detection and classification of inappropriate content in videos. For this, the proposed framework employs an ImageNet pre-trained convolutional neural network (CNN) model known as EfficientNet-B7 to extract video descriptors, which are then fed to bidirectional long short-term memory (BiLSTM) network to learn effective video representations and perform multiclass video classification. An attention mechanism is also integrated after BiLSTM to apply attention probability distribution in the network.

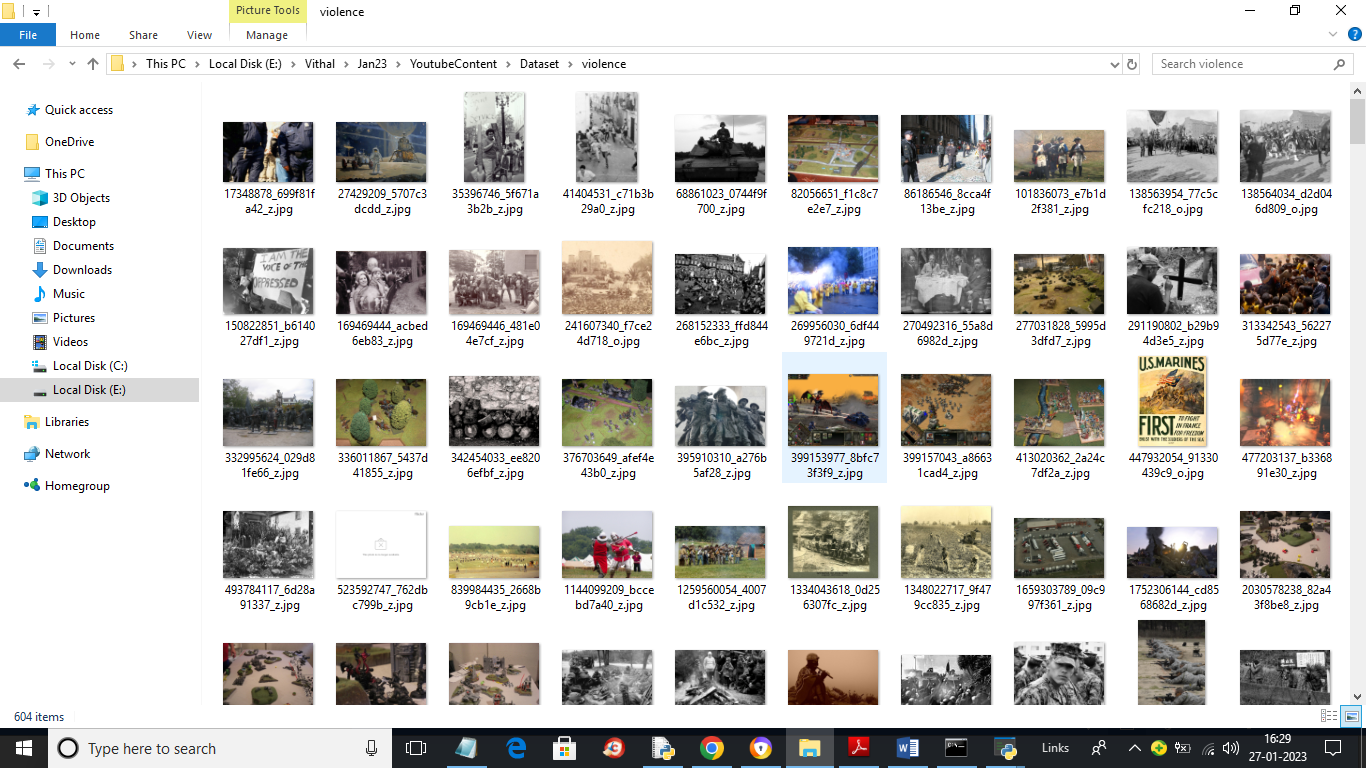
Two algorithms are trained on YouTube video annotated images such as Attention based BI-LSTM and EfficientNetB7 based BI-LSTM and in both algorithms EfficientNetB7 is giving better accuracy. On same dataset we have experiment with existing SVM algorithm but its accuracy is less than propose EfficientNetB7-BILSTM algorithm.

In propose paper author has used vulgar videos like nudity and sex so we cannot used such dataset to test above algorithms so we have used ‘Normal and Violence (fight)’ type of dataset to train above algorithms. For kids violence and fighting videos are also consider as Inappropriate Content.

To train all algorithm we have used below YouTube images consist of two folders called ‘Normal and Violence’ and below screen showing dataset details



In above screen we have two folders and juts go inside any folder to view training images



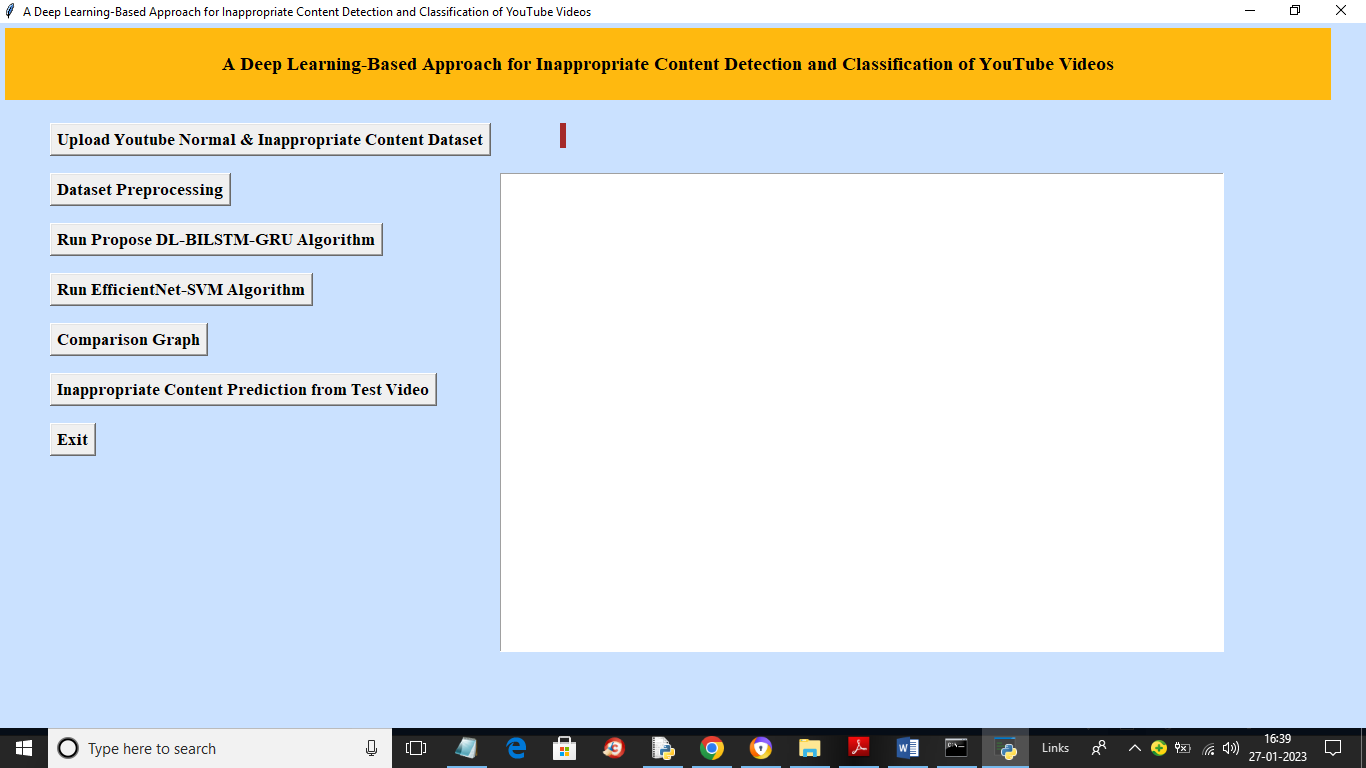
In above screen we can see violence images used to train algorithms

To implement this project we have designed following modules

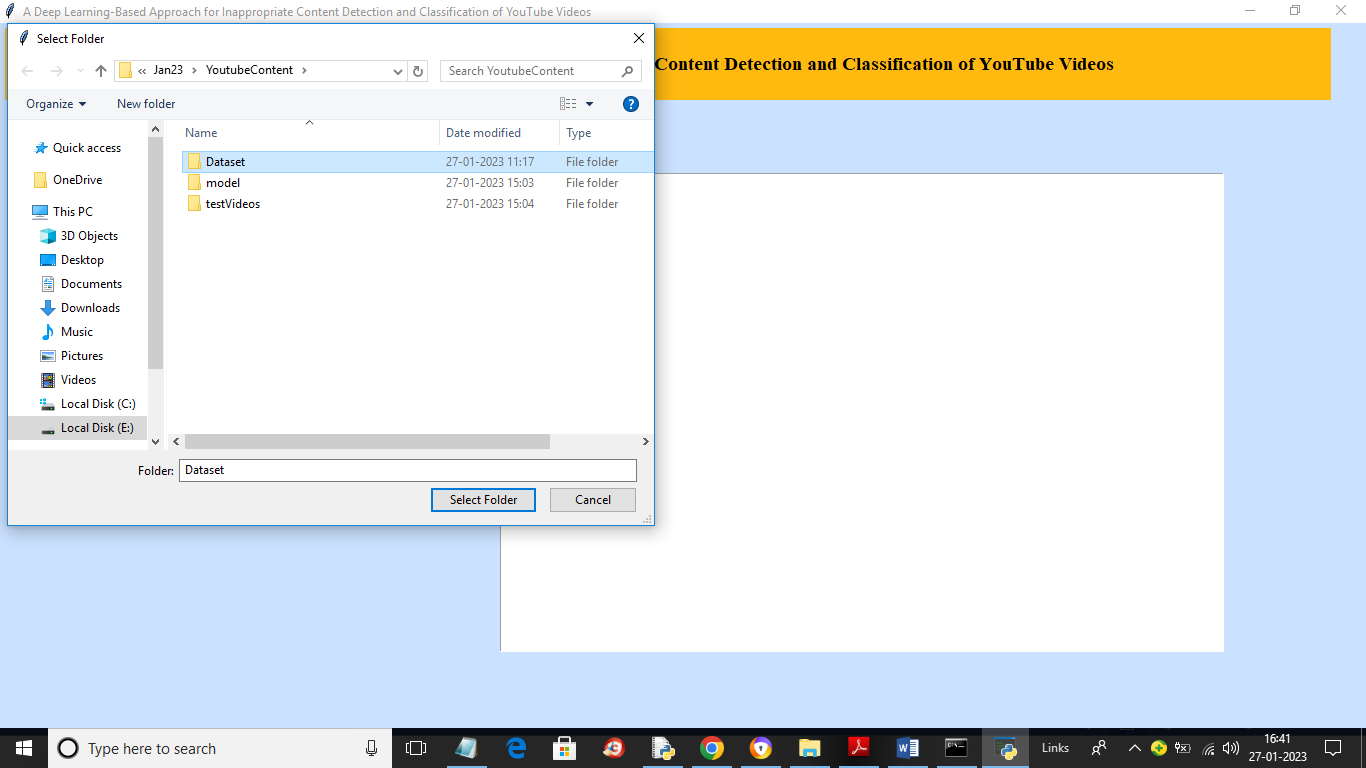
1. Upload YouTube Normal & Inappropriate Content Dataset: using this module we will upload YouTube dataset images to application
2. Dataset Preprocessing: using this module we will read all images and then resize all images to equal size and then normalize image pixel values and then shuffle the dataset
3. Run Propose DL-BILSTM-GRU Algorithm: using this module we will split dataset into train and test and then input 80% training data to Pre-Trained CNN (EfficientNetB7) algorithm to extract digital content from images and then those features will get retrained with BI-LSTM algorithm to train a model. Trained model will be applied on 20% test data to calculate prediction accuracy
4. Run EfficientNet-SVM Algorithm: EfficientNetB7 features will get retrained with existing SVM algorithm and then calculate prediction accuracy
5. Comparison Graph: using this module we will plot accuracy comparison graph between propose EfficientNetB7-BILSTM and EfficientNetB7-SVM.
6. Inappropriate Content Prediction from Test Video: using this module we will upload any YouTube and if video contains fighting or violence then application will predict as ‘Inappropriate Content’ otherwise will predict SAFE content.

SCREEN SHOTS

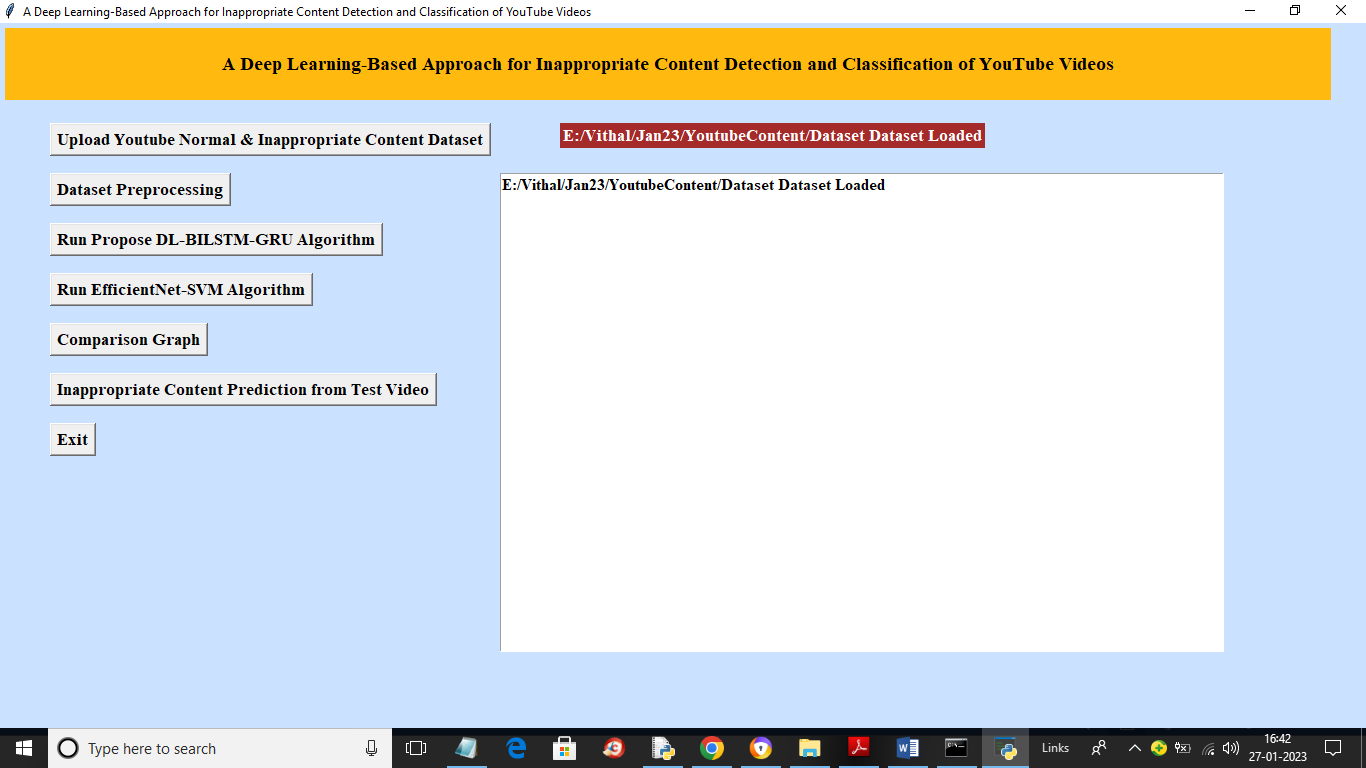
To run project double click on ‘run.bat’ file to get below screen



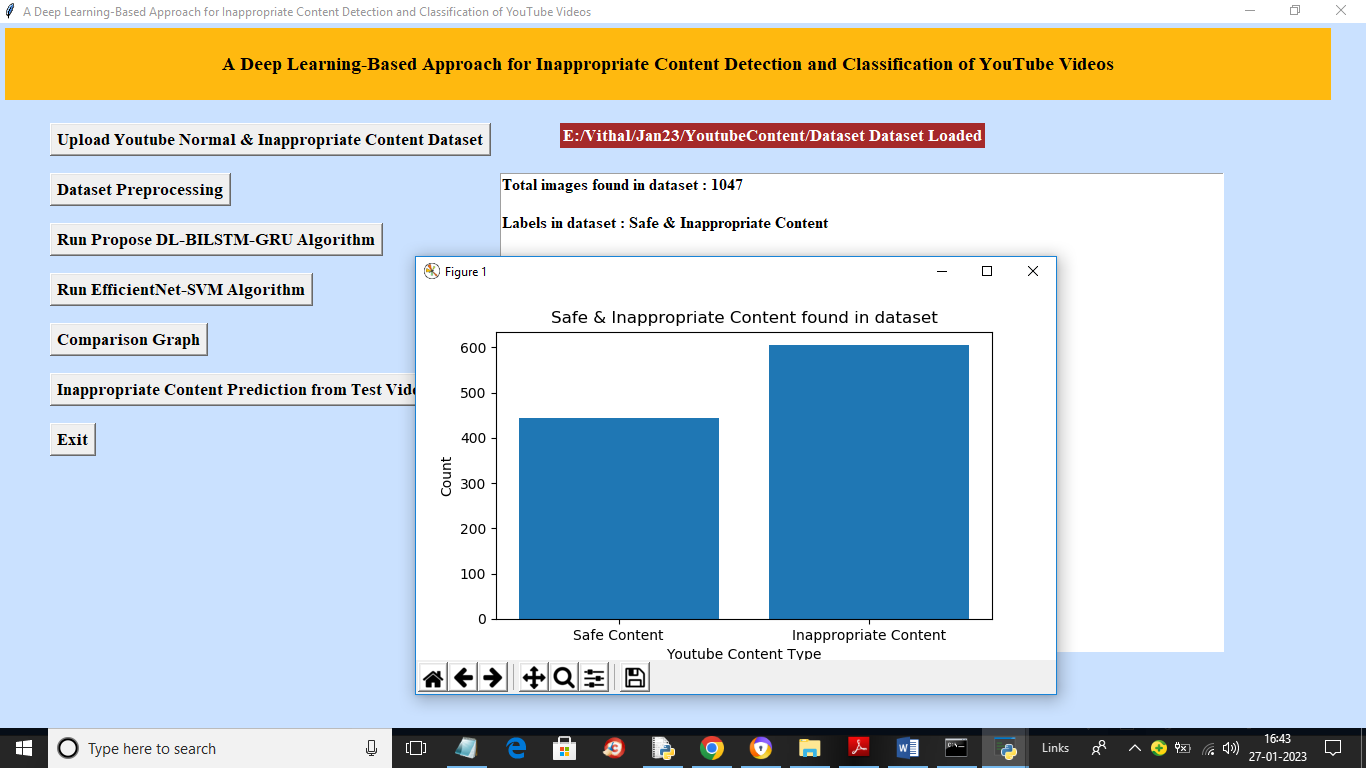
In above screen click on ‘Upload YouTube Normal & Inappropriate Content Dataset’ button to upload dataset and get below output



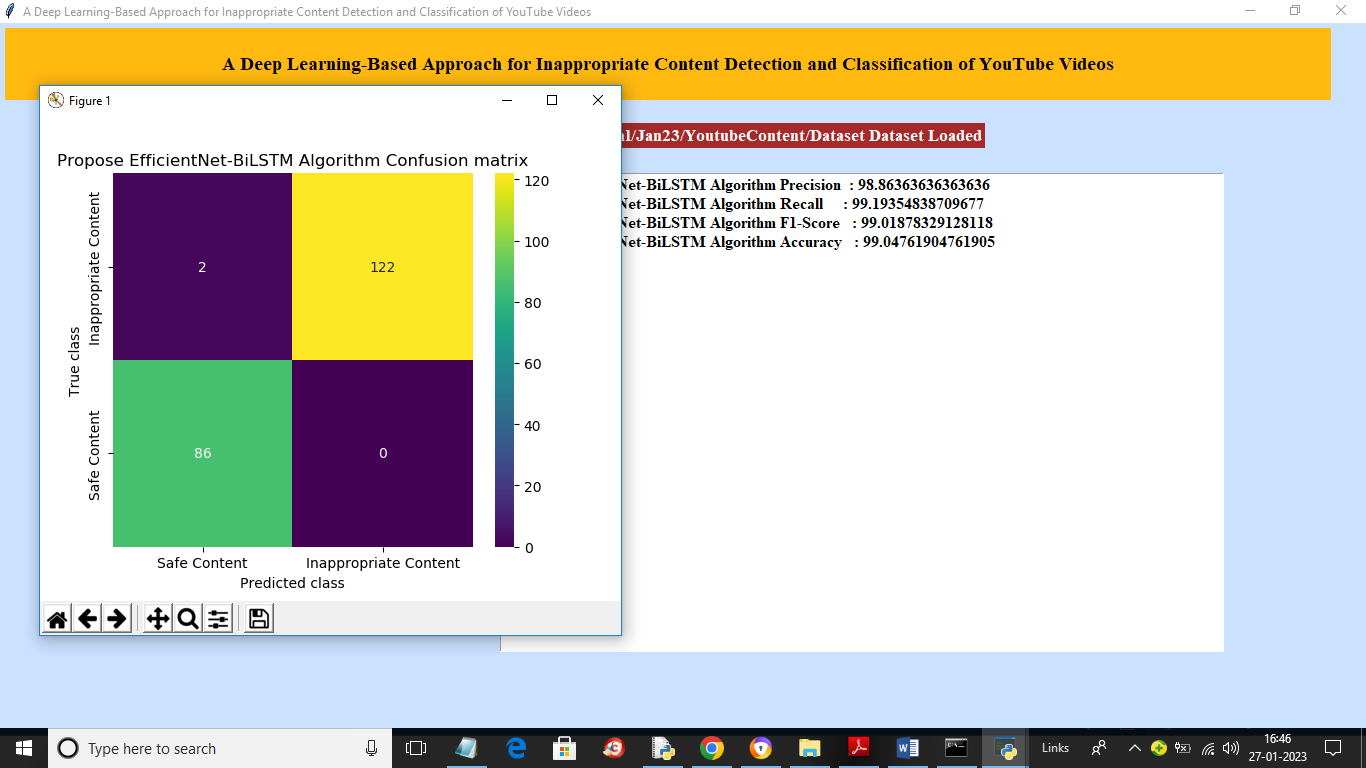
In above screen selecting and uploading entire ‘Dataset’ folder and then click on ‘Select Folder’ button to load dataset and get below output



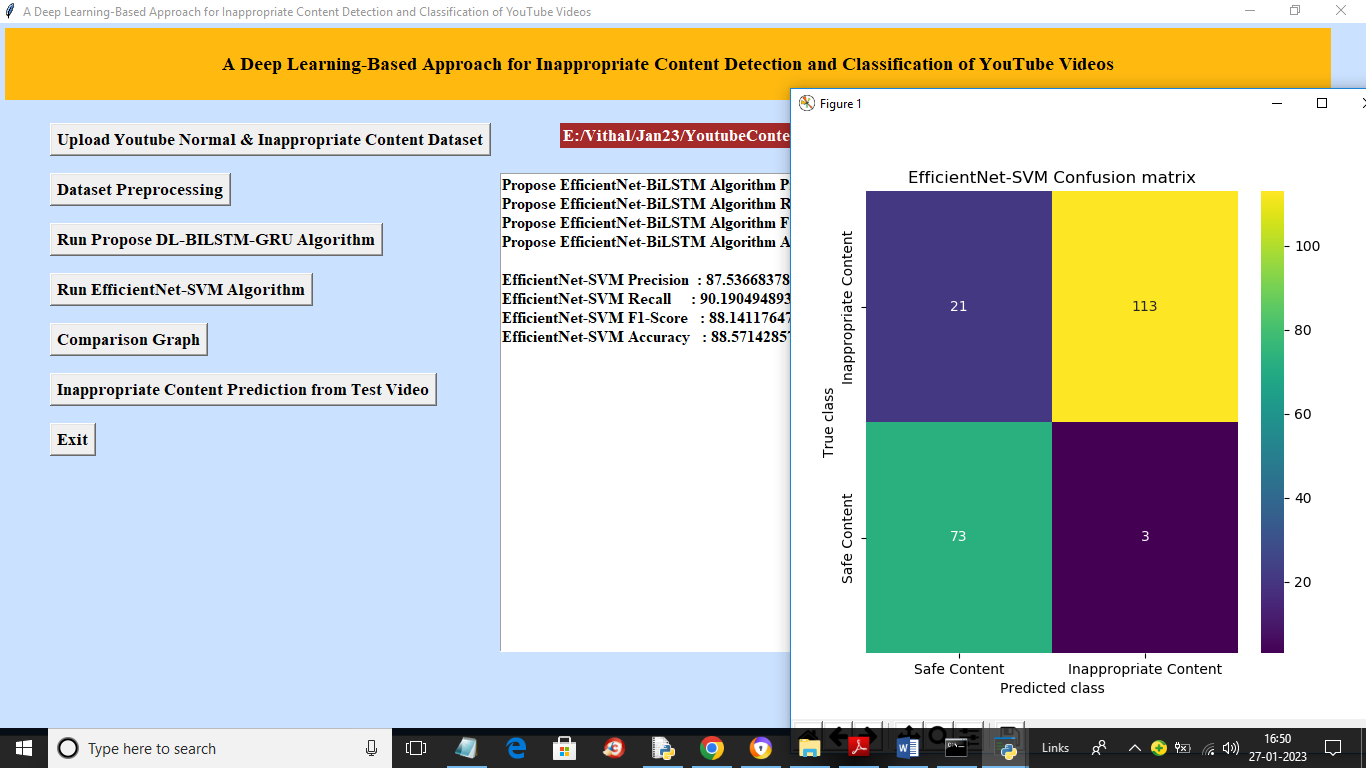
In above screen dataset loaded and now click on ‘Dataset Preprocessing’ button to read all images and then processes those images for training and get below output



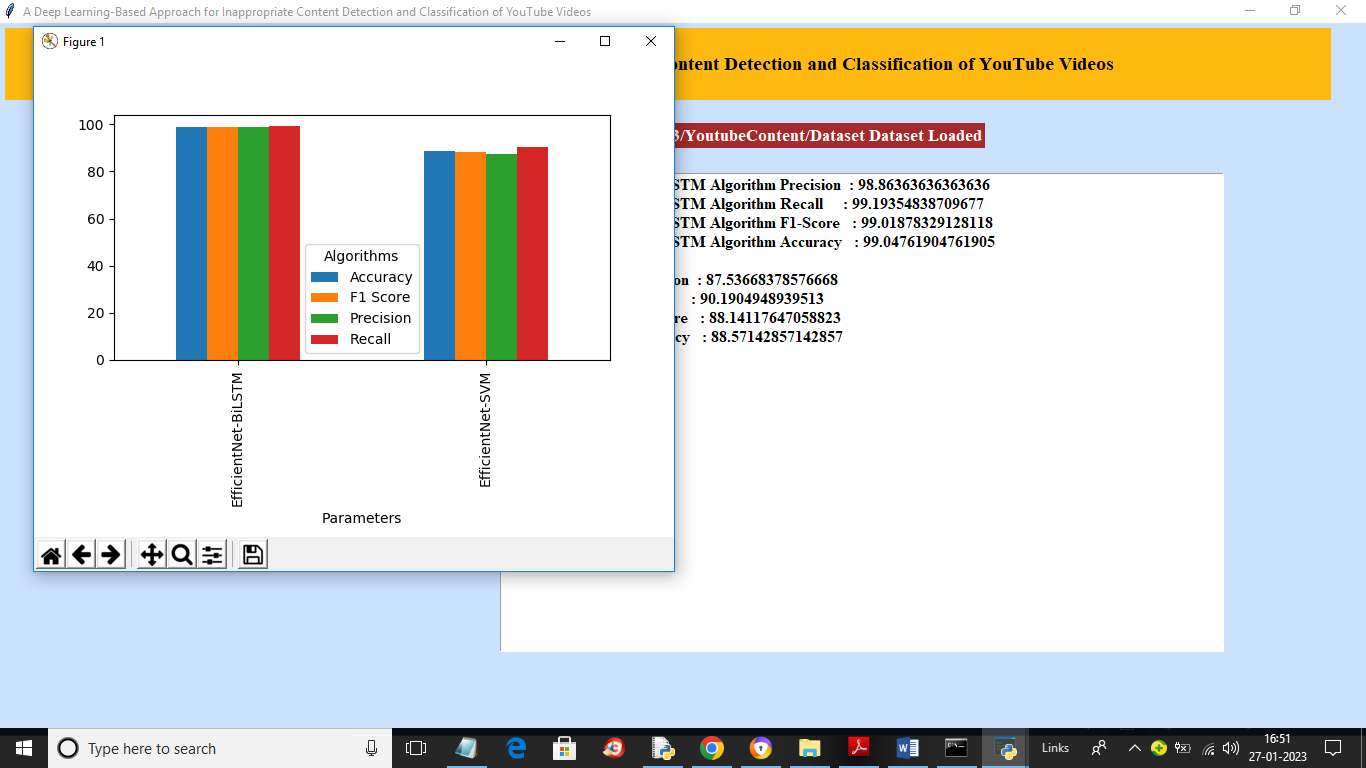
In above screen we can see dataset contains 1047 images and then in graph x-axis represents type of images such as ‘Safe and Inappropriate’ and y-axis represents count of those images. Now dataset processing completed and now click on ‘Run Propose DL-BILSTM-GRU Algorithm’ button to train propose algorithm and get below output



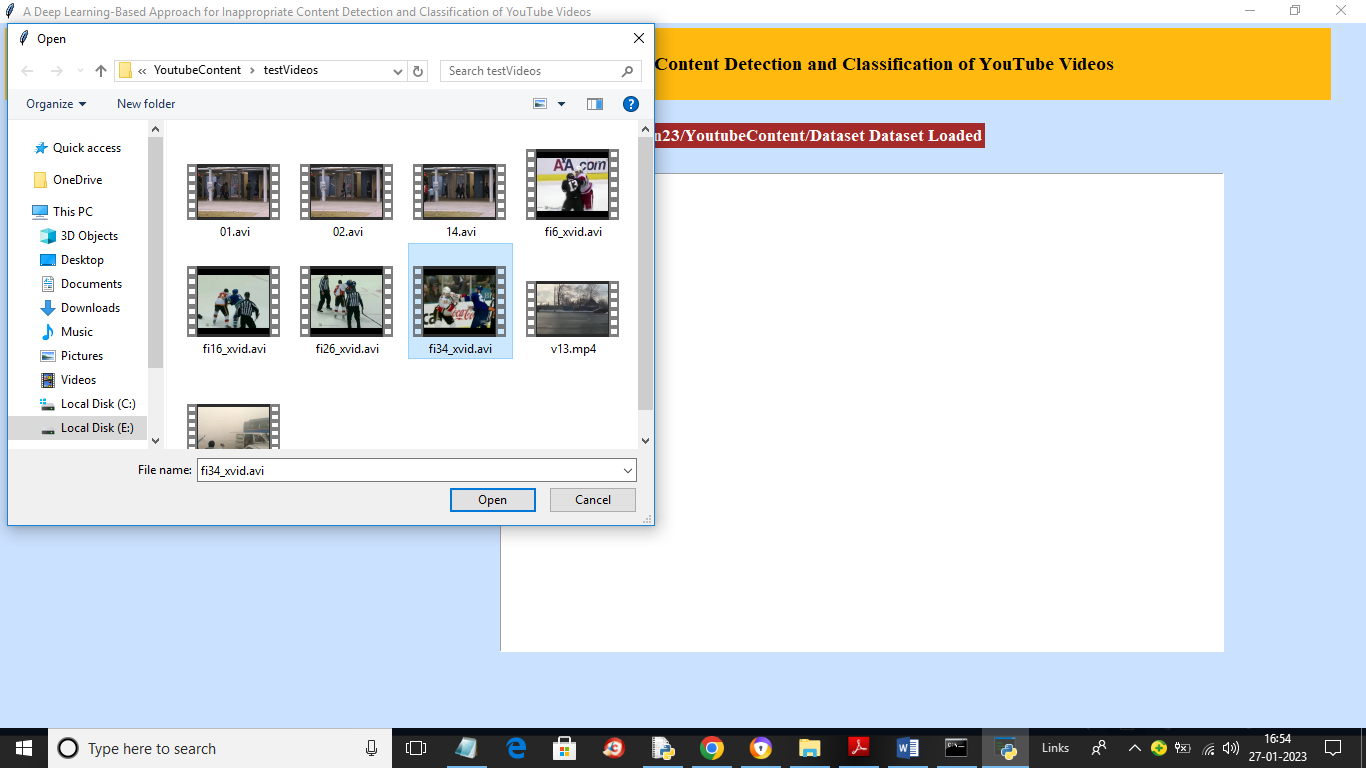
In above screen with propose EfficientNetB7-BI-LSTM we got 99.04% accuracy and in confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels and green and yellow boxes contains correct prediction count and blue boxes contains incorrect prediction count which is 2 only. Now close above graph and then click on ‘Run EfficientNet-SVM Algorithm’ button to get below output



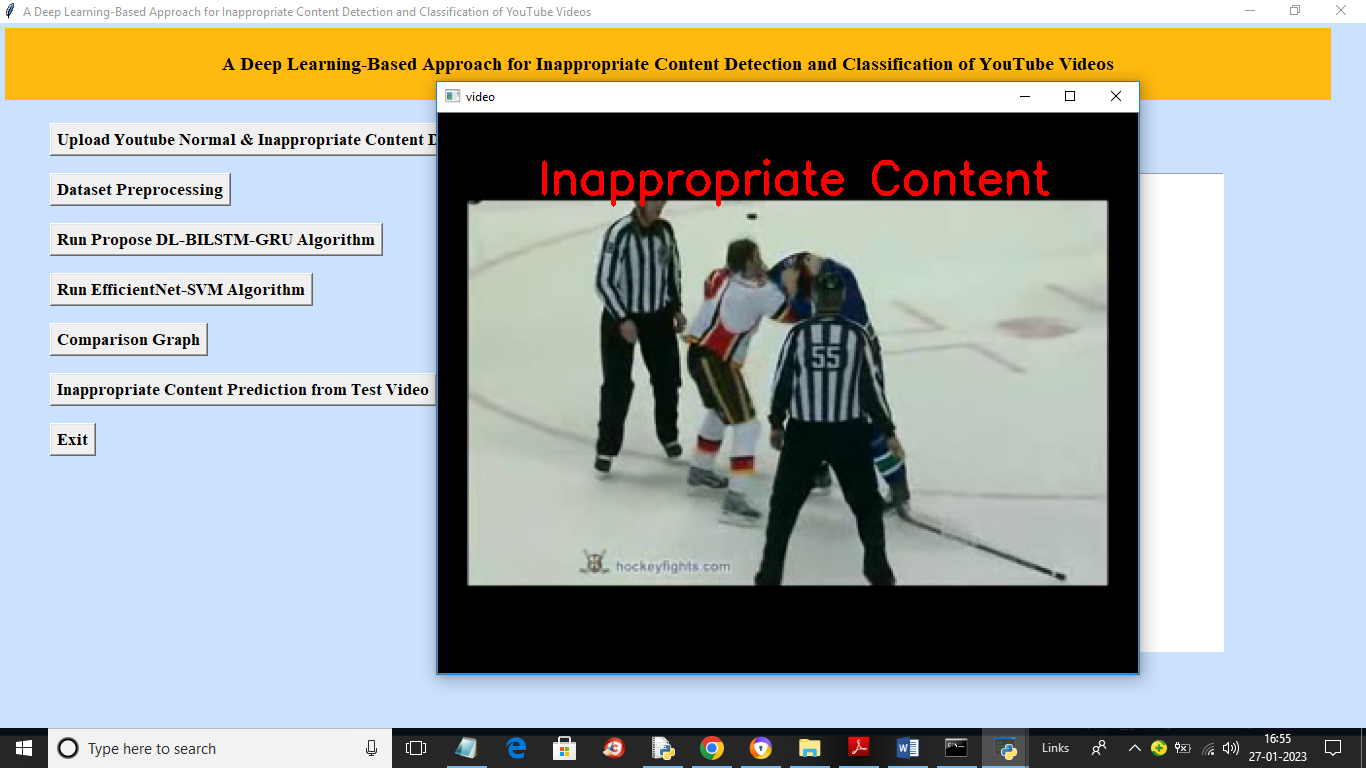
In above screen with EfficientNetB7-SVM we got 88% accuracy and in confusion matrix graph we can see in blue boxes that SVM predicted total 24 incorrect prediction so its accuracy is less. Now close above graph and then click on ‘Comparison Graph’ button to get below output



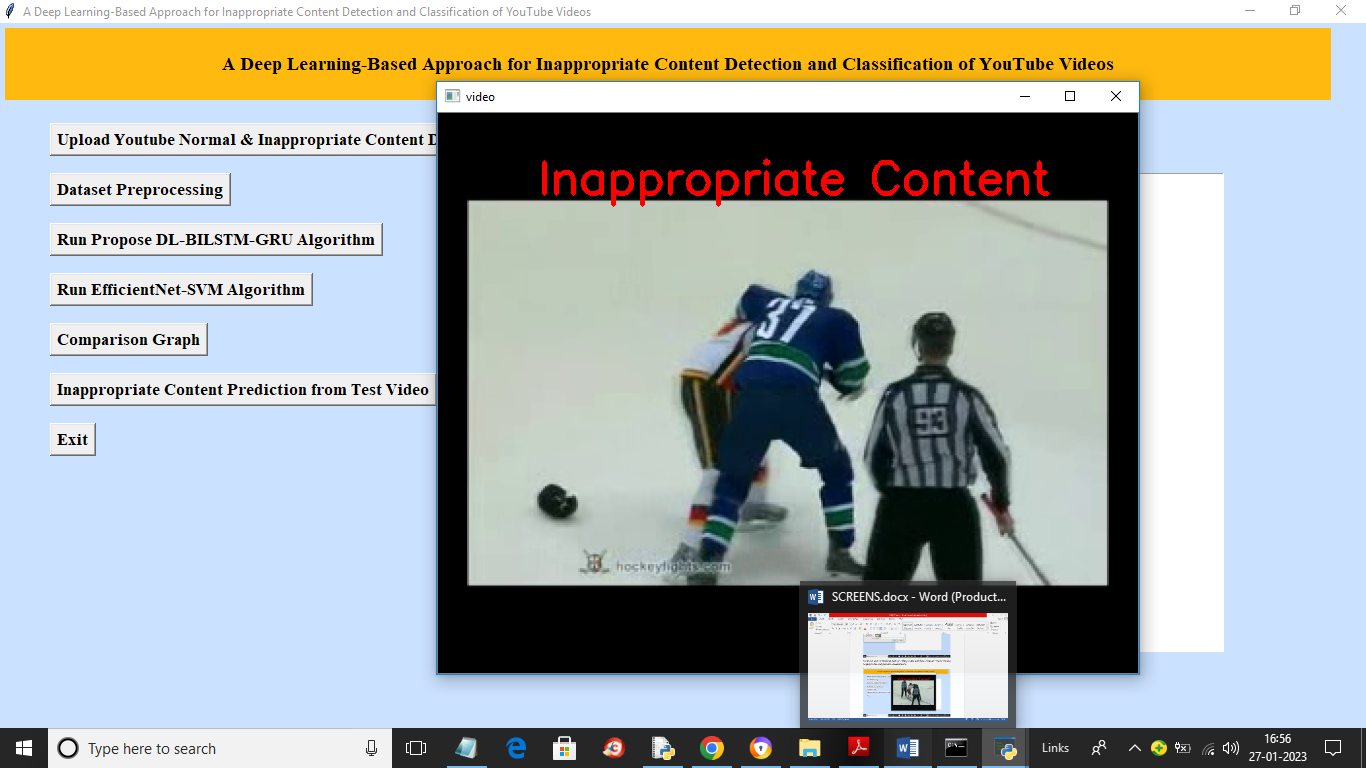
In above graph x-axis represents algorithm names and y-axis represents accuracy and other metrics in different colour bars. In both algorithms propose EfficientNetB7-BI-LSTM got high accuracy. Now close above graph and then click on ‘Inappropriate Content Prediction from Test Video’ button to upload test video and classify it as Safe or inappropriate.



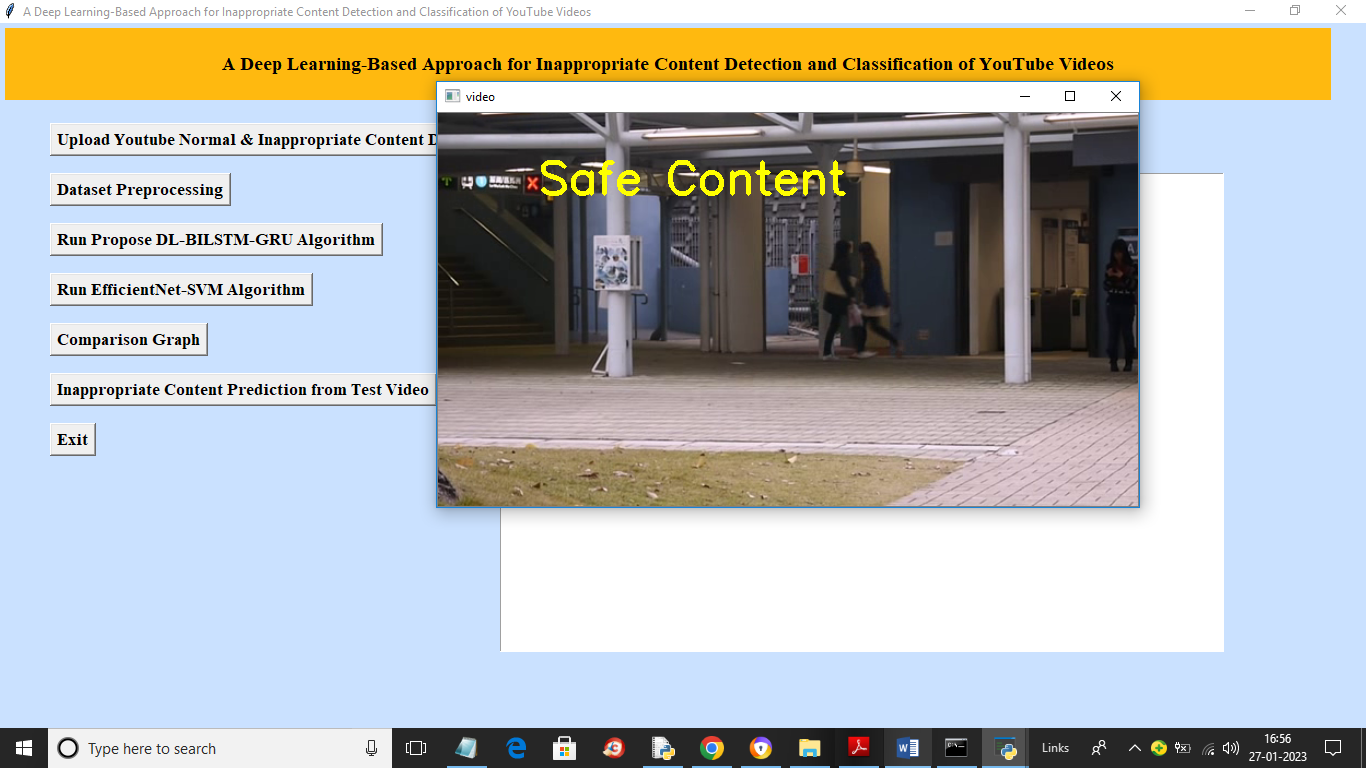
In above screen selecting and uploading video and then click on “Open’ button to play video and perform classification



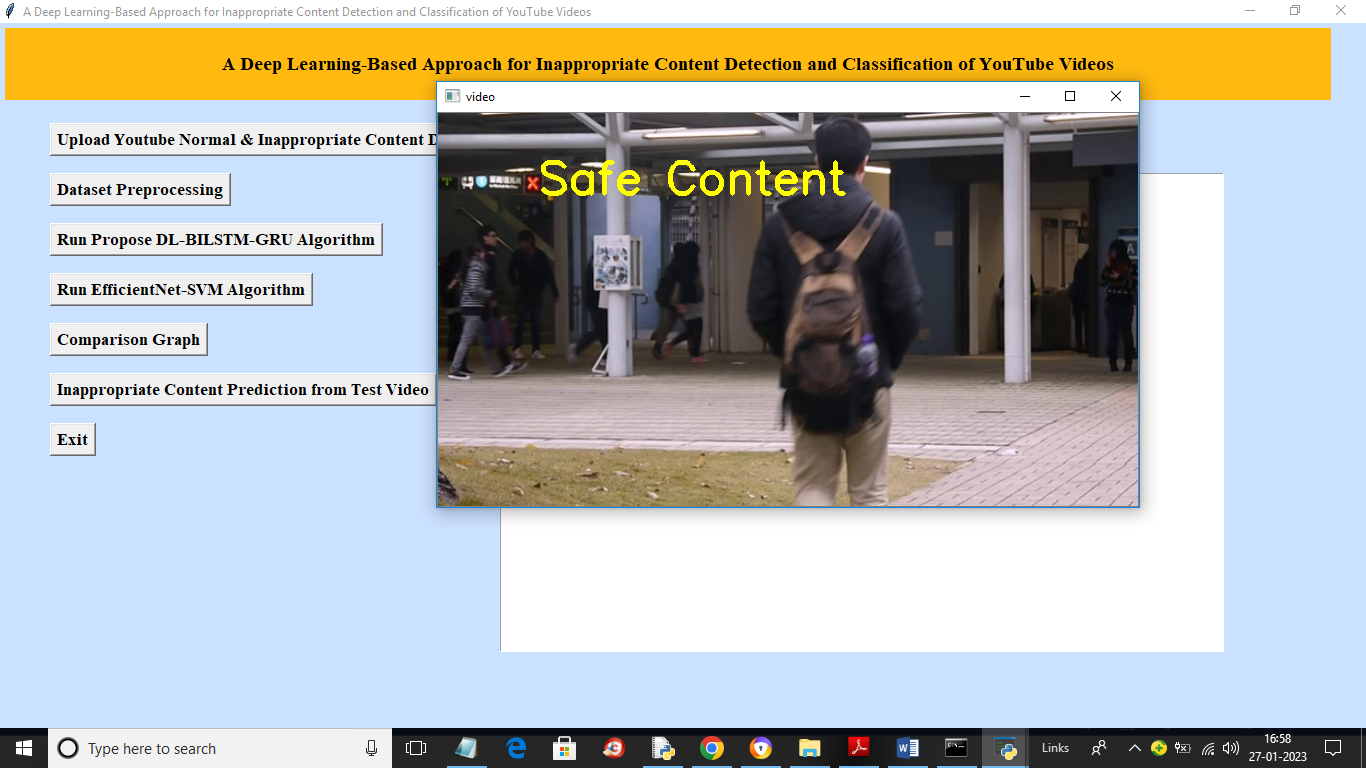
In above screen propose algorithm evaluating playing video and then detecting and classifying it as ‘Inappropriate Content’



In above video also we got classification output



In above we got result as Safe Content



In above screen we got output as Safe Content as peoples are only moving in the video.

Similarly you can upload and test other videos also