iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.44900

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Suspicious Human Activity Recognitionand Alarming System

, , Leela S[1], K.V Sai Likhita[2] Deepak Kumar[3], A Abhiram[4], Dr. Keerthika V[5]
[1 2, 3, 4]Final Year Students, [5]Associate Professor, Computer Science and Engineering, DSATM

***Abstract:** Personal safety and social stability are seriously threatened by the violent activities. A variety of methods have been tried to curb the violent activities which includes installing of surveillance systems. It will be of great significance if the surveillance systems can automatically detect violent activities and give warning or alert signals. 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. The images along with other necessary details such as time and location is sent as an alert to the concerned authority. The proposed method is a deep learning based automatic detection approach that uses Convolutional Neural Network to detect violence present in a video. But, the disadvantage of using just CNN is that, it requires a lot of time for computation and is less accurate. Hence, a pre-trained model, Mobile Net, which provides higher accuracy and acts as a starting point for the building of the entire model. An alert message is given to the concerned authorities using telegram application.*

***Keywords:** Human activity recognition · Violence detection · Alarming System · Vision-based human activity recognition · Anomaly detection*

I. INTRODUCTION

Anomaly detection in security systems is one of the instances in which human activity recognition can be used. An activity recognition system should be able torecognise the basic tasks that a human performs on a daily basis. Because of the complexity and diversity of human activities, achieving high accuracy for recognition of various activities is difficult. Different methodologies unique to the application areused to build activity models required for theidentification and classification of human activities. Video processing and machine learning groups are drawn to this research subject because it has applications in a variety of fields, including medicine and health care, human–computer interaction, crime investigation, and security systems.

Human activity recognition has applications outside of healthcare and security.Behavioral biometrics, video analysis, animation, and synthesis are all examples ofthis ongoing and open research field in computer vision. Sensors are used to interpret human behaviours involving gestures and motions of the human body. The identification of human actions requires theseinterpretations.

II. RELATED WORK

D. G. Shreyas, S. Raksha & B. G. Prasad (2020) [1] proposed that by using the intermediate result of adaptive video compression an accurate, real-time anomaly detection system is implemented. The proposed method outperforms other existing systems in terms of accuracy and timeliness.

Santosh Kumar Yadav, Kamlesh Tiwari, Hari Mohan Pandey & Shaik Ali Akbar (2021) [2] presented a privacy-preserving activity recognition and fall detection system using a single Kinect (v2) sensor and ConvLSTM. The proposed system derives geometrical and kinematic features and passes them along with the raw skeleton coordinates into deep learning networks.

Djamila Romaissa Beddiar, Brahim Nini, Mohammad Sabokrou & Abdenour Hadid (2020) [6] proposed a classification according to several criteria. It initially discusses the different applications of HAR, and the major objectives intended by these systems. Then, it presents an analysis of the used approaches in the state of the art, as well as the means used in their validation.

Giovanni Ercolano; Daniel Riccio; Silvia Rossi [3] presented a CNN-LSTM model for activity recognition working on a matrix representation of the skeleton joints

K.Ishwaryaa and A.Alice Nithyab (2021)[4] produces the detail study of the proposed model review. In that analysis various time series methods deeply analysed to correctly identify the daily activity effectible.

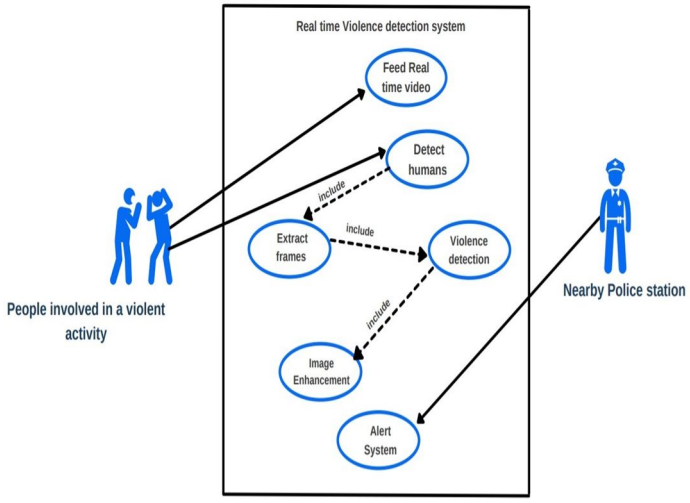
Jaspal Kumar ,M.Kulkarni Davya gupta (2016)[5] proposed an intelligent video surveillance system to detect and track the human body and to recognize the type of the daily human movement and activity is introduced.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

III. METHODOLOGY

We propose a deep learning model composed of *convolutional* and *Long Short-Term Memory* recurrent layers, which can automatically learn local features and model the relation between features. Theimplementation of such a system requires the videos from surveillance cameras to generate frames. Footage from the surveillance camera is broken down into frames. The frames are given as input to MobileNet v2 classifier for detecting violent activities in the given sequence of input frames. If no violent activity is recognized the respective frames are discarded. The violence detected frame is obtained and it is enhanced for better clarity. That frame, along with the location are sent to the nearest authorities using Telegram bot

Convolutional Neural Networks (CNN) are great for image data and Long-Short Term Memory (LSTM) networks are great when working with sequence data but when you combine both of them, you get the best of both worlds and you solve difficult computer visionproblems like video classification.



A. *Tensorflow*

We'll use a CNN to extract spatial characteristics at a specific time step in the input sequence (video), then an LSTM to find temporal relationships between frames. ConvLSTM and LRCN are the two architectures we will utilise to combine CNN and LSTM. TensorFlow may be used for each of these approaches. TensorFlow is an open source machine learning platform that runs from start to finish. It has alarge, flexible ecosystem of tools, libraries, and community resources that allow academics to advancethe state-of-the-art in machine learning and developersto quickly construct and deploy ML applications.

B. *Keras*

Keras is a lightweight deep learning Python package that may be used with Theano or TensorFlow. It was created to facilitate the implementation of deep learning models for research and development as simple as feasible. Given the underlying frameworks, it operates on Python 2.7 or 3.5 and can run on both GPUs and CPUs. It's under the MIT licence, which allows for a lot of flexibility. François Chollet, aGoogle developer, created and maintains Keras basedon four principles:

Modularity: A model can be understood as a sequence or a graph alone. All the concerns of a deep learning model are discrete components that can be combined in arbitrary ways.

Minimalism: The library provides just enough to achieve an outcome, no frills and maximizing readability.

Extensibility: New components are intentionally easy to add and use within the framework, intended for researchers to trial and explore new ideas.

Python: No separate model files with custom file formats. Everything is native Python.

Keras does not do low-level operations like tensor products and convolutions on its own; instead, it uses a back-end engine. Despite the fact that Keras supports many back-end engines, TensorFlow is its major (and default) back end, and Google is its principal backer. TensorFlow includes the Keras API as tf.keras, which, as previously stated, will become the primary TensorFlow API as of TensorFlow 2.0. The Model is the most basic data structure in Keras. The Sequential model and the Model class used with the functional API are the two major types of models accessible in Keras. Sequential Keras models The Sequential model is a linear stack of layers with relatively simple layer descriptions.

C. LSTM (Long-Short Term Memory)

Long-Short Term Memory is an artificial recurrent neural network (RNN) architecture we use in deep learning. RNN is a class of artificial neural network where connections between nodes form a directed/in directed graph along a temporal sequence, exhibiting temporal dynamic behavior. It basically uses memory(internal state) to process variable length sequences ofthe input. Unlike the normal feedforward neural networks which do not form a cycle, LSTM has feedback connections. Our project needs to process video and speech apart from just images; hence LSTMis well suited for this since it can process not only single data points but also the sequences of data. Our main focus which is basically on detecting and capturing the anomaly activities, LSTM which is well applicable for unsegmented tasks such as anomaly detection or intrusion detection is much preferred.

Here we basically need a system to store and process the data and handle the information accordingly. A common LSTM unit usually consists of a cell, an input gate, an output gate and a forget gate. The cell remembers the values over the arbitrary time intervals,whereas the 3 gates regulate the flow of theinformation in and out of the cell. Since, classifying, processing and making predictions based on the time series is an important part by which the anomalies are to be detected, LSTM plays a vital role in making thispossible.

The LSTM cell can process data sequentially and keepits hidden state through time. This algorithm helps us to connect the previous information to connect it to thepresent task which is happening or predicted to happen. Using its default behavior to remember things for long without any struggle, is a major advantage wehave in using this algorithm to our project.

$$
\begin{aligned}
f_t &= \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \\
i_t &= \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \\
o_t &= \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \\
\tilde{c}_t &= \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \\
c_t &= f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \\
h_t &= o_t \circ \sigma_h(c_t)
\end{aligned}
$$

In this Algorithm, firstly the decision of throwingaway or keeping the information is done by the forgetgate. It outputs a number between 0 and 1 for each number in cell state where 0 represents to completely get rid of info and 1 represents to keep that info completely. In the next step the input layer will decidewhat values to update and a vector for new candidate state is created and both are combined to create an update to the state. Now we update the old state. We multiply it with old state, forgetting things we decidedto forget before. Finally, we decide what we output based on the cell state.

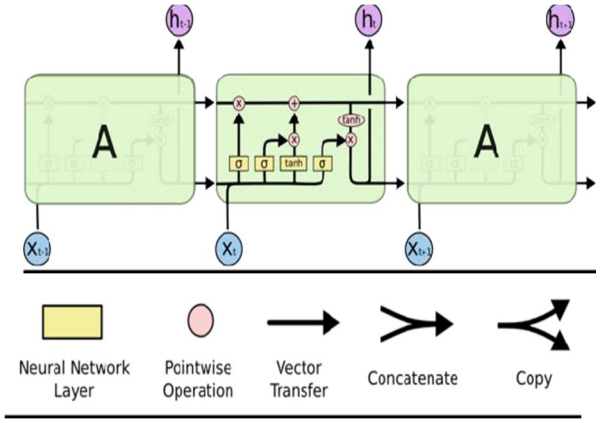


Figure 3.1 Working of LSTM Algorithm

D. CNN

Convolutional neural network models were created forimage classification challenges, and feature learning isthe process by which the model learns an internal representation of a two-dimensional input.

This similar approach can be used to recognise human activity from one-dimensional sequences of data, such as acceleration and gyroscopic data. The model learnshow to extract features from observation sequences and how to map internal features to various activity kinds.

The advantage of utilising CNNs for sequenceclassification is that they can learn directly from raw time series data, eliminating the need for domain knowledge to manually build input features. The model should be able to learn an internal representation of the time series data and, in theory, perform similarly to models trained on a version of thedataset with artificial features.
This section is divided into 4 parts; they are:

1) Load Data
2) Fit and Evaluate Model
3) Summarize Results
4) Complete Example

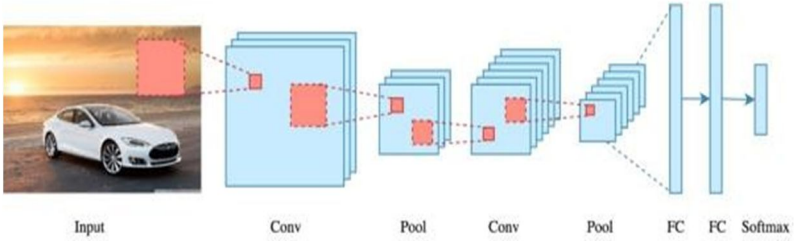


Figure 3.2 Working of CNN Algorithm

Multiple layers of artificial neurons make up convolutional neural networks. Artificial neurons are mathematical functions that calculate the weighted sum of various inputs and output an activation value, similar to their biological counterparts. Each layer creates many activation functions that are passed on tothe next layer when you input an image into a ConvNet.
Basic features such as horizontal or diagonal edges areusually extracted by the first layer. This information ispassed on to the next layer, which is responsible for detecting more complicated features like corners and combinational edges. As we go deeper into the network, it can recognise even more complex elements like objects, faces, and so on.

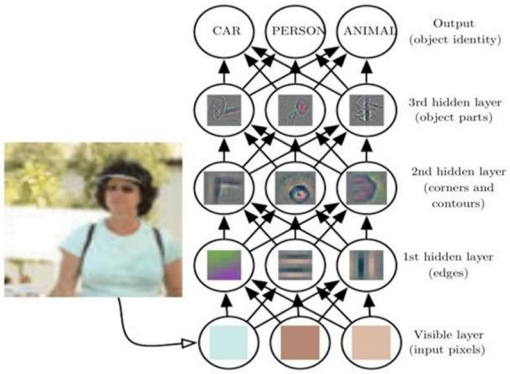


Figure 3.3 Illustration of human activity detection model

E. Dataset

The dataset contains 1000 video clips which belongs to two classes, violence and non-violence respectively. The average duration of the video clips is 5 seconds and majority of those videos are from CCTV footages. For training, 350 videos each from the violent and non-violent classes are taken at each epoch.

F. Mobilenet V2

The MobileNet architecture is primarily based on depth wise sep- arable convolution, in which factors a traditional convolution into a depth wise convo- lution followed by a pointwise convolution.
The module presents a residual cell (has a residual/identity connection) with stride of 1, and a resizing cell with a stride of 2. From Figure 4, "conv" is a normal con- volution, "dwese" is a depth wise separable convolution, "Relu6" is a ReLu activation function with a magnitude limitation, and "Linear" is the use of the linear function.
The main strategies introduced in MobileNetV2 were linear bottleneck and inverted residual blocks.

In the linear bottleneck layer, the channel dimension of input is ex- panded to reduce the risk of information loss by nonlinear functions such as ReLU. It stems from the fact that information lost in some channels might be preserved in other channels. The inverted residual block has a ("narrow" -"wide"-""narrow") structure in the channel dimension whereas a conventional residual block has a ("wide" - "narrow"- "wide") one. Since skip connections are between narrow layers instead of wider ones, the memory footprint can be reduced.

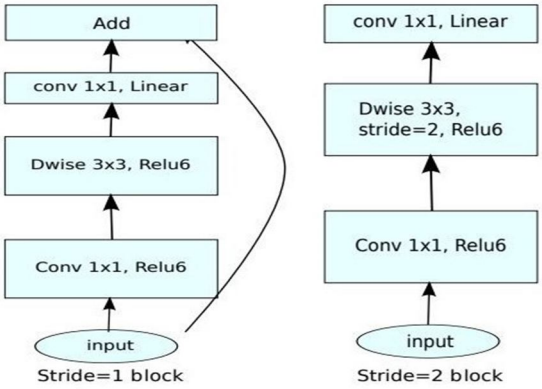


Figure 3.4 MobileNet v2 Architecture

G. Alert Module

The alert module sends alert message to the specified authority. Figure 3.5 describes the architecture of the implemented alert system. When a frame is detected true for violence, the system initialises a counter variable to one. Then it checks the subsequent 30 frames, whether if they too have violence detected true.

The counter is incremented at each consecutive frame that is true for violence. If a frame is false for violence, the counter variable is set to 0 and starts checking the consecutive frame respectively checking whether violence is recognized. On the other hand, if the violence is detected true for the 30 consecutive frames, the current time is obtained using an inbuilt python function and an alert is sent to a Telegram group that consists officials of higher authorities. The Alert message comprises of an image of the detected violent activity, current timestamp and the location where the camera is placed.

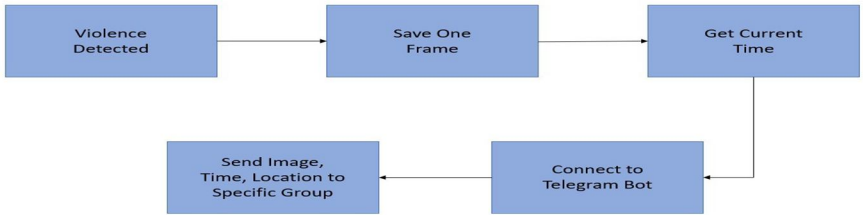


Figure 3.5 Architecture diagram of the Alert System

Figure 3.6 shows the alert message that is sent to the telegram group by the telegram bot. The concerned authorities can view the alert and take necessary actions.

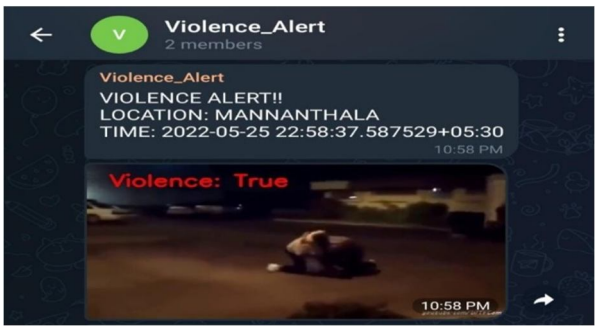


Figure 3.6 Screenshot of the alert message

IV. RESULT

Fig. 4.1 displays the training and testing accuracy and loss for the MobileNet v2 model when a dataset containing 1000 videos of average duration 7 seconds is given as input. For each epoch 350 videos from the violence class and 350 videos from the non-violence are trained. 96% accuracy was obtained on training and a respective accuracy of 95% was obtained when a
CCTV footage that was not included in the dataset was given for testing. The obtained output video frames are shown in Figure 4.3.
In the graph in Figure 4.1 the accuracy and loss comes to a constant level of incre- ment and decrement after approximately 5 epochs. The obtained confusion matrix and other evaluation parameters are shown in Fig. 4.1
A video with violence is given as input to the system.
Figure 4.3 shows one frame in the video that was labeled to have violent activity. Another video clip without violent activity was given as input. Figure 4.4 shows one frame of that video which is rightly labelled as false or violence.

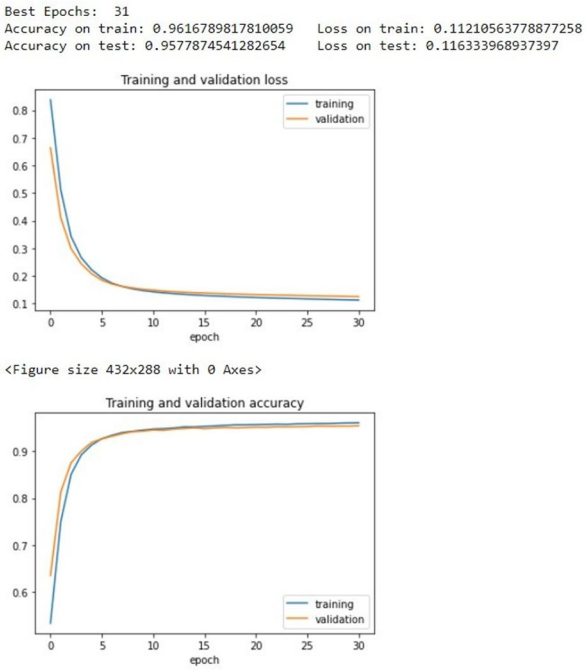


Figure 4.1 Accuracy and Error of the training set

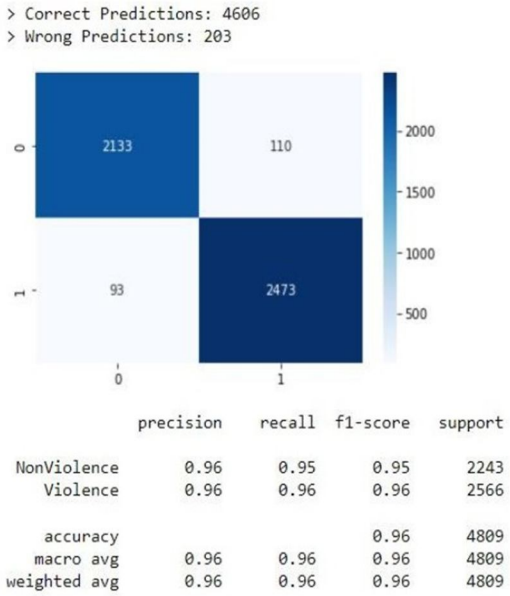


Figure 4.2 Confusion matrix of the trained model

Figure 4.3 Output frame that recognized violence

Figure 4.4 Output frame that did not recognize violence

As the comparison shown in the Fig 4.1 MobileNet v2 has shown improvement than CNN-LSTM in the violence detection task. The above shown graphs has shown that MobileNet v2 is capable of performing better than model trained using CNN- LSTM. This proves that MobileNet v2 too can become the state of the art model for Real-time Violence detection.

	precision	recall	f1-score	support
0	0.93	0.82	0.87	66
1	0.85	0.95	0.90	74
accuracy			0.89	140
macro avg	0.89	0.88	0.88	140
weighted avg	0.89	0.89	0.88	140

Figure 4.5 Evaluation metrics of the CNN-LSTM Model

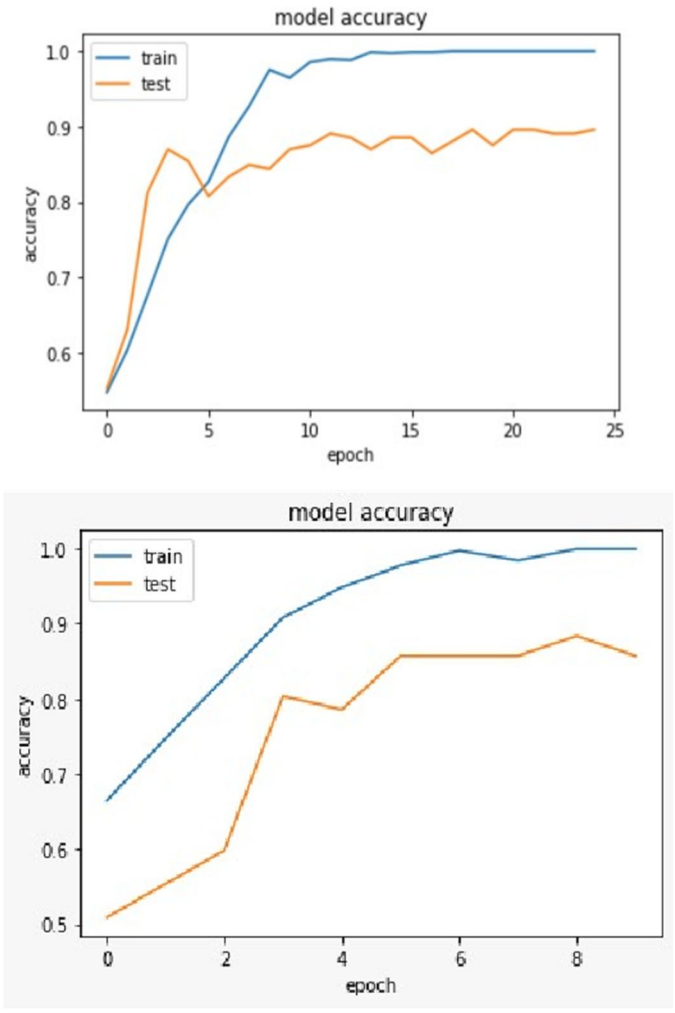


Figure 4.6 Comparison of Training and Testing accuracy of MobileNet v2 and CNN- LSTM Models

V. CONCLUSION

Theft, accidents, graffiti, fighting, chain snatching, crime, and other suspicious actions have become increasingly common in recent years. We can't rely on traditional procedures that require a human being to monitor the system on a constant basis. Our project primarily addresses this issue, which necessitates the development of an automated system that detects unusual behaviours in the environment and alerts the end user.

In this research, we use the MobileNet v2 model to offer an innovative and efficient technique for identifying violent events in real-time surveillance footage. The proposed network has a good recognition accuracy in typicalbenchmark datasets, indicating that it can learn discriminative motion saliency maps successfully. It's also computationally efficient, making it ideal for use in time-critical applications and low-end devices. Here, we had also shown the working of an Alert system that is integrated with the pretrained model. In comparison to other state-of- the-art approaches, this methodology will give a far superior option.

VI. FUTURE SCOPE

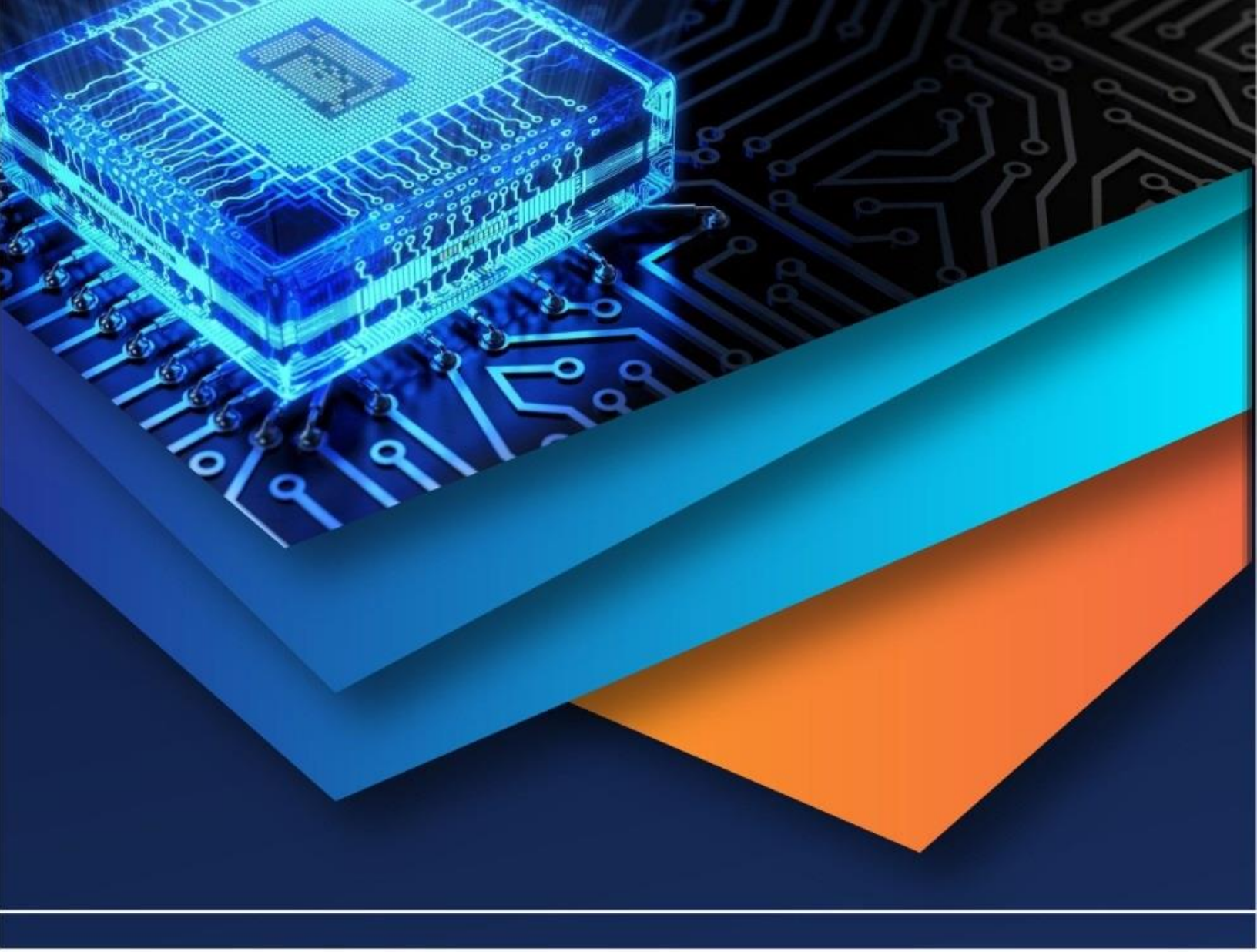
This model could be upgraded to work in multiple cameras con- nected by a single network in a concurrent fashion. A short video of the violent activitycould be incorporated along with the alert message.

This concept can be extended to numerous public locations, precise to the application surroundings like schools, colleges, airports, bus stops, hospitals and railwaystations primarily based totally on their precise requirements.

It can be integrated to right away call anambulance while road accidents are detected. Thus, using the intermediate result of adaptive videocompression an accurate, real-time anomaly detectionmachine is implemented.

REFERENCES

[1] D.G.Shreyas, S. Raksha & B. G. Prasad SN Computer Science volume 1, Article number: 168 (2020) "Implementation of an Anomalous Human Activity Recognition System"
[2] Santosh Kumar Yadav, Kamlesh Tiwari, Hari Mohan Pandey & Shaik Ali Akbar Soft Computing (2021)"Skeleton-based human activity recognition using ConvLSTM and guided feature learning"
[3] Giovanni Ercolano;Daniel Riccio;Silvia Rossi 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) Year: 2017 | Conference Paper | Publisher: IEEE "Combining CNN and LSTM for activity of daily living recognition with a 3D matrix skeleton representation"
[4] K.Ishwaryaa , and A.Alice Nithyab (2021) "Human activity recognition: A review"
[5] Mohanad Babiker, Othman Omran Khalifa, Aisha Hassan Abdalla Hashim" Automated Daily Human Activity Recognition for Video Surveillance Using Neural Network"
[6] Djamila Romaissa Beddiar, Brahim Nini, Mohammad Sabokrou & Abdenour Hadid ,Multimedia Tools and Applications volume 79, pages 30509–30555 (2020) "Vision-based human activity recognition"
[7] Department of Computer Science and Engineering, University of Ioannina, Ioannina, Greece| Computational Biomedicine Laboratory, Department of Computer Science, University of Houston, Houston, TX, USA "A Review of Human Activity Recognition Methods"
[8] Li Wei and Shishir K. Shah Computer Science Department, University of Houston, 3551 Cullen Blvd., Houston, TX 77204, U.S.A "Human Activity Recognition using Deep Neural Network with Contextual Information"
[9] Wenchao Xu; Yuxin Pang; Yanqin Yang; Yanbo Liu "Human Activity Recognition Based On Convolutional Neural Network"
[10] Florenc Demrozi; Graziano Pravadelli; Azra Bihorac; Parisa Rashidi "Human Activity Recognition Using Inertial, Physiological and Environmental Sensors: A Comprehensive Survey"



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY