VISIONSAFE

Kavana K Nayak[1*], Dr. Shashidhar Kini K[2]

[1] *Student,* [2]*Head of the Department of MCA, Srinivas Institute of Technology, Valachil, Mangalore574143, Karnataka, India*

[1]kavanathumbe@gmail.com, [2]skinimca@sitmng.ac.in

ABSTRACT

Hearing about the violent activities that occur on a daily basis around the world is quite overwhelming. Personal safety and social stability are seriously threatened by the violent activities. The whole system can be implemented with a sequence of procedures. Firstly, the system has to identify the presence of human beings in a video frame. Then, the frames which are predicted to contain violent activities has to be extracted. The irrelevant frames are to be dropped at this stage. Finally, the trained model detects violent behaviour and these frames are separately saved as images. These images are enhanced to detect faces of people involved in the activity, if possible. The enhanced images is sent as an alert to the concerned authority through email.

Keywords: ***Violence, Alert Email.***

I. INTRODUCTION

The problem of violent conduct in public areas needs to be addressed. Violence also degrades communities because it diminishes output, depreciates property, and interferes with social services. Violence is a serious problem for public health all over the world. It has an impact on people at many stages of life, including infants and the elderly.

The main goal of this research is to create a system that can identify violence in uploaded videos. It will accurately forecast any violence in the video, and if any is detected, an email alert is issued to the appropriate authority. The project uses effective algorithms and optimisations to analyse video, making it possible to quickly detect violent content as videos are uploaded.

II. LITERATURE REVIEW

Recently proposed methods for violence detection can be roughly classified into three categories visual based approach, audio-based approach and hybrid approach. Visual-Based Approach: In this approach, visual data is retrieved and represented as pertinent features. There are two sorts of features: local features and global features. Local features include things like location, velocity, form, and colour, whereas global features include things like average speed, region occupancy, relative positional fluctuations, and interactions between objects and the background.

Audio Based Approach: Violence is categorised using audio data in the audio-based approach. To differentiate gunshots, explosions, and car braking in audio, a hierarchical method based on Gaussian mixture models and Hidden Markov models is used.

Hybrid Approach: In the hybrid approach, combining auditory and visual elements is the main focus. Some methods use blood and flame detection, capturing the amount of motion, as well as the usual sounds of violent events, to identify violent situations in recordings.

III. PROPOSED SYSTEM

The goal of the proposed research on video upload violence detection is to create a reliable and precise system that can recognise the violent content

inside the videos during upload. The system can manage a variety of violent situations, including those involving physical violence, firearms, sexual content, and other types of negative behaviour. The goal of the research is to create models that generalise effectively across various forms of violence, enabling accurate detection in a variety of settings.

The aim is to contribute to the development of the system that enhances user safety, supports content moderation effects, and promotes a safer and more responsible and safer online environment when using it.

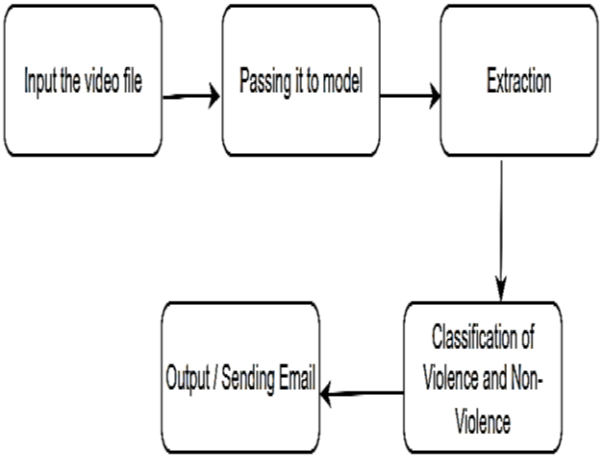

Fig 1. Context Flow Diagram of VISIONSAFE

The user will upload the video. During the uploading process it passes through the model and extract the frames present in the video. If violence found in the frame the email will be sent to the concerned authority or otherwise the non-violence is present then the output shows video is non-violence.

IV. METHODOLOGY

Long-term recurrent convolutional networks (LRCN) and convolutional neural networks (CNN) are the methodologies employed in VisionSafe. For applications involving sequential and spatial input, such as video analysis, LRCN is a hybrid model that combines the strength of both recurrent neural networks (RNN) and convolutional neural networks (CNN). The fundamental concept is to employ LSTMs to convert a series of picture embedding into a class label, sentence, probability, or whatever you require, and CNNs to learn visual features from video frames. As a result, a CNN processes the raw visual input, and the outputs are then fed into a stack of recurrent sequence models.

Convolutional neural networks (CNN) are a component of machine learning. It is one of a number of artificial neural network models used for various tasks and data sets. For deep learning algorithms, a CNN is a specific kind of network design that is used for tasks like image recognition and pixel data processing. Despite the fact that deep learning uses a variety of neural network types, CNNs are the preferred network design for detecting and classifying objects. Consequently, they are perfect for applications demanding crucial object identification and computer vision (CV) tasks.

V. EXPECTED RESULT

The expected result of violence and non-violence detection using a Convolutional Neural Network (CNN) would be a model capable of accurately classifying input data into two categories: violent and non-violent.

To achieve this, the CNN would need to be trained on a labelled dataset containing examples of both violent and non-violent content. The dataset should include a diverse range of images, videos, or any other relevant input format that represents various types of violence and non-violence.

During training, the CNN would learn to extract features from the input data using convolutional layers, followed by pooling layers to down sample the feature maps. The extracted features are then fed into fully connected layers to perform the final classification.

Once the video is uploaded through the interface, the CNN analyses the video and give the result. If the violence is found it gives the alert message to the concerned authority through email and if not found then the output will be shown as non-violence.

VI. CONCLUSION

Since the detection of hostile human actions is a prerequisite, violent action detection such as fight scene recognition has drawn the attention of computer vision researchers in recent years. However, deep representation-based transfer learning techniques have been utilised to recognise human actions including hand waving, jogging, and walking. However, utilising a deep model based on transfer learning to recognise violent sequences is rare. As opposed to other human action detection datasets that are currently accessible, the datasets used to train the model are the first of their type especially built for violent/fight action recognition. In this study, LRCN and a deep CNN model based on learnt representations are presented to detect hostile actions in films.

REFERENCES

[1] Lee MY, Khan SU, Haq IU, Rho S, Baik SW. Cover the Violence: An Innovative Deep Learning-Based Approach to Detecting Violence in Films. 2019; 9(22):4963 for Applied Sciences. https://doi.org/10.3390/app9224963J. Clerk Maxwell, Third Edition, Vol. 2, A Treatise on Electricity and Magnetism, Oxford: Clarendon, 1892, pp. 68–73.

[2] M. Ramzan and colleagues, "A Review on State-of-the-Art Violence Detection Techniques," IEEE Access, vol. 7, no. 10, 107560–107575, 2019, doi: 10.1109/ACCESS.2019.2932114.

[3] Nayak, L. Characterization of Violent Scenes Based on Audio-Visual Content. National Institute of Technology, Rourkela, Odisha, India, 2015, Ph.D. thesis.

[4] F.U.M. Ullah, A. Ullah, K. Muhammad, I.U. Haq, and S.W. Baik. Using a 3D convolutional neural network and spatiotemporal features, violence detection. 2019 sensors 19, 2472.

[5] MIC-TJU at MediaEval Violent Scenes Detection (VSD) 2014. Zhang, B.; Yi, Y.; Wang, H.; Yu, J. MediaEval 2014 Workshop Proceedings, Barcelona, Spain, 16-17 October 2014.

[6] Nievas, E.B., Suarez, O.D., Garca, G.B., and Sukthankar, R. Computer vision algorithms for violence identification in video. Springer: Berlin/Heidelberg, Germany; International Conference on Computer Analysis of Images and Patterns; 2011; pp. 332–339.

[7] A Novel Violent Video Detection Scheme Based on Modified 3D Convolutional Neural Networks by Song, W., Zhang, D., Zhao, X., Yu, J., Zheng, R., and Wang, A. 2019 IEEE Access 7, 39172–39179.