***https://doi.org/ 10.33472/AFJBS.6.5.2024. 2464-2473***

African Journal of Biological Sciences

# Human visual violence pattern recognition in video using novel and improved deep learning framework

[1*]Mahaveer Jain, [2]Mukesh Kumar

[1,2]Department of Computer Science and Engineering, Rabindranath Tagore University Raisen Madhya Pradesh India.

*[1*]profmahavir@gmail.com, [2] goutam.mukesh@gmail.com*

Article History
Volume 6, Issue 5, Apr 2024
Received: 25 Apr 2024
Accepted: 06 may 2024
doi: *10.33472/AFJBS.6.5.2024. 2464-2473*

Abstract

One of the major problems that we are facing is the increase in violent activities in society. There are many reasons for increased violent activities that need the immediate attention of society, but in this research article, we are discussing how we can detect violent behavior in surveillance videos. The surveillance cameras address the major concerns of our daily security needs and can record violent activities too. Today we are surrounded by surveillance cameras everywhere, such as in malls, streets, houses, and other public places. Surveillance cameras capture every movement of an object that comes under their range and produce huge visual data. This visual data can be analyzed to detect violent behavior manually, but it is almost impossible in this era where we are surrounded by so many cameras. This problem needs some technological intervention to detect the violence in surveillance videos as soon as it happens. Deep learning and computer vision have the potential to contribute a lot to this problem domain, specifically with pretrained models like vgg16 and vgg19. These pretrained models are already trained on a large dataset and capable of figuring out edges and shapes in scenes. We have used the improved visual geometry group 19 pretrained model to detect violence in video. The vgg19 consists of fixed sequence of convolutional layers, max pool layers and fully connected layers. Output is classified into 1000 different classes, but to fulfill the objective of this study, we have replaced the fully connected layers with our classification module to classify a video into two categories, namely violent and non-violent. The efficiency of the model is tested against three popular datasets, namely movie fights, hockey fights, and real-life violence situations. The accuracy of the proposed model is 96.25%, 98.75%, and 99.43% for the movie fight, hockey fight, and real-life violence situation datasets, respectively. The proposed model code is available on GitHub (https://github.com/profmahaveer/VGG19.git).

Keywords: Surveillance, Deep learning, Anomaly detection, Violence, VGG19

## Introduction

Humans are such an animal that has unpredictable behavior and may react differently in the same situation. There are two types of human behavior: normal behavior and abnormal behavior. Normal behavior comprises deeds that are according to social norms and do not affect the freedom of another person. Abnormal behavior altogether is against social norms, and especially violent behavior is the most abnormal behavior. Human violence is both innate and learned, influenced by biological and social factors. It also mentions that violence has evolved and taken different forms due to socio-cultural development [1] [2]. Thanks to surveillance cameras, there is a sense of security in every aspect of human life. Studies have focused on various aspects of surveillance cameras, including their spatial distribution, impact on privacy and public safety, and their role in crime prevention [3]. The expanding number of cameras has resulted in the generation of enormous amounts of visual data, causing issues in terms of processing, storage, and retrieval. [4]. Manual examination of security camera footage needs significant human work and time, which may not be possible when looking for aggressive behaviour. [5]. An automated system is required to quickly and accurately locate violent behavior in surveillance footage [6]. Violent action recognition in surveillance footage is the problem of computer vision, and deep learning has transformed computer vision by enabling significant advancements in various applications [7]. Deep learning techniques like convolutional neural networkshave expanded the boundaries of computer vision and made things possible that were supposed to be impossible in the past. Pretrained models such as VGG16, VGG19, Googlenet, and Alexnet are convolutional neural networks that are able to classify an image into many different classes because they are trained on large datasets [8] [9]. These pretrained models have the capability of identifying shapes and patterns in images. The proposed model uses the Visual Geometry Group 19 (VGG19) pretrained model for feature extraction, and we have modified this model by removing all flattened layers and replacing them with our new classification module. The performance of the proposed model has evaluted against three popular datasets.

## 1. Related Work

Violence activity detection from video can be achieved using pretrained models. Many researchers have explored deep learning techniques for this purpose. Models such as VGG-19, VGG-16, InceptionV3, and MobileNetV3 have been used as base models for feature extraction and classification [10]. Transfer learning techniques have also been applied, where the final layers of pretrained models like VGG16 are replaced to accommodate the specific requirements of violence detection [11]. Additionally, the use of the MobileNetV2 architecture with the CNN algorithm and OpenCV has been proposed for violence recognition [12]. Additionally, smart networks that use 3D convolutions to describe the dynamic interactions between humans and objects have been studied for the identification of violence in surveillance footage [13]. Key frame extraction techniques have been used to remove duplicate frames, and the pretrained VGG16 architecture has been employed for violence detection and recognition [14]. Deep learning and natural language processing (NLP) algorithms can be used to identify offensive, hostile, and hate speech in the audio channels of surveillance cameras, as well as unusual objects and situations in videos [15]. A neural network (CNN) can be trained with certain features obtained from appearance, motion speed, and picture representation as input to identify contextual patterns of violence in every video frame [16]. By combining different features and streams, the proposed approach can accurately detect violent acts in CCTV video streams [17]. The use of pretrained models allows for accurate results while using more modest computational resources, making violence detection more accessible for common surveillance systems.

## 2. The Proposed Work

The proposed model consists of 3 phases namely Datasetacquisition and preprocessing, feature extraction and classification. Acquisition of dataset comprises video to frame conversion, Frame resizing and frame normalization. We have used VGG19 pertained model for feature extraction and Classification is done by replacing the top layer of VGG19 model to classify a frame into 2 categories 0 (Non violence) and 1 (Violence) [18].Figure Idepict the proposed model's block diagram.

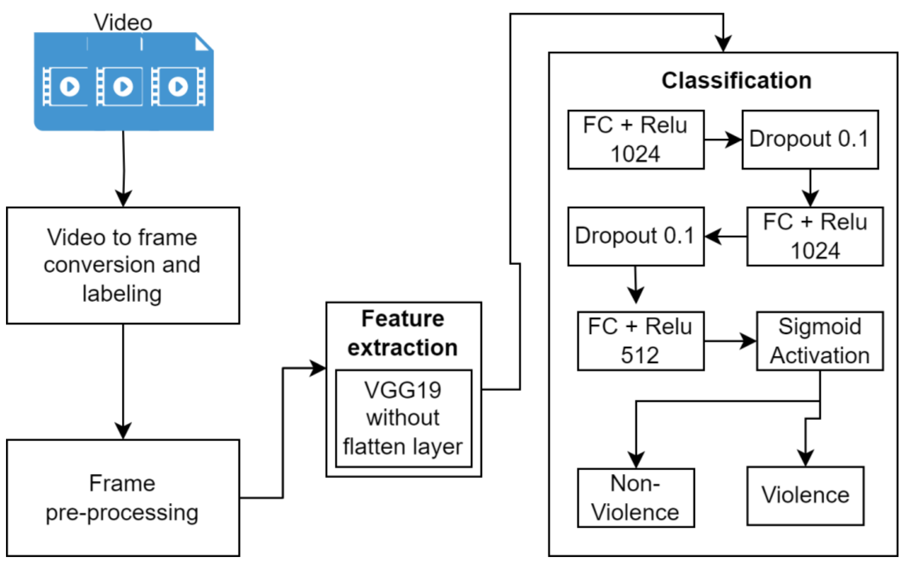



**Fig. I Block diagram of the proposed model**

### 2.1 Dataset acquisition and preprocessing

We acquired three popular datasets, namely the hockey fight, movie fight, and real-life violence situation datasets. The dataset needs to be labeled and annotated to accurately classify the instances of violence. This involves manually reviewing each data instance and categorizing it as either violent or non-violent. Moreover, it is also crucial to balance the dataset with an equal representation of both violent and non-violent samples to avoid biased prediction results. Once the dataset is prepared, it can be further preprocessed, including techniques like normalization and resizing of images, before being used to train a CNN model for violence detection.

#### 3.1.1 Video to Frames conversion

Video-to-frame conversion refers to the process of extracting individual frames from a video file. It involves breaking down a video into a series of still images, commonly known as frames, which can be used for various purposes. This conversion allows for more detailed analysis and evaluation of each frame, enabling researchers, filmmakers, and image processing algorithms to scrutinize specific moments within a video. Video-to-frame conversion plays a crucial role in applications such as video editing, motion tracking, object recognition, and computer vision. By converting videos to frames, we can unlock a wealth of information and gain a deeper understanding of the visual content within the video. We converted these videos into frames by selecting 1 frame per second. The video-to-frame conversion process is depicted in Figure II.

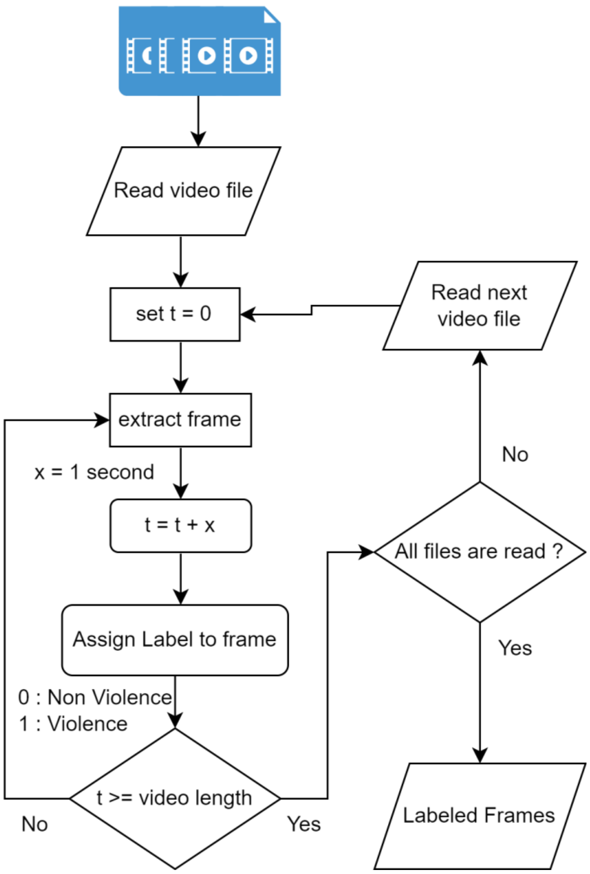



**Fig. II. A typical process of video to frame conversion**

#### 3.1.2 Frame Resizing

The VGG19 model has a specific and fixed size input requirement, so to fulfill this requirement, we need to resize frames. By resizing the images, we ensure that their dimensions remain consistent and compatible with the model input requirements. This procedure involves utilizing algorithms like bilinear or bicubic interpolation to adjust the size of the images while preserving their aspect ratios. Resizing images also reducescomputational complexity by decreasing the input size. Overall, image resizing plays a crucial role in preparing the dataset for violence detection, enabling vgg19 to effectively learn and recognize patterns associated with violent content.

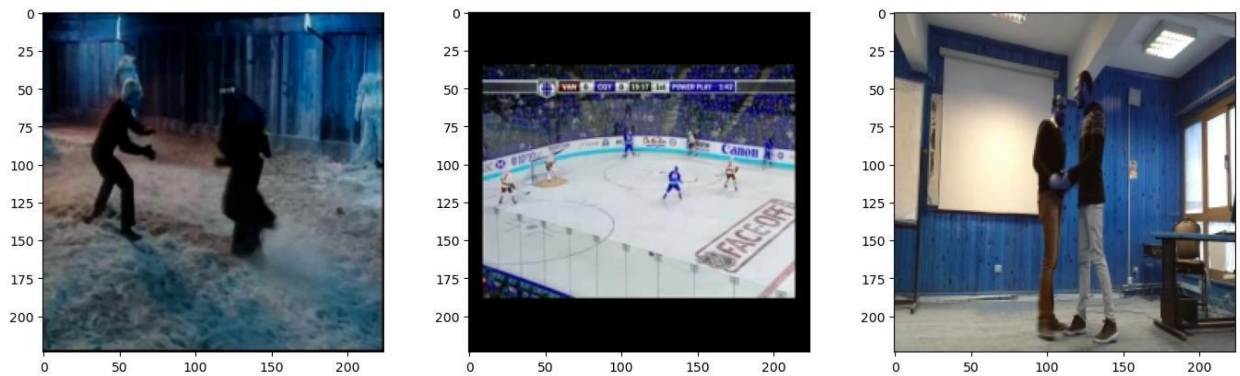
$$\boldsymbol{Resized\ Frame} = \mathbf{224\ x\ 224}$$

### 3.1.3 Frame Normalization

Normalization is the process of transforming the features of the data to a common scale or range. It is typically performed to ensure that different features of the data have equal importance and do not dominate the analysis based on their magnitude. Normalization is essential in many machine learning and data analysis tasks as it helps in achieving better results and avoids bias towards specific features. It allows different frames or data points to be compared and analyzed effectively. The most commonly used technique for frame normalization is feature scaling, where each feature is rescaled to fall within a specific range. This can be done using various methods such as min-max scaling, standardization, or normalization by statistical measures like mean and standard deviation. Min-max scaling reduces each feature to a range of 0 to 1, while keeping the proportional relationship between the values. Normalization is particularly important when dealing with features that may have different units, scales, or distributions. By bringing them to a common scale, it enhances the performance and interpretability of various analytical and machine learning algorithms. A frame pixel may vary from 0 to 255, so to normalize all the pixels in a range, we have divided the pixels by 255. It can be represented as follows:

$$\boldsymbol{Normalized\,Frame = \; Resized\,Frame/\,255}$$

The sample images of datasets are shown in Figure III.



(a) (b) (c)

**Fig. III. Dataset sample images (a) Movie fight (b) Hockey fight (c) Real life violence situation**

### 2.2 Feature Extraction

Feature extraction is done by VGG19 pretrained model. VGG19 stands for visual geometry group 19 [19]. VGG19 is a pretrained model trained on a large dataset and capable of identifying different shapes and patterns [20]. The structure of VGG19 follows a fixed sequence of 16 convolutional and 5 max-pool layers [21]. The flatten layers of this model are excluded to fulfill the objective of this study and replaced with a new classification module that classifies a frame into 0 (non-violence) and 1 (violence) labels.

Typically, VGG19 is based on the artificial neural network paradigm and organized into a sequence of convolutional layers, pooling layers, and output layers [22] [23]. The functions of these layers are given in the form of mathematical equations, i.e. how a neural network converts input into output [24].

$$F[i,j] = (I * K)_{[i,j]} \qquad \text{.......I}$$

The convolutional operation between frame ($I$) and kernel ($K$)is given in equation I. Size of the frame $I$is$m_1$ x $m_2$ x $m_c$ , where $m_1$ , $m_2$ and $m_c$ denotes height, width and number of channels (3 for an RGB image and 1 for gray scale image) respectively. $K$ is the kernel filter of size $n_1$ x $n_2$ x $n_c$ that will apply to the image with a stride of 1. Where $n_1$, $n_2$ and $n_c$denote height, width, and number of channels in the kernel respectively. The kernel and frame should have equal number of channels ($m_c$ should be equal to $n_c$) [25] [26].

$$C = Conv\,(I,K) = \; \varphi_a\;(F) \qquad \text{.......II}$$

$\varphi_a$ is the activation function (Relu) and the operation of activation is shown in equation II [27].

$$P = \varphi_p\,(\mathrm{C}) \qquad \text{.......III}$$

$\varphi_p$ is the Pooling operation that will produce the output P as given in equation III [28][29].

$$X = \sum_i W_i P_i + b' \qquad \text{.......IV}$$

Flattened vector $P$ is combined with weight vector $W_i$ to produce the output $X$ as given in equation IV [30] [31].

$$Z = g(X) \qquad \text{.......V}$$

Finally,the output $X$ is converted into class labels 0 (Non violence) or 1 (Violence) with the help of activation function $g$ (Sigmoid) as given in equation V.

### 2.3 Classification

Now we need to classify a frame as non-violent or violent, depending on the extracted features. For this purpose, we have designed a new classification module that consists of 4 fully connected layers and 2 dropout layers. The output layer uses the sigmoid activation function to classify a frame into 0 (non-violence) or 1 (violence) labels [32] [33]. We used different layer settings during training and testing of the model and captured the best setting shown in Figure IV.

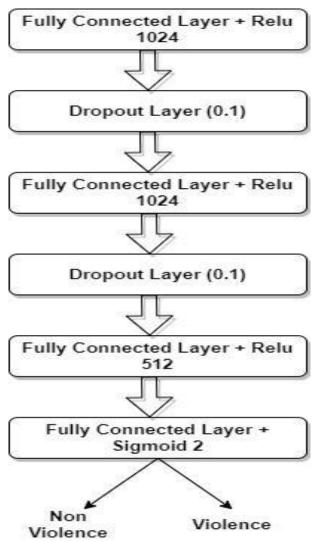



**Fig. IV. Frame classification module**

## 3. Result and Discussion

The proposed model's performance is evaluted against three popular datasets, namely the hockey fight [35], movie fight [35], and real-life violence situation [36]. The datasets contain videos organized by non-violence and violence labels. We have split datasets into train (80%) and test (20%). Training data is again divided into training and validation dataset in the ratio of 80% and 20%. We evaluated accuracy, precision, recall, and specificity metrics [34] to assess the proposed model's performance. The mathematical equations to calculate the above metrics are shown in equations VI, VII, VIII, and IX.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad \text{.......VI}$$

$$Precision = \frac{TP}{TP + FP} \qquad \text{.......VII}$$

$$Recall/Sensitivity = \frac{TP}{TP + FN} \qquad \text{.......VIII}$$

$$Specificity = \frac{TN}{TN + FP} \qquad \text{.......IX}$$

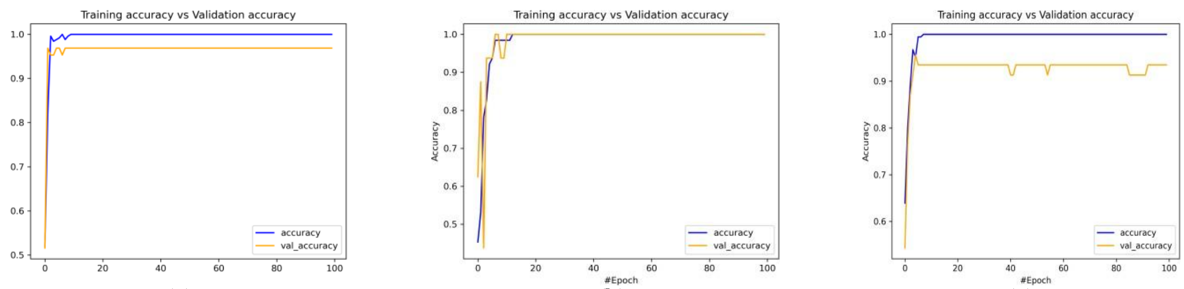
TP, TN, FP, and FN stand for true positive, true negative, false positive, and false negative, respectively.

Hyper parameters used during training are as follows.

**Table. I. Hyper Parameters of Experimental Setup during training and testing**

| # | Feature | Specific Value |
|---|---|---|
| 1 | Learning Rate | 0.0001 |
| 2 | Optimizer | Adam |
| 3 | Hidden Layer Activation Function | Rectified linear Unit |
| 4 | Output Layer Activation Function | Sigmoid |
| 5 | No of Epochs | 100 |

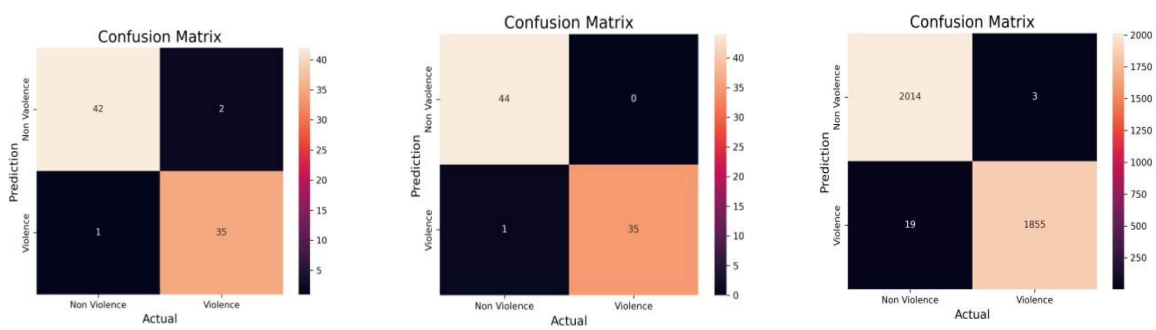
The model is trained on all three datasets for 100 epochs, one by one, and their performance metrics are captured. The training vs. validation accuracy graph is shown in Figure IV.




(a) (b) (c)

**Fig. V. Training accuracy vs validation accuracy graph for (a) Movie fight (b) Hockey fight and (c) RLVS datasets**

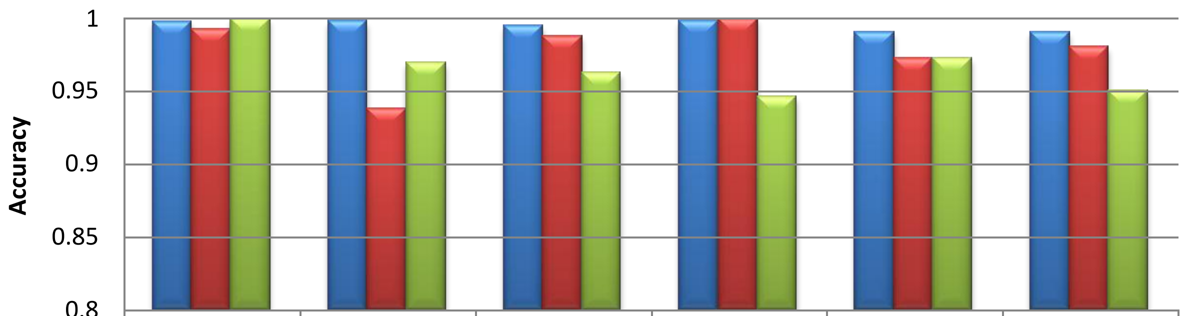
The model is tested on all three datasets, and a confusion matrix is drawn to evaluate its performance. Figure VI shows the confusion matrix.




(a) (b) (c)

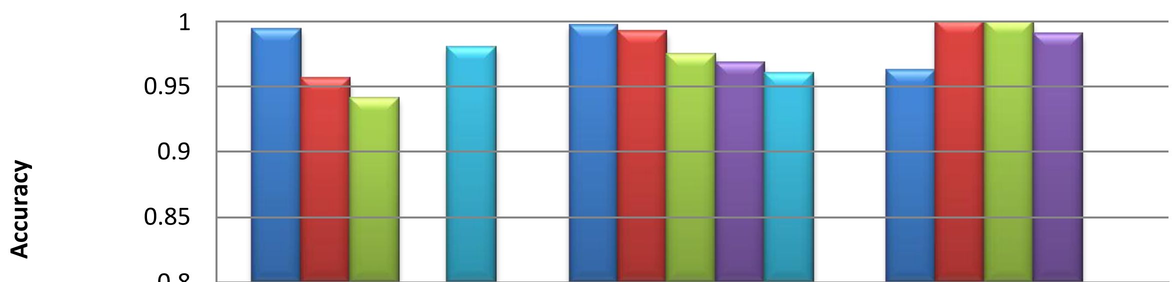
**Fig. IIII. Confusion matrix for (a) Movie fight (b) Hockey fight and (c) RLVS dataset**

The proposed model outperforms, and their performance metrics are shown in Figures VI and VII. The model is compared with existing models to showcase the efficacy of the proposed model. The performance of the proposed model for the RLVS dataset is 99.43%, as shown in figure VII.




| | Training Accuracy | Validation Accuracy | Testing Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|---|---|
| RLVS | 0.9975 | 0.9981 | 0.9943 | 0.9983 | 0.9898 | 0.99 |
| Hockey | 0.9922 | 0.9375 | 0.9875 | 1 | 0.9722 | 0.98 |
| Movie | 1 | 0.9688 | 0.9625 | 0.9459 | 0.9722 | 0.95 |

**Fig. IVI. Training, validation, testing, precision, recall accuracy and F1-score of the proposed model.**




| | RLVS | Hockey | Movie |
|---|---|---|---|
| proposed model | 0.994 | 0.997 | 0.962 |
| OF [42] | 0.956 | 0.992 | 1 |
| Pseudo-OF [42] | 0.941 | 0.975 | 1 |
| Bilinski [41] | | 0.968 | 0.99 |
| Ullah [38] | 0.98 | 0.96 | |

**Fig. VI. Comparison of Proposed model accuracy with existing models**

## 4. Conclusion

The use of pre-trained CNN models to identify violence has shown to be an extremely successful approach. These algorithms, trained on big datasets, can effectively recognize violent activities and distinguish them from nonviolent behavior in a variety of scenarios. Violence detection systems have achieved significant success in real-time video analysis by combining deep learning techniques such as CNNs, allowing for rapid response and prevention of violent situations. By utilizing the capability of these pre-trained models, the computational cost of training a violence detection system is greatly decreased, making it more feasible for widespread applications. Furthermore, the versatility of pre-trained CNN models enables the transfer learning technique, in which the model is fine-tuned or re-trained on a smaller dataset specialized to violence detection.Thepre-trained CNN models have shown their efficiency in this field of violence detection. These models have training on large datasets, are capable of accurately identifying violent actions and distinguishing them from non-violent behavior in various contexts. This method not only improves the model's efficiency, but also allows for customization for certain scenarios and environments. Pre-trained CNN models have showed considerable potential, but there is still opportunity for improvement. More research and development are needed to fine-tune the algorithms, minimizing false positives and improving their ability to detect subtle kinds of violence. Additional efforts should be made to improve the models' ability to detect violence across multiple domains, including distinct cultural contexts and languages. The use of pre-trained CNN models for violence detection has proven to be valuable and helpful in achieving the overall goal of creating safer environments. With further developments and refining, these models have the potential to revolutionize the field of violence detection and make substantial strides towards a more secure and harmonious society.

**Statements and Declarations**

**Acknowledgement**

I extend my sincere thanks to my Guide Dr. Mukesh kumar for providing support and guidance during this study.

**Competing Interest**

The authors of this article are not associated with any organization or entity that has interest in this topic and material presented here.

**Funding Information**

The author did not receive any funds to help with the preparation of this manuscript.

**Author Contribution**

This study develops a new and novel method of violence detection in the field of vision.

**Data and Code Availability**

The datasets described in this article is openly available and we have given references of datasets in the reference section. The code of the developed model is available on the GitHub [44].

**Informed Consent**

I understand the basic goals, potential hazards, and research methodology, and I give my consent to participate in this study.

## 5. References

1. LENŢA, O. E., & CORMOŞ, V. C. (2017). Aggression and Violence, Corrosive Factors of Humanity. Postmodern Openings/DeschideriPostmoderne, 8(1).
2. Anderson, C. A., & Huesmann, L. R. (2007). Human aggression: A social-cognitive view. The Sage handbook of social psychology, 259-287.
3. Cho, K., & Cho, D. (2022). Autonomous driving assistance with dynamic objects using traffic surveillance cameras. Applied Sciences, 12(12), 6247.
4. Sheng, H., Yao, K., & Goel, S. (2021, July). Surveilling surveillance: Estimating the prevalence of surveillance cameras with street view data. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (pp. 221-230).
5. Khan, A., Rehman, S., Waleed, M., Khan, A., Khan, U., Kamal, T., ... & Marwat, S. N. K. (2018). Forensic video analysis: passive tracking system for automated Person of Interest (POI) localization. IEEE Access, 6, 43392-43403.
6. Sitara, K., &Mehtre, B. M. (2019). Automated camera sabotage detection for enhancing video surveillance systems. Multimedia Tools and Applications, 78(5), 5819-5841.
7. Le, N., Rathour, V. S., Yamazaki, K., Luu, K., & Savvides, M. (2022). Deep reinforcement learning in computer vision: a comprehensive survey. Artificial Intelligence Review, 1-87.
8. Kaur, R., Kumar, R., & Gupta, M. (2021, December). Review on Transfer Learning for Convolutional Neural Network. In 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (pp. 922-926). IEEE.
9. Dhaka, V. S., Meena, S. V., Rani, G., Sinwar, D., Ijaz, M. F., & Woźniak, M. (2021). A survey of deep convolutional neural networks applied for prediction of plant leaf diseases. Sensors, 21(14), 4749.
10. Singh, S., & Tyagi, B. (2023). Computational Comparison of CNN Based Methods for Violence Detection.
11. Shubber, M. S. M., & Al-Ta'i, Z. T. M. (2023). Detection of Violence Behavior using Deep Learning Technique. Academic Science Journal, 1(2), 188-206.
12. Rathi, S., Sharma, S., Ojha, S., & Kumar, K. (2023, March). Violence Recognition from Videos Using Deep Learning. In Proceedings of International Conference on Recent Trends in Computing: ICRTC 2022 (pp. 69-77). Singapore: Springer Nature Singapore.
13. Huszar, V. D., Adhikarla, V. K., Négyesi, I., &Krasznay, C. (2023). Toward fast and accurate violence detection for automated video surveillance applications. IEEE Access, 11, 18772-18793.
14. Aarthy, K., & Nithya, A. A. (2022, October). Crowd Violence Detection in Videos Using Deep Learning Architecture. In 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon) (pp. 1-6). IEEE.

15. Kumari, V., Memon, K., Aslam, B., & Chowdhry, B. S. (2023, May). An effective approach for violence detection using deep learning and natural language processing. In 2023 7th International Multi-Topic ICT Conference (IMTIC) (pp. 1-8). IEEE.
16. Appavu, N. (2023, April). Violence Detection Based on Multisource Deep CNN with Handcraft Features. In 2023 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET) (pp. 1-6). IEEE.
17. Vladu, C., Prodan, L., &Iovanovici, A. (2022, November). Resource Constrained, Fast Convergence Training for Violence Detection in Video Streams. In 2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI-MACRo) (pp. 000239-000244). IEEE.
18. Rajab, M. A., Abdullatif, F. A., &Sutikno, T. (2024). Classification of grapevine leaves images using VGG-16 and VGG-19 deep learning nets. TELKOMNIKA (Telecommunication Computing Electronics and Control), 22(2), 445-453.
19. Zakaria, N., & Hassim, Y. M. M. (2024). A Review Study of the Visual Geometry Group Approaches for Image Classification. Journal of Applied Science, Technology and Computing, 1(1), 14-28.
20. Kumari, P., Ranjan, P., & Srivastava, P. (2024). Classification and Assessment of Visual Content of Medicinal Plants Using CNN. In Handbook of Research on Innovative Approaches to Information Technology in Library and Information Science (pp. 126-147). IGI Global.
21. Manimurugan, S., Karthikeyan, P., Aborokbah, M., Narmatha, C., & Ganesan, S. (2024). Breast cancer diagnosis model using stacked autoencoder with particle swarm optimization. Ain Shams Engineering Journal, 102734.
22. Butt, U. M., Letchmunan, S., Hassan, F. H., Zia, S., & Baqir, A. (2020). Detecting video surveillance using VGG19 convolutional neural networks. International Journal of Advanced Computer Science and Applications, 11(2).
23. Manimurugan, S., Karthikeyan, P., Aborokbah, M., Narmatha, C., & Ganesan, S. (2024). Breast cancer diagnosis model using stacked autoencoder with particle swarm optimization. Ain Shams Engineering Journal, 102734.
24. Krichen, M. (2023). Convolutional neural networks: A survey. Computers, 12(8), 151.
25. Cao, J., Li, Y., Sun, M., Chen, Y., Lischinski, D., Cohen-Or, D., ... & Tu, C. (2022). Do-conv: DepthwiseW. Yang, J. Zhang, J. Caiand Z. Xu, “HybridNet: Integrating GCN and CNN for skeleton-based action recognition”, Applied Intelligence, 2023.
26. Singh, B., Patel, S., Vijayvargiya, A., & Kumar, R. (2023). Analyzing the impact of activation functions on the performance of the data-driven gait model. Results in Engineering, 18, 101029.
27. Tao, Z., XiaoYu, C., HuiLing, L., XinYu, Y., YunCan, L., & XiaoMin, Z. (2022). Pooling operations in deep learning: from “invariable” to “variable”. BioMed Research International, 2022.
28. Zafar, A., Aamir, M., Mohd Nawi, N., Arshad, A., Riaz, S., Alruban, A., ... &Almotairi, S. (2022). A comparison of pooling methods for convolutional neural networks. Applied Sciences, 12(17), 8643.
29. Shafi, I., Din, S., Khan, A., Díez, I. D. L. T., Casanova, R. D. J. P., Pifarre, K. T., & Ashraf, I. (2022). An effective method for lung cancer diagnosis from ct scan using deep learning-based support vector network. Cancers, 14(21), 5457.
30. Meng, X., Shi, N., Li, G., Zhang, G., Li, W., Zhu, N., & Li, M. (2022). On-demand reconfigurable incoherent optical matrix operator for real-time video image display. Journal of Lightwave Technology, 41(6), 1637-1648.
31. Sangeetha, S. K. B., Kumar, M. S., Deeba, K., Rajadurai, H., Maheshwari, V., & Dalu, G. T. (2022). An empirical analysis of an optimized pretrained deep learning model for COVID-19 diagnosis. Computational and Mathematical Methods in Medicine, 2022.
32. Karacı, A. (2022). VGGCOV19-NET: automatic detection of COVID-19 cases from X-ray images using modified VGG19 CNN architecture and YOLO algorithm. Neural Computing and Applications, 34(10), 8253-8274.
33. Talaat, F. M., Aljadani, A., Badawy, M., &Elhosseini, M. (2024). Toward interpretable credit scoring: integrating explainable artificial intelligence with deep learning for credit card default prediction. Neural Computing and Applications, 36(9), 4847-4865.
34. Nievas, E.B., Suarez, O.D., Garciaand,G.B., Sukthankar, R. “Hockey fight detection dataset”, pp. 332–339, Aug. 2011, [Online]. Available: http://visilab.etsii.uclm.es/personas/oscar/FightDetection/
35. Soliman,M.M., Kamal, M.H., El-Massih Nashed,M.A., Mostafa,Y.M., Chawkyand, B.S., Khattab, D. Dec. 2019.
36. Li, H., Wang, J., Han, J., Zhang, J., Yang, Y., & Zhao, Y. (2020). A novel multi-stream method for violent interaction detection using deep learning. Measurement and Control, 53(5-6), 796-806.
37. Ullah, F. U. M., Ullah, A., Muhammad, K., Haq, I. U., & Baik, S. W. (2019). Violence detection using spatiotemporal features with 3D convolutional neural network. Sensors, 19(11), 2472.

38. Asad, M., Yang, J., He, J., Shamsolmoali, P., & He, X. (2021). Multi-frame feature-fusion-based model for violence detection. The Visual Computer, 37(6), 1415-1431.
39. Song, W., Zhang, D., Zhao, X., Yu, J., Zheng, R., & Wang, A. (2019). A novel violent video detection scheme based on modified 3d convolutional neural networks. IEEE Access, 7, 39172-39179.
40. Bilinski, P., & Bremond, F. (2016, August). Human violence recognition and detection in surveillance videos. In 2016 13th IEEE international conference on advanced video and signal based surveillance (AVSS) (pp. 30-36). IEEE.
41. Bilinski, P., & Bremond, F. (2016, August). Human violence recognition and detection in surveillance videos. In 2016 13th IEEE international conference on advanced video and signal based surveillance (AVSS) (pp. 30-36). IEEE.
42. Rendón-Segador, F. J., Álvarez-García, J. A., Enríquez, F., & Deniz, O. (2021). Violencenet: Dense multi-head self-attention with bidirectional convolutional lstm for detecting violence. Electronics, 10(13), 1601.
43. Khan, S. U., Haq, I. U., Rho, S., Baik, S. W., & Lee, M. Y. (2019). Cover the violence: A novel Deep-Learning-Based approach towards violence-detection in movies. Applied Sciences, 9(22), 4963.
44. https://github.com/profmahaveer/VGG19.git