A Hybrid CRNN Model for Multi-Class Violence Detection in Text and Video

Premanand Ghadekar[1*], *Kunjal* Agrawal[1], *Adwait* Bhosale[1], *Tejas* Gadi[1],*Dhananjay* Deore[1] and *Rehanuddin* Qazi[1]

[1]Vishwakarma Institute of Technology, Pune, India

Abstract. Gender-based violence is a critical issue that not only poses a threat to physical safety but also has significant impacts on mental health. Shockingly, up to 1 billion children aged 2-17 years are estimated to have experienced gender-based violence globally, making it a pressing concern for the machine learning and deep learning communities to address. To end this, a novel approach has been proposed in the form of a Convolutional Neural Network and bi-directional LSTM (CRNN) to classify three types of violence present in both text and video data, thereby making the internet a safer space for individuals. The proposed approach utilises two datasets consisting of 400 and 600 samples each for videos and text, respectively, to improve the precision and accuracy of the model. The use of a Convolutional Recurrent Neural Network framework combined with LSTM layers has resulted in an accuracy of 97% on text and 96% on videos, surpassing the performance of existing RNN models. Additionally, the inclusion of dropout and regularizer layers has helped the model avoid overfitting and generalise better on unseen data. Overall, the CRNN-based approach presents a promising solution to the problem of gender-based violence detection, with the potential to significantly improve the safety of individuals online. By leveraging the power of machine learning and deep learning, we can contribute towards creating a safer and more equitable world for all.

1 Introduction

Gender-based violence (GBV) is a pervasive and complex social phenomenon that significantly impacts the mental and physical health of victims. While international organisations have recognized the importance of addressing GBV and many countries have implemented policies and programs to prevent and respond to it, much work remains to be done. Machine learning and deep learning communities have a critical role in developing practical approaches to detecting and preventing GBV. Violence is a leading cause of death and disability worldwide, particularly among young people. Deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have shown promising results in detecting different forms of violence in text and videos.

*Corresponding author: adwaitbhosale04@gmail.com

Recent research has focused on developing novel approaches to classify violence accurately in text and videos while addressing data imbalance and noise. One method involves using multimodal techniques that combine text and video data with improving the model's accuracy. Additionally, researchers have explored the effectiveness of different deep-learning architectures and optimization techniques for violence classification. Such models are trained on large datasets containing both text and video data, enabling them to learn complex patterns and features indicative of violent behaviour.

There are still challenges to overcome in developing effective violence detection models, such as dealing with the variety of languages and dialects used in text data and the complexity of visual cues in video data. However, ongoing research in this area holds promise for helping to prevent and respond to GBV. By improving the accuracy and precision of violence detection models, machine learning and deep learning communities can play a crucial role in mitigating the negative impacts of GBV on individuals and communities worldwide.

2 Literature Review

In [1], the authors have presented the techniques of violence detection. Violence detection will be carried out in three categories: namely Traditional Machine Learning approach, using Support Vector Machine(SVM), and Deep Learning. A review of the methods for violence detection and datasets like Movies (consisting of 200 clips), Hockey (1000 clips), Media (10,000 clips), etc. In [2], the authors have proposed a model that detects violence in videos captured by video surveillance cameras. The proposed model has a UNET-like network model using MobileNet as an encoder. This is followed by an LSTM network for temporal feature extraction. They made use of three datasets which include RWF200, Movie fights, and Hockey fights dataset. In the article [3], The authors propose a system architecture that combines edge computing and computer vision techniques to detect violent behaviour in real-time, which can provide an early warning system for potential safety hazards The results of the study demonstrate that the proposed system achieves high accuracy and low latency, making it a promising solution for violence detection in IoT-based industrial surveillance networks. The increasing need for video surveillance systems with automatic violence detection capabilities has led the authors [4] to focus on enhancing existing violence detection methods. To this end, they proposed a new feature descriptor called Histogram of Optical Flow Magnitude and Orientation (HOMO), which they have implemented using MATLAB. To evaluate the performance of the proposed method, the authors have used two benchmark datasets. The comparison of HOMO with other descriptors on these datasets shows that HOMO performs satisfactorily. In [5], the authors have proposed a system that uses Convolutional Neural Network(CNN) to recognize physical violence actions.

The model detects various bullying actions like kicking and punching. In [6], the authors have proposed a model that uses convolutional 3D networks for feature extraction and classification. The paper focuses on campus violence detection. The authors gathered the data for the creation of campus data by performing certain violent actions and daily - activity videos that would help in the classification of violence. The network consists of convolutional and max pooling layers. The hidden layer makes use of the ReLU activation function to get values as either 0 or 1. They achieved an accuracy of 92.00% for their model. In [7], the authors have created a dataset consisting of violent and non-violent videos. They made use of the CNN model to classify content as violent or non-violent. Extracted features were then fed into the LSTM network. In [8], the authors made use of the hybrid model consisting of AlexNet and SqueezeNet networks. The Convolution Long Short Term Memory (ConvLSTM) is used to extract precise features from a video, which is then classified using

the softmax classifier. In [9], the authors made a comparison of different video classification approaches and techniques based on the features. They presented different performance metrics like accuracy, and f1 score for video classification and discussed its applications as well. The paper [10] presents a solution for the detection of fights, aggressive motions, and violent scenes in live video streams using a 3D Convolutional Neural Network.

The proposed method has demonstrated significantly improved performance compared to existing techniques, as evidenced by its promising results on three challenging benchmark datasets: Hockey Fight, Crowd Violence, and Movie Violence. In [11], the authors have proposed a combination of CNN and LSTM. They proposed two different models for text classification which are NA-CNN-COIF-LSTM and NA-CNN-LSTM. The combination of CNN by not using activation function with Long Short Term Memory (LSTM) has better performance. In [12], the authors have used the toxic comments dataset and used and compared two different models namely Glove +CNN, Glove +CNN +LSTM, based on testing and training loss and accuracy, which concludes that the first combination gives the best performance required and the second combination doesn't perform that well. In [13], the authors have proposed a system to detect gender-based violations on Twitter messages generated in Mexico. They downloaded 1,857,450 Twitter messages for the creation of the dataset and were manually labeled as positive, negative, or neutral messages. The authors performed minimal pre-processing on the dataset and thus the initial messages were converted to a numeric-format vector. They also studied different feature extraction methods like CountVectorizer, TfidfVectorizer, and Hashing Vectorizer. In [14], the authors have proposed a women's abuse detection method using CNN where they detect the male and the female present in the location. In [15], the author surveys the recent advances in the use of deep learning techniques for detecting violence in various settings. The authors discuss the different types of violence and the challenges associated with detecting them.

They provide an overview of different deep learning models, including CNNs, RNNs, LSTM networks, Capsule Networks, and GANs. The authors also review different datasets used for training and evaluating violence detection models and discuss the performance of various models on different modalities, including video, audio, and text. In [16], the authors have presented various techniques to detect violence. They have categorized these techniques as using Machine Learning, Deep Learning, and Support Vector Machine. In [17], the authors proposed a VGG19 Convolutional Neural Network, where they extract the frames from the input videos and label the objects in the frame that show abnormal behaviour. The system is used for the detection of crimes. In [18], the authors have proposed a weakly supervised method to detect spatial and temporal actions that are violent in the videos. They have used the fast - RCNN architecture that extracts the spatiotemporal information. A summary of the papers is as shown in table 1.

Table 1. Literature survey of all the referred papers

Reference (Year)		Outcome
[1]	2019	The paper reviews various techniques for detecting physical, verbal, and visual violence, including audio-based, video-based, and multimodal approaches, and discusses the challenges and limitations associated with each method. The authors emphasize the need for further research in this field and suggest that deep learning and other advanced techniques could lead to more accurate and reliable violence detection methods in the future.

[2]	2022	Proposed a novel approach for detecting violent events in surveillance videos. The authors develop a two-stage system that first extracts features using a convolutional neural network (CNN) and then applies a decision tree algorithm to classify the features as either violent or non-violent. The proposed system achieves high accuracy in detecting violent events while minimizing false positives, and it outperforms several other state-of-the-art violence detection methods in terms of speed and efficiency.
[3]	2021	This paper presents a novel approach to violence detection in industrial surveillance networks using AI-assisted edge vision, which has significant potential for improving safety in industrial environments. The paper provides a valuable contribution to the field of industrial informatics by demonstrating the potential of AI-assisted edge vision for improving safety and security in industrial settings.
[4]	2019	Proposed a new approach for detecting violent events in videos using optical flow. The authors develop a system that first computes the optical flow vectors between consecutive video frames and then extracts features from these vectors using a convolutional neural network (CNN). The proposed system applies a support vector machine (SVM) to classify the extracted features as violent or non-violent.
[5]	2022	Proposed a system for detecting physical violence among students using surveillance cameras and convolutional neural networks (CNNs). The proposed system achieves high accuracy in detecting physical violence among students while minimizing false positives, and it outperforms several other state-of-the-art violence detection methods in terms of accuracy and computational efficiency
[6]	2021	Proposed a system for detecting violence on a university campus using surveillance cameras and artificial intelligence (AI) techniques. The proposed system uses a deep neural network (DNN) to extract features from video frames and then applies a support vector machine (SVM) to classify the features as violent or non-violent.
[7]	2019	Proposed a system for detecting violence in videos using pretrained deep learning models. The authors evaluate the performance of several state-of-the-art deep-learning models, including ResNet, Inception, and VGG, in detecting violence in videos. They also investigate the impact of fine-tuning these models on violence detection accuracy.
[8]	2022	Proposed a system for detecting violence in videos using a fusion technique that combines deep features from multiple convolutional neural networks (CNNs). The proposed system uses three CNN models to extract deep features from video frames, which are then combined using a fusion technique to classify the frames as violent or non-violent.
[9]	2020	Provided a comprehensive review of video classification methods and techniques, along with their findings, performance, challenges, limitations, and future research directions. The authors discuss various deep learning-based methods for video classification, including 2D CNNs, 3D CNNs, and recurrent neural networks (RNNs), and highlight their strengths and weaknesses

[10]	2020	The paper proposes a novel method for violence detection in videos by combining 3D convolutional neural networks (CNN) and support vector machines (SVM). The 3D CNN extracts spatio-temporal features from the video frames, while the SVM is used for classification. The proposed approach outperforms existing methods on the challenging UT-Interaction dataset, achieving an accuracy of 90.65%.
[11]	2019	Proposed a hybrid model that combines Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to classify text documents. The CNN component of the model is used to extract features from the text, while the LSTM component is used to capture the sequential nature of the text data.
[12]	2018	Proposed a hybrid deep learning model called LSTM-CNN for text classification, which combines the strengths of Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN). The LSTM is used to capture the long-term dependencies in the input text, while the CNN extracts local features from the text. The outputs of these two networks are concatenated and fed into a fully connected layer for classification.
[13]	2021	Proposed a deep neural network (DNN) model to detect gender-based violence (GBV) in Twitter messages. The authors use a dataset of tweets containing GBV-related keywords and apply natural language processing techniques to preprocess the text data. They then train a DNN model using a combination of convolutional and recurrent neural networks to classify tweets as either GBV or non-GBV.
[14]	2020	Presented a deep learning-based approach for detecting women abuse in video surveillance. The proposed method uses a pre-trained convolutional neural network (CNN) to extract features from the video frames, which are then fed into a long short-term memory (LSTM) network for temporal modeling. The model is trained and evaluated on a dataset of videos depicting different types of women abuse, and achieves an accuracy of 92.4%.
[15]	2019	Provided a comprehensive review of the recent advances in the use of deep learning techniques for detecting violence in various settings. The authors discuss the different types of violence and the challenges associated with detecting them. They provide an overview of different deep learning models, including CNNs, RNNs, LSTM networks, Capsule Networks, and GANs. The authors also review different datasets used for training and evaluating violence detection models and discuss the performance of various models on different modalities, including video, audio, and text.
[16]	2022	Presented a comprehensive review of the state-of-the-art techniques for violence detection. The authors survey various methods for detecting violence in different contexts, including surveillance videos, social media, and online gaming. The review covers both traditional machine learning-based approaches and more recent deep learning-based methods. The review serves as a valuable resource for researchers and practitioners working in the field of violence detection and related areas.
[17]	2020	Proposed a method for detecting video surveillance cameras in a given scene using

		VGG19 convolutional neural networks. The authors use a pre-trained VGG19 model and fine-tune it on a small dataset of surveillance images to recognize the features of cameras, such as their shape and color. The proposed method achieved high accuracy in detecting cameras in a variety of scenes, including indoor and outdoor environments, and the authors suggest that it can be used for automated surveillance system deployment and monitoring.
[18]	2022	Proposed a weakly supervised approach for violence detection in surveillance videos. The proposed method uses a pre-trained convolutional neural network (CNN) to extract features from the video frames, and then employs a novel pooling strategy to generate a violence score for each frame. The pooling strategy is designed to capture the spatial and temporal characteristics of violence in a weakly supervised setting, where only video-level labels are available.

3 Methodology

Initially, the text based classification process and the proposed algorithm are explained in detail. The dataset which is prepared, contains 850 records while each record consists of text and its corresponding violence type (label). For videos, there are around 280 videos, approximately 80-100 videos for each class.

3.1 Data Preprocessing

a) The text preprocessing techniques such as stemming, lemmatization and Tokenization (as seen in fig. 1) are applied on the sentences and also used by LabelEncoder to encode the class labels.
b) Mapped the vocabulary to the integer value by making use of the StringLookup functionality of keras, which will not perform any splitting or transformation on the input string.
Later a InceptionV3 feature extractor with weights set to imagenet was being applied to the data.

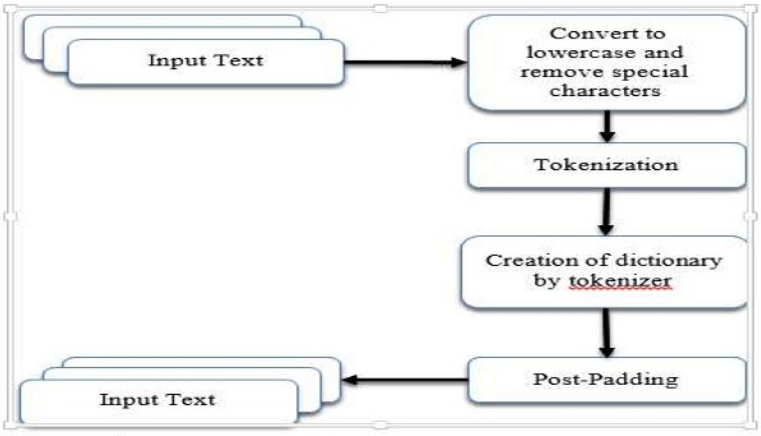

Fig. 1. Stages of pre-processing

3.2 Splitting of the Dataset

For text data, the split ratio is 80:20 so that 80% i.e., 680 rows are used for training set and 20% i.e 170 rows for the testing set. Shuffled the entire dataset wherein allotted 200 videos for training and 100 videos for testing purpose.

3.3 Model building and Compilation

Algorithm 1: Convolutional Recurrent Neural Network for text classification

1. Let X be the input sequence of length L and Y be the target output with C classes.
2. Preprocess the input text data by mapping it to a sequence of word embeddings: X = {x_1, x_2, ..., x_L} where x_i ∈ R^d.
3. Initialize a 1D convolutional layer with k filters, each with size h × d, where h is the height of the filter and d is the dimension of the word embeddings. Denote the convolutional layer output as H, where
4. H ∈ R^{L-k+1}×n.
5. Apply a max-pooling operation to the output of the convolutional layer: P = maxpool(H), where P ∈ R^{1×n}.
6. Feed the output of the max-pooling layer into a recurrent layer with m hidden units, such as a GRU or LSTM layer: R = recurrent(P), where R ∈ R^{1×m}.
7. Perform temporal max-pooling over the recurrent layer output to obtain a fixed-length representation of the input sequence: T = maxpool(R), where T ∈ R^{1×m}.
8. Connect the output of the temporal max-pooling layer to a fully connected layer with softmax activation to output the probability distribution over the C classes: Y_pred = softmax(WT+b), where Y_pred ∈ R^{1×C}, W ∈ R^{m×C}, and b ∈ R^C.

For textual data (fig. 2), the various models such as LSTMs, Bidirectional LSTMs and CNN model are used generally, but the proposed algorithm makes use of CNN+Bidirectional-LSTMs combined architecture which gives more desirable and accurate results. The model comparison is shown in table no. 2. The CNN+ Bidirectional LSTMs is constructed in the following manner: a series of convolutional 2d layers and max pooling 2d layers is created and then a concatenate layer for combining all the MaxPooling 2d layer outputs. Now the text features which are extracted by CNN architecture are given to further Bi-LSTMs layers and finally an output Dense layer with neurons equal to the number of classes. Later the model is compiled using Adam optimizer and loss of sparse categorical cross entropy. The mathematical equations for the cell state, candidate cell state, and final output are given in the equation 1, 2 and equation 3. The flow
of the proposed system on textual data is shown in fig. 3.

$$c - t = \tanh(wc[ht - 1, xt] + bc) \quad (1)$$

$$ct = ft * ct - 1 + it * c{\sim}t \quad (2)$$

$$ht = ot * \tanh(ct) \quad (3)$$

where ct is cell state(memory) at timestamp(t),
$c{\sim}t$ represents candidate for cell state at timestamp(t)

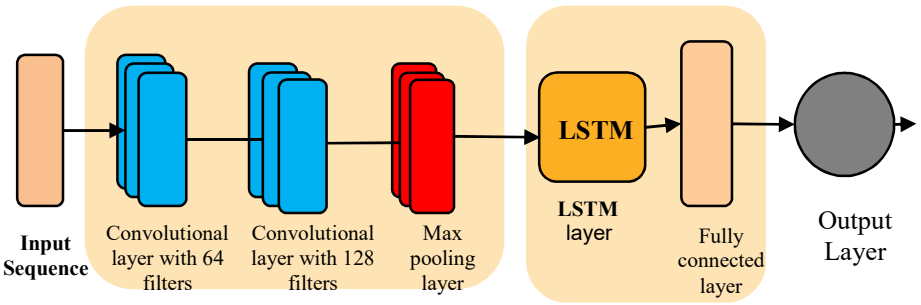

Fig. 2. Proposed model hybrid model of CNN and Bidirectional LSTM architecture for text classification

Algorithm 2: Convolutional Recurrent Neural Network for video classification

. **Input:** A video sequence of length T, with each frame having dimension H x W x C

2. Convolutional Layers:

a. Apply a set of convolutional filters with size k x k x d to each frame, resulting in a feature map with size H x W x F1

b. Apply max pooling with size p x p to the feature map, resulting in a downsampled feature map with size (H/p) x (W/p) x F1

c. Apply a set of convolutional filters with size k x k x d to the downsampled feature map, resulting in a feature map with size (H/p) x (W/p) x F2

d. Apply max pooling with size p x p to the feature map, resulting in a downsampled feature map with size (H/p^2) x (W/p^2) x F2

3. Recurrent Layers:

a. Reshape the downsampled feature maps into a sequence of vectors, each with dimension F2

b. Feed the sequence of vectors into a set of recurrent layers (e.g. LSTM or GRU) with hidden dimension h and output sequence length T/h

4. Fully Connected Layers:

a. Apply a fully connected layer to each output of the recurrent layers, resulting in a sequence of feature vectors with dimension F3

b. Apply a final fully connected layer with softmax activation to classify the video into one of C classes.

For videos, there is an extraction of vocabulary for every input sentence by using the label processor. Built a CRNN model wherein passed the input vector initially to the GRU, Dense

and Dropout layer. Wherein Dropout helped to reduce the overfitting of the model. Later compiled the model with the loss of sparse categorical crossentropy which OneHotEncoded the vectors made use of Adam optimizer. Applied the model on the training data and evaluated the model's F1 score and accuracy on the testing data.

4 Flowchart

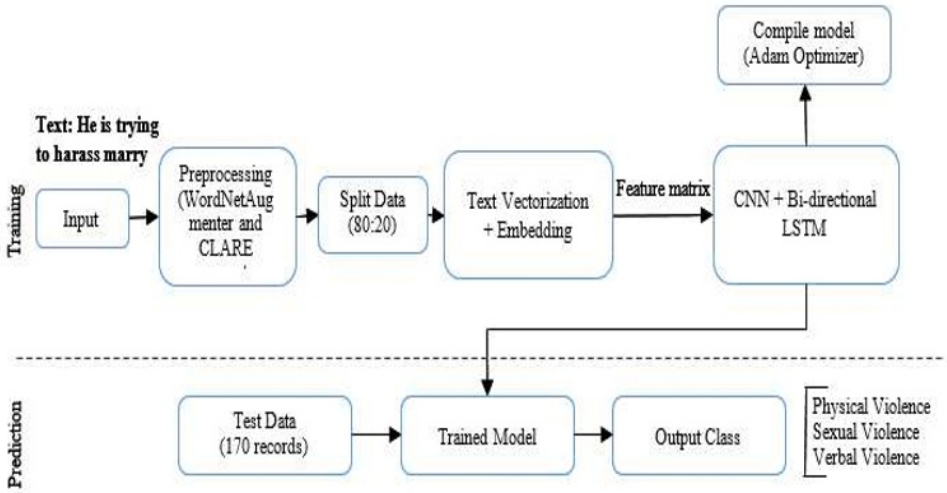

Fig. 3. Proposed project flow diagram for Text Classification

5 Experimentation

First for the text based classification, the base models such as LSTM model is used which gives an F1 score of 0.80 and CNN based text classification with a F1 score of 0.82. Then the combination of CNN + Bidirectional LSTM combined model is used which gives a F1 score of 0.90 (fig. 5). The model uses Adam optimizer for model compilation and loss function of sparse categorical cross entropy.

Table 2: Model comparison for Text Classification models

Model	F1 Score
LSTM model	0.80
Bi-LSTM model	0.86
CNN model	0.82
CNN + Bi-LSTM model	0.90

The video classification has been tackled using a Convolutional Recurrent Neural Network (CRNN) architecture. The CRNN model comprises bi-directional Long Short-Term Memory (LSTM) layers which have effectively improved the accuracy of the predictions.

The model has also been equipped with dropout and regularizer layers to avoid overfitting. The model has achieved an impressive accuracy of 96%, with a minimal loss of 0.8.

6 Results and Discussion

The results verify that the proposed approach reaches higher accuracy with precise output. For the text classification, the hybrid model of CNN and Bidirectional LSTM(fig. 4) is used. Here, the features from both the directions are combined and considered for further analysis. The model trains the input text data twice with the help of forward and backward directions. The accuracy on the testing data for text as the input is 97%. In order to extract high-level information from videos, CNN models are fed the video's pictures. The RNN layer's output is connected to a fully connected layer to produce the classification output after the features have been provided to it. The Convolution Recurrent Neural Network(CRNN) performs better for motion based activities. It extracts the correlation between the images by keeping in mind the past frames and their features. The accuracy for the training dataset was 97% whereas on the testing dataset it was found to be 96%. The violence which has a higher percentage amongst the others is predicted based on the input text features.

```
In [27]: sent=["He is gonna kill and beat them tommorow morning at 10AM!"]
         print(prediction(sent))

         1/1 [==============================] - 0s 57ms/step
         ['Physical_violence']
```

Fig. 4. Output image of CNN+Bi-LTSM text classification

Fig. 5. Output image of CRNN video classification

CRNN provides better results when compared to transfer learning. The major reason being the number of layers and classes the latter has. The model is trained on IMAGENET which has 1000 classes and layers more than 500. This becomes computationally expensive and increases the training time.

7 Conclusion

The paper highlights the detection of violence in text, images, and videos with the help of various deep learning algorithms. A hybrid model of CNN and Bidirectional LSTM (combined architecture) is used so that Bi-LSTM can utilize the information from both sides for better understanding. Recurrent neural networks are used for text classification which

initially starts with pre-processing like removing of punctuations followed by feature extraction. These features help the model identify and understand the violence. The proposed system provides good accuracy with no overfitting. The created system is beneficial to be used in surveillance systems and social media applications which are prone to violence and harassment.

The proposed system involves detection of violence in text and videos. The dataset which is being created consists of various videos depicting the different types of violence like physical, sexual and emotional. Though the results that are drawn from the proposed approach are quite precise and beneficial to be used in surveillance systems, there is always a need for improvement. A dataset consisting of audios which depict any kind of violence based on the phonic information. These audios will help in detecting violence or harassment for example: if someone is trying to emotionally blackmail a person or abuse him, it will automatically be detected, and appropriate action will then be taken. Thus, addition of audios to the dataset will make an appropriate system that can be further used in various applications to avoid various kinds of violences before any major mishap.

References

1. M. Ramzan et al., "A Review on State-of-the-Art Violence Detection Techniques," in IEEE Access, vol. 7, pp. 107560-107575, 2019, doi: 10.1109/ACCESS.2019.2932114.
2. Vijeikis, Romas, Vidas Raudonis, and Gintaras Dervinis. 2022. "Efficient Violence Detection in Surveillance" Sensors 22, no. 6: 2216
3. F. U. M. Ullah et al., "AI-Assisted Edge Vision for Violence Detection in IoT-Based Industrial Surveillance Networks," in IEEE Transactions on Industrial Informatics, vol. 18, no. 8, pp. 5359-5370, Aug. 2022, doi: 10.1109/TII.2021.3116377.
4. Javad Mahmoodi, Afsane Salajeghe, A classification method based on optical flow for violence detection, Expert Systems with Applications, Volume 127, 2019, Pages 121-127, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2019.02.032.
5. John Clement Suladay Escobanez and Benilda Eleonor Comendador. 2022. Student Physical Violence Detection using Convolutional Neural Networks. In Proceedings of the 12th International Conference on Information Communication and Management (ICICM '22). Association for Computing Machinery, New York, NY, USA, 34–38.
6. Ye, Liang, Tong Liu, Tian Han, Hany Ferdinando, Tapio Seppänen, and Esko Alasaarela. 2021. "Campus Violence Detection Based on Artificial Intelligent Interpretation of Surveillance Video Sequences" Remote Sensing 13, no. 4: 628.
7. Sumon, Shakil & Goni, Raihan & Hashem, Niyaz & Shahria, Md Tanzil & Rahman, Mohammad. (2019). Violence Detection by Pretrained Modules with Different Deep Learning Approaches. Vietnam Journal of Computer Science. 7. 10.1142/S2196888820500013.
8. Mohammed, Heyam & Elrefaei, Lamiaa. (2022). Detecting Violence in Video Based on Deep Features Fusion Technique.
9. Islam, Md & Sultana, Shanjida & Roy, Uttam & Al, Jubayer. (2020). A review on Video Classification with Methods, Findings, Performance, Challenges, Limitations and Future Work. Jurnal Ilmiah Teknik Elektro Komputer dan Informatika. Vol 6, No 2 (2020). 47-57. 10.26555/jiteki.v6i2.18978.
10. Simone Accattoli, Paolo Sernani, Nicola Falcionelli, Dagmawi Neway Mekuria & Aldo Franco Dragoni (2020) Violence Detection in Videos by Combining 3D Convolutional Neural Networks and Support Vector Machines, Applied Artificial Intelligence, 34:4, 329-344, DOI: 10.1080/08839514.2020.1723876.

11. Y. Luan and S. Lin, "Research on Text Classification Based on CNN and LSTM," 2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 2019, pp. 352-355.
12. Mukul Anand , Dr.R.Eswari . “Classification of abusive comments in social media using deep learning” ,(ICCMC 2019) IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4.
13. Castorena CM, Abundez IM, Alejo R, Granda-Gutiérrez EE, Rendón E, Villegas O. Deep Neural Network for Gender-Based Violence Detection on Twitter Messages. Mathematics. 2021; 9(8):807
14. Sandhiya, R., & Prassad, A.R. (2020). Women Abuse Detection in Video Surveillance using Deep Learning.
15. Dandage, V., Gautam, H., Ghavale, A., Mahore, R., & Sonewar, P.A. (2019). Review of Violence Detection System using Deep Learning.
16. Milon Biswas, Afjal Hossain Jibon, Mim Kabir, Khandokar Mohima, Rahman Sinthy, Md. Shamsul Islam a and Monowara Siddique. State-of-the-Art Violence Detection Techniques: A review, Asian Journal of Research in Computer Science 13(1): 29-42, 2022; Article no.AJRCOS.79063 ISSN: 2581-8260
17. Umair Muneer Butt, Sukumar Letchmunan, Fadratul Hafinaz Hassan, Sultan Zia and Anees Baqir, “Detecting Video Surveillance Using VGG19 Convolutional Neural Networks” International Journal of Advanced Computer Science and Applications(IJACSA), 11(2), 2020
18. Choqueluque-Roman D, Camara-Chavez G. Weakly Supervised Violence Detection in Surveillance Video. Sensors. 2022; 22(12).