# Harmful Act Detection Using Deep Learning

**Sourabh Gulaiya, Tushar Kaushik, Vaibhav Tomar, Prashant Singh, Jagbeer Singh**

(sourabh.gulaiya.cs.2019@miet.ac.in, tushar.kaushik.cs.2019@miet.ac.in, vaibhav.tomar.cs.2019@miet.ac.in, prashant.singh.csl.2019@miet.ac.in, jagbeer.singh@miet.ac.in)

Meerut Institute of Engineering and Technology, Meerut

**Abstract**

Harmful recognition is a process of recognizing harmful activities in public, private places through video monitoring. Harmful recognition can be used to monitor harmful act in public places through video monitoring. In this paper, convolutional neural network(CNN) algorithm is used with Mobilenet v2 architecture and OpenCV for harmful act recognition. In this model/system, we use various real lifebased videos to train our model/system in which half of the data-set contains harmful videos and another half contains safe videos. The proposed model fulfill the accuracy of 90%.

**Keywords:** Mobilenet V2, Harmful recognition,supervised learning, convolutional neural network(CNN) algorithm, Computer Vision, Deep learning, harmful.

## 1 Introduction

Harmful act recognition using deep learning is a difficult task in computer science domain. Computer vision provides various approach to work on video and images data-sets.[1] In this research paper, harmful act recognition system is proposed which takes videos as input, it may contain audio or may not and recognizes harmful activities in frames resulting to from the input video and shows each frame as safe or harmful as output. This research paper aims on taking frames from the input video coming from CCTV, recorded videos or other sources.[2] After trying some common approaches for harmful act recognition, CNN + Mobilenet v2 is selected for this paper.

**MobileNetV2** used on mobile devices. It helps in classification of images, for example: from a dog's image it will detect it a dog. It take image of 224*224 pixels which have three colour channel. In CNN we use filters/kernels which run on a image and calculate dot product. Every filter takes out different features from image. [3] CNN is satisfactory at picking up on designs in the input image, such as gradients, lines circles, or even characteristics faces and eyes. [4] This research paper uses CNN as major working algorithm to train our model. CNN is used to extract characteristicsfrom frames, learn from images and report them as harmful or safe for humans [6][7].

The main task of this paper is to cut human intervention into the monitoring of videos collected from various public or private places CCTV cameras to recognise the harmful activities happening there.[5] We aim that this paper will help in facilitateharmful act recognition and activities happening in computer vision and deep learning domains.[6]The remaining portion of this paper contains Section 2 consists Literature Review, Section 3 contains Proposed Methodology, Section 4 contains Result Discussion and Section 5 contains conclusion [8][9][10].

## 2.Literature review

With the increasing number of monitoring cameras in metropolitan cities, huge number of videos can be stored.While there are insufficient human resource for monitoring many screens at one time. The techniques of video understanding to recognize harmful behaviour.[7] In this research statistical analysis morphological and threshold techniques are used to process the pictures obtained from sample of harmful videos. Convolutional Neural Network (CNN) model will used for the same. There are a number of researches on this topic which uses different classification techniques.[8] This problem can be solved by using (support vector machine) SVM and (k-nearest neighbour) KNN also, but CNN with MobileNetV2 is used in this paper to solve this problem [11][12].

In December 2019, Solimanintroduced a model which used a pretrained VGG16 on ImageNet as spatial feature extractor than it moves to Long Short-Term Memory (LSTM) as sequences which havecompletely connected layers for classification cause.
The data-set used was used for real time harmful situation which includes 2000 videos splited into 1000 harmful videos and 1000 nonharmful videos. The presented models achieved a precision of 88.2%. [12] [13]

## Dataset

In this paper, Real Life Harmfulact data-set is used which consist of videos having harmful and other are safe video clips containing normal activities. Both type of videos are placed in different directories. This dataset contains 1000 harmful videos and 1000 non-harmful videos. In this paper, only 350 harmful and 350 nonharmful video clips are used to train the model due to lack of memory [14][15] [16][17].

## 3. Proposed Methodology

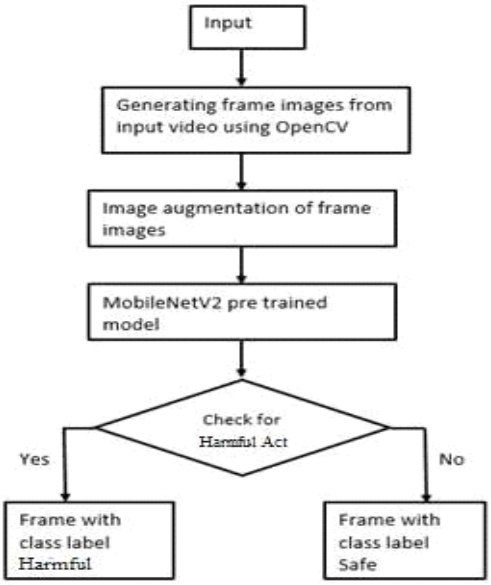


Figure 1: Block Diagram(ERD)

**The above block diagram shows**and describe the control flow of various steps which includes frame generation,image augmentation and examine for harmful and labelling frames with their respective class after classification [18][19][20][21]

### Step 1: Data-set splitting

In this paper, 70% of our data-set videos are used for training and 30% of videos are used for validation. This means 245 harmful videos and 245 non-harmful videos are used for training purpose and 105 harmful and 105 non-harmful videos are used for validation purpose [22][23][24].

### Step 2:Processingof data-sets

Here frames are generated from video clips using Computer Vision tool OpenCv2. The frames are augmented and preprocessed further. Augmentation is done to magnify the size of dataset to conquer the problem of overfitting. Image frames are taken out from the video clips and each frame is stored in the data-set. Then the size of frame is changed into 128*128*3 to decrease computational time [25][26].

### Step 3: Neural network model development

The data-set is breaked into testing andtraining data-set. Then a MobileNetV2 pre-trained model with CNN classifiers for the classification of frames is equipped with training dataset. Each frame is feed into a neural network. It goes through these layers: 1)Zero padding layer, 2)Convolutional layer, 3)Batch normalization layer, 4)Sigmoid activation layer, 5) Maxpooling layer (two times), 6)Flatten layer, 7)Dense fully connected layer with one neuron [27][28][29].

**Step 4: Experimentation and training on data-sets**

Training and testing are done on MobileNetV2 pre-trained model with CNN classifier on the data-set to do the prediction accurately. The model is directed for 50 epochs(stages) and loss and accuracy plots are constructed. The accuracy of the model is calculated [30][31][32][33].

**Convolutional neural network (CNN)**

They are made up of node levels, each of which includes an input layer, one or more hidden layers, and an output layer. Each node has a threshold and weight that are connected to one another. Any node whose output exceeds the defined threshold value is activated and begins providing data to the network's uppermost layer. Otherwise, no data is transmitted to the network's next tier [14,15].

• Convolutional neural networks outperform other neural networks when given inputs like images, voice, or audio, for example. Convolutional, pooling, and completely linked layers are their three primary types of layers [16,17].

CNN algorithm with MobileNetV2 already trained model which will be used to take outcharacterstics from frames, on using Real Time harmful act situation data-set with this model total no. of 1281 trainable parameters are acquired [18].

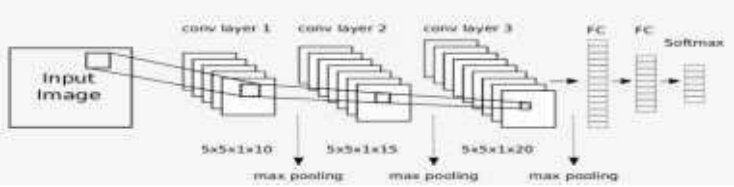


**Fig 2**: Convolutional Neural Networks Layer

**Configuration of layers**

Input layer extract an image frame of size 128*128 with 3 colours channels. Then it goes to first cnn layer with an image size of 64*64 with 32 parameters. Then furtherreached to the normalization layer and then ReLU layer. After that neural network is expanded depth wise and the same process is repeated for 16 blocks and after the size of 4*4 with 1280 is obtained. Then it further moved through global average pooling 2D layer and lastly in the dense layer 1281 trainable parameters are obtained [34][35].

**4.Experimental Results**

After training our model, the model is capable to generate frames from video clips and recognise whether it is having harmfulact or it is safe.

**4.1 Confusion Matrix**

It is used to show the performance of a classification model.

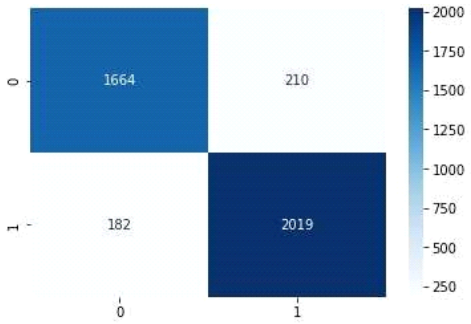


**Fig 3 :** Model'sConfusion Matrix

**The above figures shows that**the confusion matrix of our model, whichstates that the true negative values are 1664, true positive values are 2019, false positive values are 210 and false negative values are 182.

### 4.2 Output

Below are the screenshots of our output which shows the label on eachframeas harmful or safe.


**Fig 4: H**armful act detected

Thefig 4 shows that the output frame with harmful label is in red color which shows that the above image frame contains harmful act.


**Fig 5 :** Safe action detected

The fig 5 shows that the output frame with no harmful act and safe label in green colorthat shows the above image frame does not contain harmful act.

### 4.3 Classification of Report

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| NonViolence | 0.90 | 0.89 | 0.89 | 1874 |
| Violence | 0.91 | 0.92 | 0.91 | 2201 |
| accuracy | | | 0.90 | 4075 |
| macro avg | 0.90 | 0.90 | 0.90 | 4075 |
| weighted avg | 0.90 | 0.90 | 0.90 | 4075 |

**Fig 6 :** Classification of report**(accuracy=90%)**

### 4.4Loss in Validation and Training

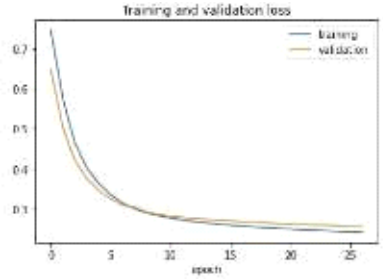


**Fig 7:**Loss in Validation and training

**The fig 7** shows the loss in validation and training of our model ,where orange line depicts loss in validation and blue line depicts loss in training.

### 4.5 Accuracy in Training and Validation

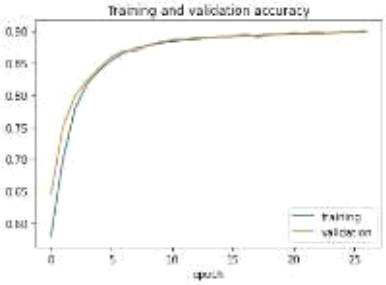


**Fig 8:**Accuracy in Training and validation

**The fig 8**depicts the accuracy in training and validationof our model ,where orange line depicts validation accuracy and blue line depicts training accuracy.

### 5.Conclusion

From above model we have concluded that our model takes video as input and breaks it into frames and then label them as harmful or safe in red and green colour respectively by using CNN algorithm and gets an accuracy of 90 percent till now. This model is able to recognise harmful act from CCTV, recorded videos based on training data set. We hope that this model will help in reducing crime and harassmentat public/private places by monitoring all the videos simultaneously.

### 6. References


- Author: AurelienGeron, is a machine learning consultant wrote a book “Hands on machine learning” dated : 25, April, 2017.
- Author: Gary Bradski and Adrian Kaehlerof book “Learning Opencv”, editor: Mike loukides, printing date: September, 2008.
- Author:Dr. Simon j.d.princefaculty member in computer science in

London university, of book “Computer vision: models learnigs and inference”, published on June, 2012.

. Sharma and R. Baghel, "Video Surveillance for Violence Detection Using Deep Learning," Adv. Data Sci. Manag

- Author: ChristopherM.bishop is assistant director at microsft research, of book “Pattern recognition and machine learning”, published on August, 2006.
- Mohamed Elgendy is a author of book “Deep Learning For vision System”, published on November, 2020.
- O Theobald is a author of book, “Introduction to Machine Learning”, published on June, 2017.
- Author: Andrew Trask is a Ph.D. student at oxford university, of book “Grokking Deep learning” published on January, 2019.
- https://youtu.be/88ndrvAZpE8 HYPERLINK "https://youtu.be/88ndrvAZpE8" “youtube” Abto violence detection technology is aimed to help ensure public safety through visual crowd surveillance.
- Trevor Hastie, Robert Tibshirani, Jerome Friedman are author of book “The Elements Of Statistical Learning”.
- Data sets on www.kaggle.com, videos containing harmful video data set.
- W3schools: https://www.w3schools.com/ai/ai_neural_networks.asp HYPERLINK "https://www.w3schools.com/ai/ai_neural_networks.asp".
- Sharma, S., & Sodhi, J. S. (2014). Implementation of biometric techniques in social networking sites. International Journal of Security and Its applications, 8(6), 51-60.
- Jain, S., & Sharma, S. (2018, August). Application of data warehouse in decision support and business intelligence system. In 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT) (pp. 231-234). IEEE.
- Gulati, S., & Sharma, S. (2020). Challenges and responses towards sustainable future through machine learning and deep learning. Data Visualization and Knowledge Engineering: Spotting Data Points with Artificial Intelligence, 151-169.
- Puri, A., & Sharma, S. (2020, June). Risk management in software engineering using big data. In 2020 International Conference on Intelligent Engineering and Management (ICIEM) (pp. 63-68). IEEE.

- Srivastava, S., & Sharma, S. (2019, January). Analysis of cyber related issues by implementing data mining Algorithm. In 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 606-610). IEEE.

- Smriti, P., Srivastava, S., & Singh, S. (2018, February). Keyboard invariant biometric authentication. In 2018 4th International Conference on Computational Intelligence & Communication Technology (CICT) (pp. 1-6). IEEE.

- Srivastava, S., & Singh, P. K. (2022). Proof of Optimality based on Greedy Algorithm for Offline Cache Replacement Algorithm. International Journal of Next-Generation Computing, 13(3).

- Narayan, Vipul, and A. K. Daniel. "Novel protocol for detection and optimization of overlapping coverage in wireless sensor networks." Int. J. Eng. Adv. Technol 8 (2019).

- Narayan, Vipul, and A. K. Daniel. "FBCHS: Fuzzy Based Cluster Head Selection Protocol to Enhance Network Lifetime of WSN." ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal 11.3 (2022): 285-307.

- Narayan, Vipul, et al. "To Implement a Web Page using Thread in Java." (2017).

- Mall, Pawan Kumar, et al. "FuzzyNet-Based Modelling Smart Traffic System in Smart Cities Using Deep Learning Models." Handbook of Research on Data-Driven Mathematical Modeling in Smart Cities. IGI Global, 2023. 76-95.

- Irfan, Daniyal, et al. "Prediction of Quality Food Sale in Mart Using the AI-Based TOR Method." Journal of Food Quality 2022 (2022).

- Awasthi, Shashank, et al. "A Comparative Study of Various CAPTCHA Methods for Securing Web Pages." 2019 International Conference on Automation, Computational and Technology Management (ICACTM). IEEE, 2019.

- Narayan, Vipul, A. K. Daniel, and Ashok Kumar Rai. "Energy efficient two tier cluster based protocol for wireless sensor network." 2020 international conference on electrical and electronics engineering (ICE3). IEEE, 2020.

- Narayan, Vipul, et al. "E-Commerce recommendation method based on collaborative filtering technology." International Journal of Current Engineering and Technology 7.3 (2017): 974982.

- Choudhary, Shubham, et al. "Fuzzy approach-based stable energy-efficient AODV routing protocol in mobile ad hoc networks." Software Defined Networking for Ad Hoc Networks. Cham: Springer International Publishing, 2022. 125-139.

- Narayan, Vipul, and A. K. Daniel. "Energy Efficient Protocol for Lifetime Prediction of Wireless Sensor Network using Multivariate Polynomial Regression Model." Journal of Scientific & Industrial Research 81.12 (2022): 1297-1309.

- Smiti, Puja, Swapnita Srivastava, and Nitin Rakesh. "Video and audio streaming issues in multimedia application." 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2018.

- Srivastava, Swapnita, and P. K. Singh. "HCIP: Hybrid Short Long History Table-based Cache Instruction Prefetcher." International Journal of Next-Generation Computing 13.3 (2022).

- Smriti, Puja, Swapnita Srivastava, and Saurabh Singh. "Keyboard invariant biometric authentication." 2018 4th International Conference on Computational Intelligence & Communication Technology (CICT). IEEE, 2018.

- Mall, Pawan Kumar, et al. "Early Warning Signs Of Parkinson's Disease Prediction Using Machine Learning Technique." Journal of Pharmaceutical Negative Results (2023): 2607-2615

- Narayan, Vipul, et al. "Deep Learning Approaches for Human Gait Recognition: A Review." 2023 International Conference on Artificial Intelligence and Smart Communication (AISC). IEEE, 2023.

- Narayan, Vipul, et al. "FuzzyNet: Medical Image Classification based on GLCM Texture Feature." 2023 International Conference on Artificial Intelligence and Smart Communication (AISC). IEEE, 2023.

- Pentyala, S., Liu, M., & Dreyer, M. (2019). Multi-task networks with universe, group, and task feature learning. arXiv preprint arXiv:1907.01791.