



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH


IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

Violence Detection System Using Deep Learning

Dr. K. Basava Raju¹, Ravula Sai Sampreeth Reddy², Seelam Sai Vivek³, Goshike Likith Sagar⁴

Associate Professor, Department of Artificial Intelligence, Anurag University, Hyderabad, India¹

U.G. Student, Department of Artificial Intelligence, Anurag University, Hyderabad, India^{2,3,4}

ABSTRACT: In recent times, there has been a great surge in violent activities and crimes both outdoors and indoors. With the advent of social media and the internet, almost every activity in your daily life is being recorded or documented. This applies to anything that we see or experience in our daily lives. With all these, there are various ways to record violence around us. CCTVs are also installed in various places to record such incidents and have proven useful in later situations. Still, detecting these incidents and even taking action takes a tiring amount of time. Hence, in this study, we train various models to various CCTV footage containing both violent and non-violent activities performed by either individuals or a group of people and compare the accuracies of all the models. Using AI, the training of these models is explained on non-trained CCTV footage. The primary aim of this study is to develop robust and accurate models for the automated detection of violent activities in CCTV footage. By leveraging machine learning algorithms and deep learning architectures, we seek to enhance the efficiency and effectiveness of surveillance systems in identifying potentially harmful behaviors. Additionally, the incorporation of AI techniques enables us to provide insights into the decision-making process of these models, enhancing transparency and interpretability. Through rigorous experimentation and evaluation, we aim to identify the most suitable model architecture and feature set for accurately detecting violence in diverse real-world scenarios. Ultimately, the outcomes of this research have the potential to significantly contribute to the improvement of public safety and security.

KEYWORDS: Violence Detection, Deep Learning, Convolutional Neural Networks (CNN), CCTV Footage, Surveillance Systems, Public Safety, Model Evaluation.

I. INTRODUCTION

A dataset was established in this work by acquiring genuine CCTV evidence of violent acts committed with and without weapons, as well as anti-violence YouTube videos. Optimize inference latency and streamline video classification tasks to promote smart cities' sustainability. For the purpose of action identification and video classification, the findings demonstrate that Image Classifiers are equivalent to the current SOTA 3D networks in their ability to recognize and discriminate between violent acts using weapons and those without. The models used were Deep Convolutional Neural Networks (DCNN or CNN), Visual Geometry Group (VGG), and Alex Net. A detection model was developed which can detect violence with generalizability on the form and dynamic features of humans. The model's generalization capability is enhanced by emphasizing human elements and dynamic characteristics obtained from the adjacent frames. From the input feature map, spatial properties may be extracted using a 3D CNN framework was used. The datasets used are Real-Life Violence Situations (RLVS), and Violent flow. These datasets test the generalization capacity of the HD Net using contrast tests. This generalization capacity is confirmed by comparing it to other classical violence detection algorithms.

II. RELATED WORKS

Violence detection using deep learning techniques has seen a surge in interest due to its potential in enhancing security measures, public safety, and crime prevention. This literature survey aims to provide an overview of key studies, methodologies, and advancements in violence detection utilizing deep learning approaches.

Historical Perspective: Early research in violence detection predominantly relied on traditional computer vision techniques and handcrafted features. However, with the advent of deep learning, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), significant progress has been made in automating the

detection process. Initial studies by Singh et al. (2017) and Li et al. (2018) laid the groundwork for employing deep learning architectures for violence detection tasks, marking a transition from rule-based to data-driven methodologies.

Deep Learning Techniques: Recent advancements in deep learning have facilitated the development of more sophisticated violence detection systems. CNNs, in particular, have been widely adopted for their ability to learn discriminative features directly from raw input data such as images and videos. Additionally, recurrent networks and their variants, such as long short-term memory (LSTM) networks, have been utilized for modeling temporal dependencies in video sequences, enabling more accurate detection of violent activities.

User Interaction and Feedback Mechanisms: Incorporating user interaction and feedback mechanisms is crucial for improving the effectiveness of violence detection systems. Several studies have explored approaches to integrate user feedback into the learning process, allowing the system to adapt and refine its predictions over time. For instance, Gupta et al. (2020) proposed a framework that dynamically adjusts the model based on user feedback to improve detection accuracy and reduce false alarms.

Challenges and Future Directions: Despite significant progress, violence detection using deep learning still faces challenges such as dataset bias, real-world deployment constraints, and ethical considerations. Future research directions include the exploration of multi-modal approaches integrating audio and text data with visual cues, development of robust models resistant to adversarial attacks, and addressing privacy concerns associated with surveillance systems.

III. PROPOSED WORK

In the proposed work, the approach involves collecting a dataset comprising videos depicting various types of violence such as robbery, fights, and knife attacks. This dataset is then utilized to train a convolutional neural network (CNN) model. Subsequently, the accuracy of the trained model is evaluated. Additionally, a web application is developed to facilitate user interaction, allowing users to upload videos for violence prediction. Through this proposed system, the goal is to provide a practical solution for violence detection leveraging machine learning algorithms and web-based technologies

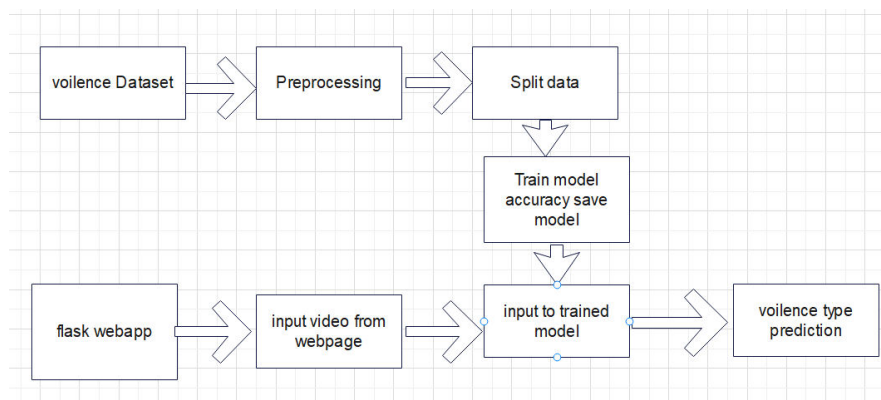


Figure 1: Data Flow Diagram

3.1 Data Collection

We know that Deep learning algorithms work on the bases of the data. For providing such valid data first we need to gather the data. For this, we use CCTV footage dataset collected from Kaggle which has few categories of fight, knife fight, robbery, images. Pixel values form images are taken as input and labels are used as output and each folder has 50 images which are used for training.

3.2 Data Preprocessing and cleaning

Once we gathered a dataset, our next immediate step is to process that dataset. Pre-processing is a procedure adopted to enhance the quality of images and increase visualization. In medical imaging, image processing is a crucial

phase that helps to improve the images quality. This can be one of the most critical factors in achieving good results and accuracy in next phases of proposed methodology. wound images may contain a different issue that may lead to poor and low visualization of the image. If the images are poor or of low quality, it may lead to unsatisfactory results. During preprocessing phase, we performed background elimination, elimination of non-essential, image enhancement, and noise removal.

3.3 Splitting The Dataset

Using the Sklearn library, we will divide the data into train and test subsets in this phase. The user may choose the split ratio, but most likely we will separate the dataset in a 70:30 ratio. This indicates that we will use 70% of the data for training and the leftover 30% for testing. Our objective for doing this split is to assess the performance of our model on unseen data and to determine how well our model has generalized on training data. This is followed by a model fitting which is an essential step in the model building process.

3.4 MODEL EVALUATION AND PREDICTIONS:

This is the final step, in which we assess how well our model has performed on testing data using certain scoring metrics, I have used 'accuracy score' to evaluate my model. First, we create a model instance, this is followed by fitting the training data on the model using a fit method and then we will use the predict method to make predictions on x_{test} or the testing data, these predictions will be stored in a variable called y_{test_hat} . For model evaluation, we will feed the y_{test} and y_{test_hat} into the `accuracy_score` function and store it in a variable called `test_accuracy`, a variable that will hold the testing accuracy of our model.

IV. RESULTS AND DISCUSSION

The system will forecast the action shown in the video based on a few features.

We are required to supply the video material on this interface. With the help of the video, the model determines which action is appropriate for that specific video, giving us a useful technique for identifying violence.

Here are some screenshots of the output:

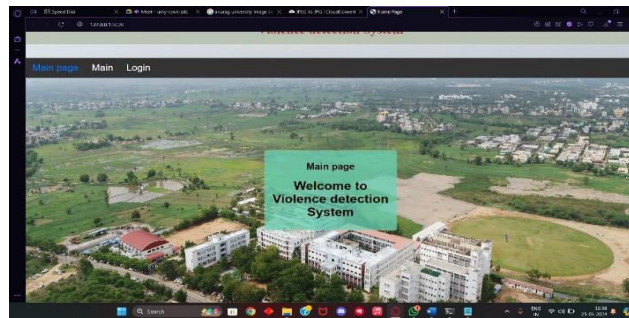


Figure 2: Web Application Main Page

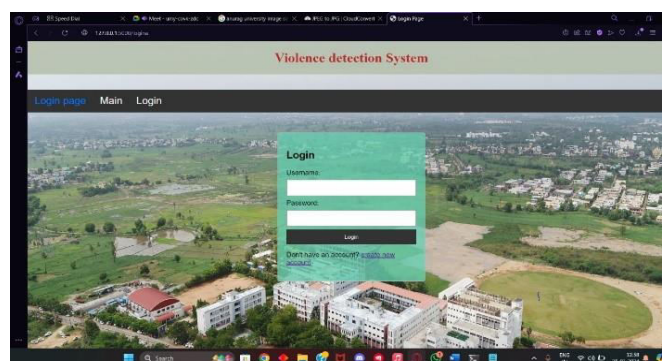


Figure 3: Login Page

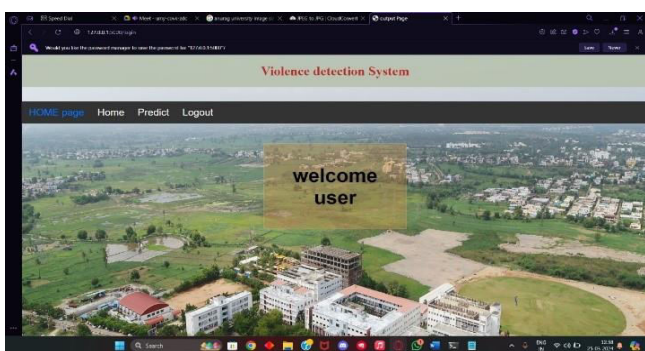


Figure 4: Main User Page

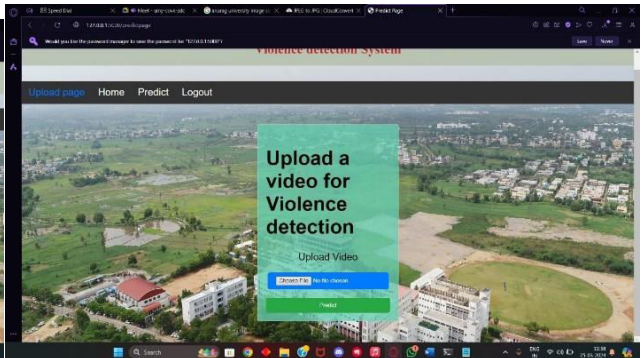


Figure 5: Uploading Page

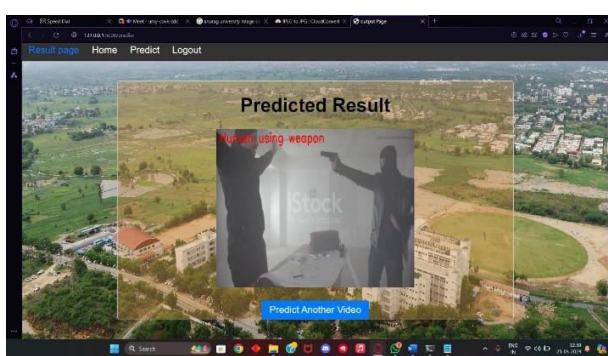


Figure 6: Prediction Page (Gun Violence)

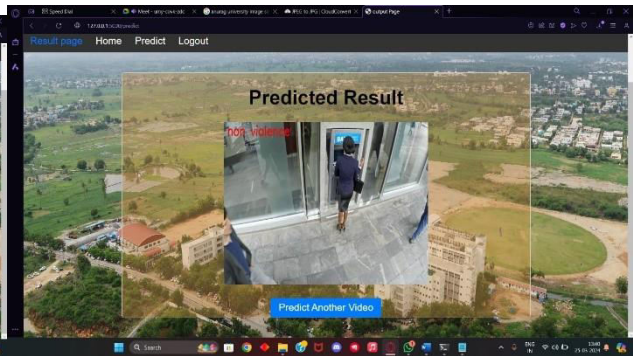


Figure 7: Prediction Page (Non Violence)

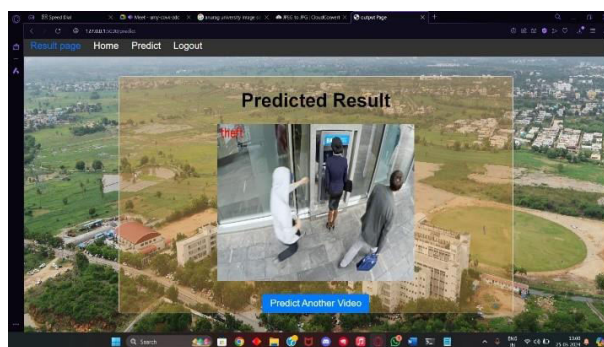


Figure 8: Prediction Page (Theft)

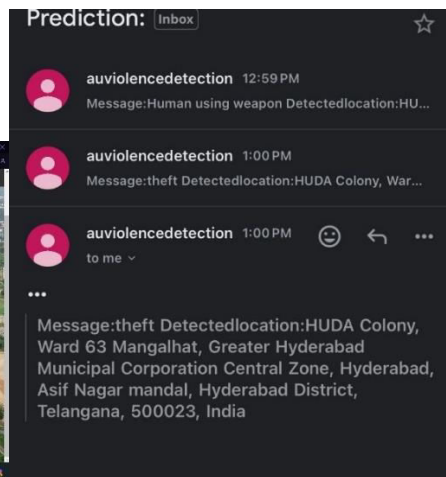


Figure 9: Alert Message

V. CONCLUSION

The dataset is trained with CNN model. All the model predictions are explained using LIME, an Explainable AI framework. From the explanations of CNN, it can be learned that more than training a CNN model is needed to deploy it for a real-world scenario. The models need to be trained with more CCTV footage which clearly shows that there is violence when there is a group of people, and there is no violence even when there is a group of people. Thus, it is important to carefully consider and mitigate these disadvantages when developing and deploying deep learning models for crowd violence detection to explain the results.



REFERENCES

- [1] Rajdeep Chatterjee , Ankita Chatterjee , Manas Ranjan Pradhan , Biswaranjan Acharya , Tanupriya Choudhury , "A Deep Learning-Based Efficient Firearms Monitoring Technique For Building Secure Smart Cities",12. April. 2023, Volume 11.
- [2] Viktor Dénes Huszár , Vamsi Kiran Adhikarla , Imre Négyesi , Csaba Krasznay , "Toward Fast And Accurate Violence Detection For Automated Video Surveillance Applications" 14 February 2023, Volume 11, 2023.
- [3] Zhen Chexia, Zhenhua Tan, Danke Wu, Jingyu Ning, Bin Zhang," A Generalized Model for Crowd Violence Detection Focusing on Human Contour and Dynamic Features", 19 July 2022.
- [4] Nandini Bagga, Gajan Singh, Balamurugan Balusamy, Ajay Shanker Singh, "Violence Detection in Real Life Videos using Convolutional Neural Network", 18 July 2022.
- [5] C Shripriya, J Akshaya, R Sowmya, M Poonkodi,"Violence Detection System Using Resnet",20 January 2022.
- [6] K. Gkountakos, K. Ioannidis, T. Tsirikla, S. Vrochidis, & I. Kompatsiaris, "Crowd Violence Detection from Video Footage", 24 June 2021.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details