National College of Ireland

Detection of Violent activities in the Cloud Computing Environment using Gated Recurrent Unit

MSc Research Project
Cloud Computing

Honey Rajendra Gugale
Student ID: x20133685

School of Computing
National College of Ireland

Supervisor: Divyaa Elango

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Honey Rajendra Gugale
Student ID:	x20133685
Programme:	Cloud Computing
Year:	2021
Module:	MSc Research Project
Supervisor:	Divyaa Elango
Submission Due Date:	31/01/2022
Project Title:	Detection of Violent activities in the Cloud Computing Environment using Gated Recurrent Unit
Word Count:	6198
Page Count:	19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	31st January 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Detection of Violent activities in the Cloud Computing Environment using Gated Recurrent Unit

Honey Rajendra Gugale
x20133685

Abstract

Over the years, violence has been increasing among different communities and people. The surging apprehension has demanded various approaches to counter the violence. Around the world, the authorities have implemented various approaches, from multiple rules and regulations to technologies. However, the measures implemented for decades were not enough to hinder the violence. There are multiple types of violence, such as shoving, thrashing, biting, piercing knives, shooting with a gun, etc. Therefore, in our study, we have introduced a novel approach by implementing advanced deep learning architectures—the study aimed for an optimal model to automatically detect violence through the surveillance system. Various DNN algorithms were considered, such as GRU, RNN, and LSTM. Each algorithm was implemented and assessed using specific evaluation metrics. The best-performing model was considered in the cloud interface for further implementations.

1 Introduction

Nowadays, violence among different communities or people has seen a more significant surge worldwide. Violence is generally seen as physical harm such as fighting, killing, pushing, etc. A statistics by United Nations Office on Drugs and Crime (UNODC) in the world, the Americas region had the highest rate of violence, standing at 16.3% of the pace of the world of countries by intentional homicide rate (n.d.). As violence can be both mentally and physically, we have limited the studies only to detecting physical-based violence in our paper. Violence can affect the nation both financially and ethically. Every country aims to become a secure place without any violence. Crowed marketplaces or public areas such as stadiums and procession arenas are prone to physical violence. Furthermore, the use of dangerous weapons can worsen the situation. Therefore, the intimation to the authorities at the right time is necessary to hinder the violence and other losses. In scenarios such as crowded public places, it could be a tedious task for the authorities to monitor and control the violence at the right time. With the current technological advances, a suitable and efficient system has to be proposed.

In the current era, a proper protocol to monitor violations is required to cease the growing rate of physical breaches. Generally, the monitoring system is based on human interventions, with multiple drawbacks. Primarily, recruiting security can cost the authorities higher financial burdens. Moreover, inattentiveness of human monitoring can also cause delayed initiatives which can be a tremendous hurdle. Utilization of this system can also

make the person's identification hard in case of the crowd in public areas. Indeed, this system has drawbacks, but it also does have its advantages. Therefore, a need for a better system brought in CCTV cameras, although this also had human interventions. But the most outstanding merits were that the locations could be monitored remotely. This system had some plus points over the previous one but needed human interventions. Therefore, a better and efficient system was required to overcome these challenges. Another system utilizing Machine Learning (ML) algorithms was introduced to replace human interventions during violence monitoring. This approach could be utilized by integrating it with the existing camera system to detect violence. Various ML approaches such as K-Nearest Neighbour (KNN), Naïve Bayes were utilized, but they had some limitations.

The ML approach for this detection model was considered efficient as it replaced human interventions. It was also economically viable and feasible than the previous model. But, using ML approaches required a large amount of dataset to train the model. Moreover, it utilized intense computational power for the process. Therefore, a better and robust model was necessary. The introduction of the Deep Learning (DL) approach for violence detection can overcome the challenges followed by implementing ML approaches. With the implementation of the Deep Learning approach, it could be trained with a smaller dataset and lower computational energy. Various Deep Learning approaches can be utilized, such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory Networks (LSTMs), Deep Boltzmann Machine (DBM), Deep Beliefs Network (DBN), etc. In our study, we have only compared the three approaches, RNN, LSTM, and GRU. Researchers have extensively utilized the Deep Learning algorithms in their studies because of their potential to precisely classify and analyze the images and video sequence. Multiple researchers' studies showed and entrenched the average detection output of 97 percent through different Deep Learning algorithms. Therefore, in our paper, we have studied and compared the best and efficient models among the RNN, LSTM, and GRU for violence detection.

1.1 Research Question

- How the Deep learning algorithms can detect the violence activities from the video surveillance camera ?
- Which deep learning architecture accurately identifies the violence behaviour from video data ?

2 Literature Review

This section will discuss some of the studies by researchers in the related domain. This section is further divided into subsections: physical violence, object detection, image processing, deep learning approach for violence detection online, and Deep learning-based violence detection for the public domain.

2.1 Physical Violence Approach

Butts et al. (2016) evaluated a public health model to reduce Gun Violence which is called Cure Violence. Cure violence acts as a remedy to break down violent behavior using guns without any force and punishment. The cure violence model was developed by a renowned physician Gary Slutkin at the University of Illinois at Chicago. Furthermore, this model is enhanced and handled by the same team at the university. The paper studied the various occurrence and scrutinized the outcome of the implementation of cure violence to overcome gun violence. The overall result was credible, but it had challenges that the results fluctuate on multiple factors such as environment, rate of violence, etc. In another study by AlBuhairan et al. (2017) assessed the impact on mental health and academic performance due to physical violence and bullying among the students. This can be considered a serious issue and a threat to the public health system. The study in the paper showed that more males than female students are prone to bullying and physical violence. Moreover, these students had a higher chance of getting into depression and anxiety, with an additional impact on their performance in academics. It also showed the importance of building a peculiar relationship between a parent and a child.

2.2 Object Detection Approach

S. and Tamilselvan (2016) studied the approach of Object Detection. This approach has multiple applications in the real world, out of which implementation in a video surveillance system is the most. This paper provided a detailed overview of object detection and regarded it as a set of multiple approaches. The systematic object detection approach includes environment modeling, motion segmentation, object classification. Both motion segmentation and object classification are further divided into two classes. Motion segmentation includes Background Subtraction and Optical Flow, whereas Object Classification includes Shape-based method and Motion-based method. The object detection approach also requires favorable inputs for efficient output. This paper showed the limited development in the domain of the object detection approach. Zhao et al. (2019) reviews the deep learning approach for object detection in his paper. The paper stated that object detection depends on two factors, namely, object localization and object classification. Furthermore, the conventional object detection approach depends on three phases, Informative Region Selection, Feature Extraction, and Classification. The most common deep learning approach known for object detection is the CNN algorithm. Objection Detection can be implemented in either two ways, called region-based and classification-based. The deep learning framework for this has been regarded as the most robust approach and efficient than the previous models. Through this framework, it can be utilized for various real-time applications such as face detection, pedestrian detection, violence detection, etc.

Szegedy et al. (2016) studied a robust model for object detection using a deep learning algorithm. The implementation of the Deep Learning approach was regarded as a powerful model by the researchers. This paper introduces an enhanced approach using Deep Neural Network (DNN) to present a regression procedure for object detecting box masks. Furthermore, a multi-scale interference procedure for higher resolution object detection with lesser cost. But the limitation that occurred for this study was higher computational time. Therefore, future research could aim for a lower computation cost and time framework. This framework must operate under a single network for multiple

classes. Mandhala et al. (2020) researched Object Detection Employing Machine Learning for Visually Impaired People. The best objection detection algorithm was evaluated against various pre-trained methods, including YOLOv3, Retina Net, and YOLO Tiny. Object detection is critical for identifying and classifying things from pictures or videos. This research could help optically handicapped persons recognize items and rhythms insight. Complications for visually impaired persons can be addressed with an optimal approach. By transforming raw pictures, this algorithm identifies precision depending on the X-Y axis and identifies the individual via voice. In the paper, an optimal technique for object identification is proposed, which assesses several parameters for productivity.

2.3 Image Processing Approach

Hoang (2018) analyzed the approach of Image Processing for the recognition of wall defects utilizing a Machine Learning algorithm. Maintenance and early Detection of wall defects are necessary for building safety. Furthermore, the Detection of wall defection can help the construction entities to provide security, timely deliverance, and satisfactory quality to their customers. This paper described a robust model for wall defect recognition. It utilized steerable filters and projection integrals to extract features from the sequenced dataset. ML algorithms such as Support Vector Machine (SVM) and least-squares SVM were implemented to standardize the segmentation and categorization into five phases, longitudinal crack, transverse crack, diagonal crack, spall image, and intact wall. The overall accuracy rate of this model achieved 85.33 percent. Although, the model could be enhanced, and the accuracy result could be increased. Patil (2021) investigated picture segmentation using a Convolutional Neural Network method in a similar study. The pictures of cats and dogs have been utilized in this project as the sample. The investigation findings were produced using a configurable neural network with the CNN architecture and Keras API. The cat and dog picture collection were taken from the Google database. The algorithm used here can classify pictures with many cycles and parameters. With a lattice filtration of 256, the model's reliability averaged 92.5 percent. This demonstrated that a higher solid system and additional method development might improve yield performance.

iteref10 investigated the data augmenting of Image Processing utilizing DL algorithms in his study. Image enhancement aided in improving analysis result precision. The traditional method of boosting is effective. However, approaches such as Generative Adversarial Network (GAN), Neural Network Augmentation, VGG 16, and Small Net will improve the simulation results. The study also proposed combining the method to improve the categorization problem. This method can assist in overcoming the difficulties associated with generalization error and imbalanced data. The article also demonstrated how to use a backpropagation algorithm to identify and evaluate a wide variety of characteristics from a dataset. Xin and Wang (2019) research proposes a new thorough neural network training parameter for optimum period minimal segmentation inaccuracy relying on loss backpropagation algorithm assessment. Several DL algorithms, including Support Vector Machine, K-Nearest Neighbour, Nave Bayes, Random Forest, and Decision Tree, were integrated. Additionally, cross-entropy and M3CE were used to fine-tune the sample for effective results on the deep learning sample: the Modified National Institute of Standards and Technology database (MNIST) and the Canadian Institute for Advanced Research (CIFAR-10). According to the report's comparative statistics,

the CNN approach has a better efficiency rate of 99.68 percent for the learning set and 83.67 percent for the validation set. Despite this, the RF method requires a minimal calculation period.

2.4 DL approach for Violence Detection through Online

Segun Taofeek Aroyehun (2018) investigated the identification of hostility in media platforms. Tons of articles are published on media platforms at periodic times, and the researchers' suggested approach recognizes each article involving violence uploaded on media platforms. The researcher utilized DNN algorithms and primary function samples for learning and instances with bogus labeling. The Naive Bayes (NB) log-count ratio features with (NB-SVM) method employed in SVM was determined to be the optimum for categorization applications. This serves as the report's benchmark paradigm. The implementation of 7 DNN frameworks is what makes this technique unique. In another paper by U. Dikwatta (2019), they researched violence detection through a social media review. This paper showed that the conventional machine learning approach was outcasted by the novel deep learning approach, as the recognition of violence through either text or images. Therefore, deep learning frameworks like You Only Look Once (YOLO) with word embedding architecture showed a robust and efficient performing algorithm. The preliminary stage in this model is to balance and sort out both positive and negative data from the corpus. Then a text and object detecting module is implemented for the text embedding. In the final stage, a classification model is utilized and implemented to detect violence.

Subramani et al. (2018) in this paper assessed an efficient model of domestic violence crisis identification through Facebook posts utilizing a deep learning approach. Domestic Violence is a significant public health concern in various societies and communities. This paper showed a robust deep learning algorithm to detect domestic violence posts and provide timely supports to the victim with the help of the dedicated authorities. Various deep learning algorithms were implemented and compared, which conversely outcasted the performance of conventional machine learning techniques. The researchers could achieve 94 percent of accuracy through the deep learning approach. This system could help the authorities, medical team, practitioners to detect and support domestic violence victims. Similarly, Sumon et al. (2019) utilized a pre-trained deep learning algorithm to detect Violence through the videos of YouTube. It used pre-trained models such as VGG16, VGG19, and ResNet50. Furthermore, the transfer learning technique such as the LSTM network is combined with the conventional methods to extract the features from the images. The dataset used in this study were sourced from various video-sharing sites like YouTube, Facebook, and Twitter. These pre-trained models were compared to the conventional CNN approach to show the performance ratio. With overall comparison, ResNet50 showed the best performing and efficient model. Moreover, the researchers aim to study a similar approach for implementing it at CCTV cameras and Unmanned Aerial vehicles (UAV).

2.5 DL-based Violence Detection for the Public Domain

Baba et al. (2019) studied the deep learning approach for violence detection in urban areas. A system for smart cities was proposed that could identify the person's behavior

using a Deep Neural Network (DNN). Furthermore, the conventional and DNN approach are compared to provide the most accurate model. The paper utilized the two most publicly available surveillance databases, BEHAVE and ARENA. The DNN algorithm was combined with a well-defined motion feature to build a robust model. Also, the convolutional layers are reduced with the lesser dataset for higher performance without compromising efficiency. Finally, the sensor network was combined with the deep learning approach to make it cost-effective and real-time efficient. In the paper, Singh et al. (2020) used deep learning techniques to implement a platform that identifies or recognizes threats in videos autonomously. The suggested method dynamically detects violence and distinguishes these pictures from regular ones. The primary is a CNN, while the second is an RNN. The footage is first transformed into pixels, followed by the formation of the initial CNN inception and max pool layer, followed by the clustering of extracted features into a common context, and lastly, the RNN is employed. Eventually, in the next phase, the categorization outcome is produced.

In similar research by Yao and Hu (2021), the paper showed the crucial application of intelligent video surveillance for Violent Behaviour Detection (VioBD). The approach of VioBD is necessary for public safety and security. This paper discussed the challenge of violence detection and compared various approaches such as conventional, deep learning, and hybrid frameworks. Furthermore, a publicly available dataset was utilized for this model for assessing the performance. Additionally, the challenges under VioBD and its future trends were stated in the paper. Li et al. (2020) has also researched abusive interactions recognition. The researchers of this paper suggested a durable violent interactions recognition approach relying on a DL technique that uses several streams. The program's efficiency is improved by using three channels, comprising attention-based spatial RGB, temporal, and local streams. The authors presented a multi-paradigm in that territorial ConvNet plus attention, spatial block ConvNet, and temporal ConvNet were used and thus incorporated into regional unanimity. Eventually, the category value merger is used to obtain outcomes.

In the paper by Gathibandhe et al. (2021), the researchers reviewed the deep learning approach for the system of violence detection. The system architecture combines conventional hardware such as CCTV cameras and deep learning algorithms. The paper utilized the Convolutional Neural Network (CNN) approach by combining it with Long Short-Term Memory (LSTM) network and the transfer learning technique. The performance of this model was higher than the performance of the standalone utilization of the CNN approach. This model achieved higher accuracy and precision rate as well. Furthermore, the study showed that this system provides a medium to interact with deep learning models through graphical user interfaces. Zhou et al. (2017) developed a Convolutional Network model for violence detection named Fight Net, which is a long-term temporal architecture. A violence interaction dataset (VID) was proposed, sourced from public repositories relying on the keywords such as Hockey, Movies, HMBD51, and UCF101. Additionally, various techniques such as acceleration field, optical flow, RGB were utilized to extract the motion features from the dataset. Also, the combination of these techniques was evaluated to compare the efficiency. With the overall comparison of other modules with the Fight Net models, the proposed one showed higher accuracy and performance with feasible computation expense.

The comparative analysis for the various studies by different researchers is shown in Table 2.5

Paper Title	Publish Year	Method	Advantages	Future Scope / Disadvantages
A study on Object Detection	2016	Motion segmentation includes Background Subtraction and Optical Flow whereas Object Classification includes Shape-based method and Motion-based method were utilized	Efficient due to hybrid approach	Limited Development
Object detection using machine learning for visually impaired people	2021	YOLOv3, Retina Net, YOLO Tiny	This study would allow visually impaired people to identify objects and patterns in front of them. With an efficient model for this, challenges for visually impaired people canbe overcome.	Although there was amazing outputs but room for improvement is necessary.
Image Processing-Based Recognition of Wall Defects Using Machine Learning Approaches and Steerable Filters	2018	ML algorithms such as Support Vector Machine (SVM) and least-squares SVM were implemented to standardize the segmentation and categorization	The overall accuracy rate of this model achieved 85.33 percent	The efficiency could be increased with utilization of other approaches as well
Real-Time Anomaly Recognition Through CCTV Using Neural Networks	2020	Both CNN and RNN approach is utilized	Automated detection of violence through hybrid DNN approach	no disadvantages found

Table 1 : Comparison of Different Works for this Study

3 Methodology

The increasing violence around the world has urged the need for suitable approaches. The authorities need to effectively supervise and counter the violence as it harms the security and safe state. This violence can be of various types, shoving, thrashing, biting, piercing knife or shooting with a gun, etc. To effectively guard the violence among people, we have utilized technological advances to counter the issue. Using the DNN algorithms, a model is developed to detect violence and embark the authority for the action. Furthermore, to enhance the model's performance, we have integrated the cloud architecture for the convenience of approach. In this section, we will discuss the detailed methodologies of the model.

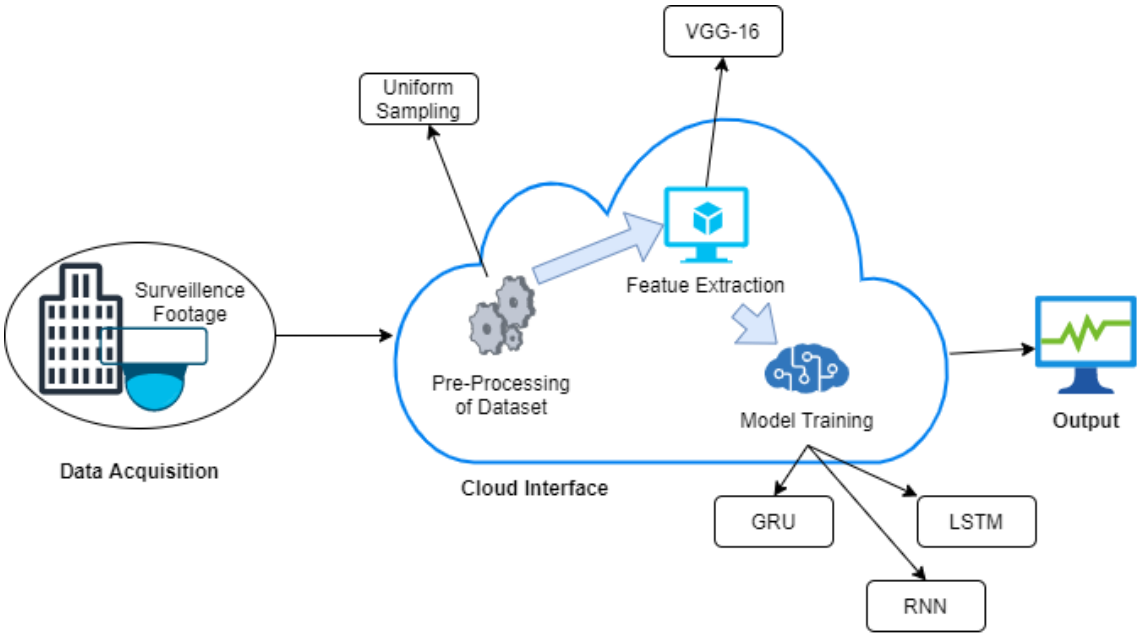

Figure 1: Proposed Methodology

3.1 Data Acquisition

We have utilized a custom-developed dataset introduced by the author, Akti, in her study for our model. The author presented her study, including the fight detection surv dataset (FDSD), at the IPTA 2019. The video sequences in this dataset were compiled from the YouTube videos and the surveillance footage. This dataset included only the fight video sequences with each 2-4 seconds snap. It consists of 300 video sequences with a balanced approach as the fight sequence, and the non-fight sequence was equally distributed, making it unbiased. Furthermore, the video sequence with no background motion was considered for convenience. In addition to this, different environmental condition was also considered.

3.2 Pre-Processing of the Dataset

In this phase, the acquired dataset is pre-processed for the model. Each video sequence consists of multiple frames as the video works under frames per second (FPS). Every video sequence may consist of 50 frames to 250 frames, but only a limited frames shall

be extracted for the model's efficacy. Here, we will have to extract a certain number of frames from the video sequence, but it may not be necessary that obtained frames consist of a fight scene. Therefore, in our model, we have utilized the method of uniform sampling, which will extract every nth frame from the video sequence. Each architecture used may have a different input size, so the cubic interpolation method is implemented to resample the extracted frames. This method effectively resizes the pixels of the acquired samples or the extracted frames. Furthermore, the video sequences are converted into the RGB format for further implementation.

3.3 Feature Extraction

After the pre-processing phase, the feature engineering method is implemented to extract the dominant and crucial features of the given samples. In our model, we implemented the VGG-16 framework for feature engineering. This framework is developed under the architecture of the CNN algorithm with 16 layers of pre-defined weights. These weights are trained using the ImageNet dataset. Each frame's pixels are transformed into the 224x224 pixel for inclusion in the VGG-16 framework. For the retarding the chances of overfitting, the highest convolution layer is removed. The task of VGG-16 in this phase is to imply the convolutional layer, filtration layer and then perform the extraction of feature and dimensionality reduction. Through dimensionality reduction, the high-dimensional parameters are removed, thereby saving time and computation of the model. The dimension of the feature vector is reduced to 4096 with the implementation of dimensionality reduction. Once the methods are implemented, the model's labels have been encoded using a one-hot encoding technique, which encodes the target binary feature fight and no-fight.

3.4 Training of the Model

To effectively detect the indulgence of fighting and non-fighting behavior, various DNN algorithms have been implemented. In our model, we have utilized three algorithms which are gated recurrent unit (GRU), recurrent neural network (RNN), and long-short term memory (LSTM). GRU is an algorithm developed under the RNN algorithm framework by incorporating the gated mechanism. Further, various features of GRU are related to the LSTM algorithm. On the other hand, the RNN algorithm is an artificial neural network that processes sequential samples. Similarly, the LSTM algorithm is developed under the RNN framework, although this model is more complex. These algorithms are implemented, and the binary problem is processed. Each algorithm is implemented to the video surveillance samples, and the output is then evaluated. Here, we considered the ratio of 4:1 for the training set and learning set. Out of 300 samples, the training set consisted of 240 video sequences, whereas the test set consisted of 60 video sequences.

3.5 Evaluation of the Model

After training the model, the performance of each algorithm was assessed in this section through various evaluation metrics. In our model, we have utilized the metrics such as accuracy, precision, recall, and f-1 score. Here, the accuracy of the model determined the number of correct predictions. In comparison, the precision, recall, and f-1 score metrics relied on the confusion metrics. Precision was defined as the ratio of the true positive

sample to the classified positive samples. On the other hand, recall was defined as the ratio of true positive to the actual positive samples. In contrast, the f1-score was defined as the precision and recall weighted average.

4 Design Specification

In this section, we will discuss the design and the framework of each algorithm implemented in our model.

4.1 Gated Recurrent Unit (GRU)

GRU is an algorithm developed under the framework of the RNN algorithm by incorporating the gating mechanism. This model replicates the LSTM based on performance, outcome, and characteristics. It also consists of a forget gate but lacks an output gate. However, during the implementation, GRU showed an efficacious performance with the small samples and dataset. Initially, this algorithm was introduced by Kyunghyun Cho in the year 2014. In this algorithm, there are various gating approaches such as fully gated unit and minimal gated unit with different activation functions such as sigmoid function and hyperbolic tangent. This algorithm controls the flow of information using the gating mechanism. It is simpler than LSTM as it has only a hidden state rather than the separate cell state. Therefore, it also has minimal computational time.

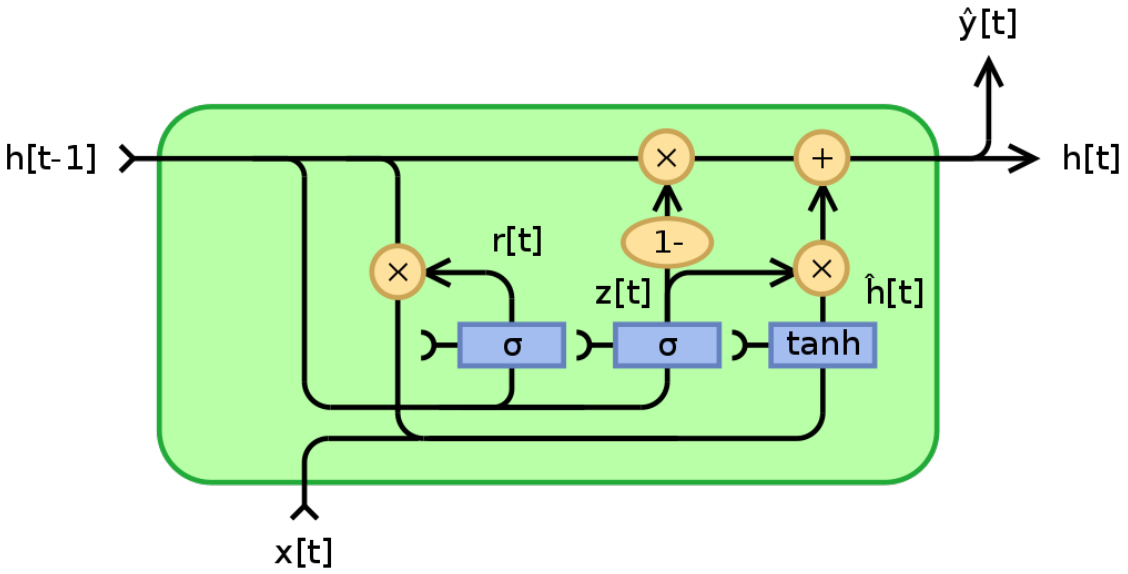

Figure 2: GRU Architecture

4.2 Recurrent Neural Network (RNN)

RNN is an algorithm under the framework of an artificial neural network. A sequential temporal scale is created during the node's connections from the directed and undirected graph. Due to this, the algorithm shows material dynamic characteristics. The features of the feedforward neural network are also reflected in the RNN algorithm. The merit here is that it can process imbalanced sequential length input using its internal memory state. There are two algorithms in the RNN algorithms: finite impulse response and

infinite impulse response. In the limited impulse response, RNN is an acyclic directed graph that can be unrolled or replaced under the supervision of the strictly feedforward neural network, whereas, in the infinite impulse response RNN, it cannot be unrolled. However, each approach inhibits the storage modules.

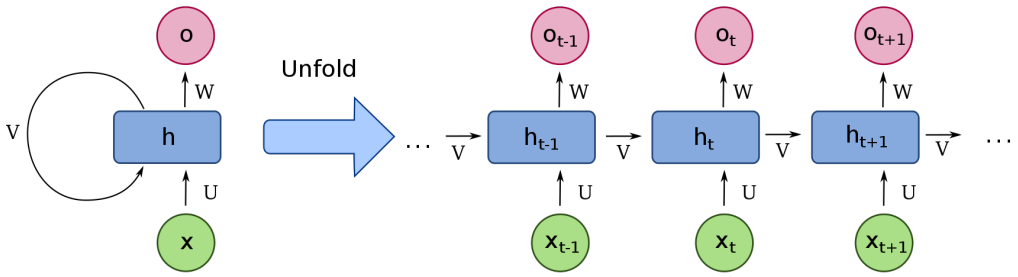

Figure 3: RNN Architecture

4.3 Long-Short Term Memory (LSTM)

Like the GRU algorithm, the LSTM algorithms also rely upon the RNN framework. LSTM consists of feedback connections. This algorithm can process the entire sequences of the samples, whether it be videos or speech. The LSTM framework consists of the input gate, output gate, and forget gate cells. LSTM algorithm overcame the challenge faced by the other similar algorithms. It can counter the vanishing gradient problem encountered in the RNN algorithm. Furthermore, the most outstanding merit over the other similar traditional approaches is the relative insensitivity to the gap length. There are different variants of this algorithm: LSTM with a forget gate, Peephole LSTM, Peephole convolutional LSTM.

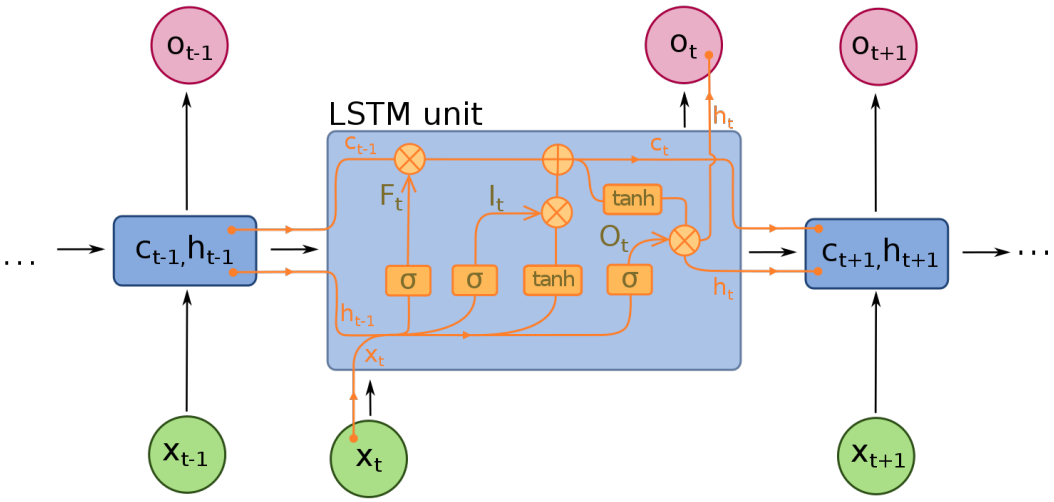

Figure 4: LSTM Architecture

5 Implementations

Through the detailed study and the implementation of this model, our preliminary objective is to detect violence from the surveillance video by integrating the trained model in the cloud framework. By utilizing various DNN algorithms, the violence is detected through multiple video sequences. We implemented three algorithms: GRU, RNN, and LSTM. Each algorithm was implemented on the acquired dataset, and the performance was precisely evaluated. Before training the model with these algorithms, the dataset was initially pre-processed and feature extracted. Here for the pre-processing, uniform sampling and cubic interpolation were implemented. During the implementation of the uniform sampling method, every 5th frame of the video sequence was considered irrespective of the total number of frames in video samples. Then for the extraction of features from the dataset, the framework of VGG-16 is utilized. After this, the algorithms were implemented, and through various metrics such as accuracy, precision, recall, and f-1score, the model was evaluated. Each algorithm implemented consisted of 4 layers: the input layer, output layers, and two hidden layers with the batch size of 5.

Furthermore, the number of epochs utilized stood at 25, which guides the model to learn from the training set. In addition to this, the tanh activation function and Adam optimizer were utilized in our model. The tanh activation function sets both input and output values from -1 to 1.

In comparison, the Adam optimizer enhances the model by effectively handling various adverse gradients. During the implementation of the model, the python programming language was considered, and various libraries were utilized such as NumPy, Pandas, TensorFlow, Keras, random, matplotlib, cv2, etc. With this, the tensor-board was utilized to visualize the samples and outcomes conveniently. Also, the seaborn library supports the visualization of the data. The library based on the computer vision called OpenCV was also utilized in this model, which was called upon through the module cv2. The best-trained model is then stored in the cloud interface for further implementations. Apart from these, to effectively run the model in the system, the following is the requisite specification of the system environment.

Operating System	Windows 10
GPU	NVIDIA Titan RTX
RAM	32 GB
Hard Disk	256 GB
Programming Language	Python
User Interface	Google Collab
Library Implemented	OpenCV, pandas, numpy, matplotlib, random, Keras, TensorFlow

Table 2 : System Specification

6 Evaluation

In this section, we have discussed the detailed evaluation of each algorithm using the evaluation metrics. Here, we have considered accuracy, precision, recall, and f-1 score metrics. There are the following two sub-sections here.

6.1 Experiment 1 / Evaluation Based on Accuracy

Here, the accuracy of each algorithm implemented is evaluated. The accuracy is the number of correct predictions over the total number of samples. We have considered three algorithms which are GRU, RNN, and LSTM. For each algorithm, the accuracy of both the learning and validation sets is discussed with the epoch set at 25. Furthermore, the input dimensions of the samples were considered as 5x4096 through the feature extraction technique by implementing the VGG-16 framework.

6.1.1 Gated Recurrent Unit (GRU)

Considering the first algorithm, the GRU is built under the framework of the RNN algorithm by incorporation the gating mechanism. For the implementation of this model, 240 samples were utilized for the learning set, and 60 samples were utilized for the validation set. This showed that the dataset was split into the ratio of 4:1. Furthermore, the epoch utilized here was set at 25 with a batch size of 5 and verbose of 2. The learning set and validation set are considered, but the model's actual performance depends on the validation accuracy. The algorithm implementation utilized the Tanh activation function and Adam optimizer. During the implementation, the highest learning accuracy achieved was 100%, whereas the highest validation accuracy stood at 93.33%. As in the figure, it was seen that at epoch less than 5, the validation accuracy was pacing more than the learning accuracy, but final epoch, the learning accuracy was more than the validation accuracy.

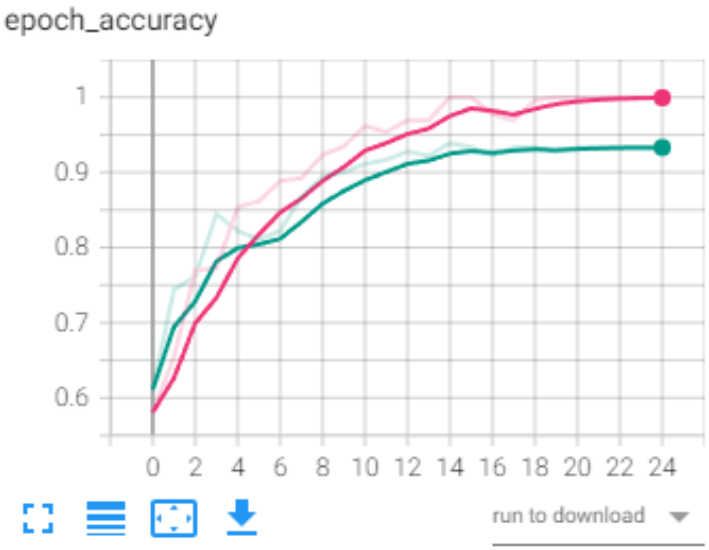

Figure 5: Accuracy of GRU Algorithm

6.1.2 Recurrent Neural Network (RNN)

In the second algorithm, a similar environment as GRU was considered for the implementations. The dataset was also split into the ratio of 4:1 for the learning set and the validation set. In addition to this, the epoch is set at 25 with the verbose at 2 and a batch size of 5. Unlike the GRU algorithm, the ReLu activation function is utilized here, and the Adam optimizer is used. For data visualization, the tensor-board was connected to the user interface. Considering the figure, it was observed that the highest learning accuracy achieved stood at 100% and the highest validation accuracy achieved stood at 87.78%. Like GRU, the learning accuracy was lower than validation accuracy during the inception epoch, but the learning accuracy paced the validation accuracy with the increasing epoch.

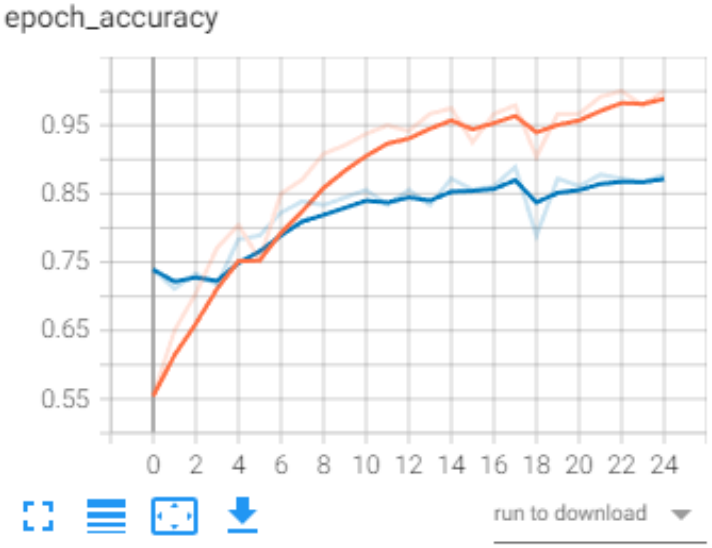

Figure 6: Accuracy of RNN Algorithm

6.1.3 Long-Short Term Memory (LSTM)

The third algorithm of our model, LSTM, is also developed under the RNN framework by incorporating the feedback connections. This algorithm consists of the memory state, which can process the model effectively. Like the other implemented algorithms, the program implementations are also the same for this algorithm. The dataset split considered here was 240 samples for the learning set and 60 samples for the validation set. In addition to this, the epoch was set at 25 with the verbose at 2 and a batch size of 5. Like, GRU algorithm, Tanh activation function, and Adam optimizer are considered for our model. The learning accuracy and validation accuracy were assessed for every epoch in our study using the graphical representation. The learning accuracy was impressive here, which stood at 100% whereas the validation accuracy achieved in the model stood at 88.89%. It was observed that with every increasing epoch, the accuracy of both the learning set and validation set increased. Although, the validation accuracy was lower than the learning set.

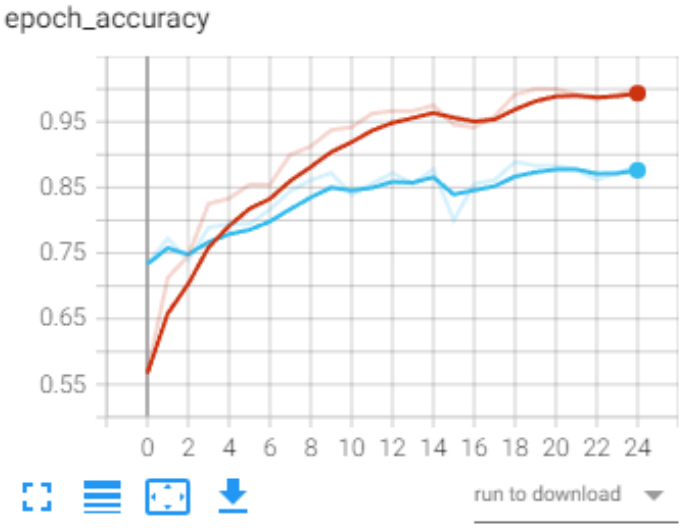

Figure 7: Accuracy of LSTM Algorithm

6.2 Experiment 2 / Evaluation Based on PRF Metrics

This section will assess the metrics such as precision, recall, and f-1 score for each algorithm implemented. These metrics are evaluated based on the confusion matrix, which states true positive, true negative, false positive, and false negative in the processed model. The algorithm considered in our models is GRU, RNN, and LSTM. The true positive assessed are 83, 75, and 77 for the algorithms, respectively, whereas the true negative assessed are 85, 83, and 82. On the other hand, false-positive assesses 8, 10, and 11, respectively, whereas false-negative assesses 4, 12, and 10, respectively.

6.2.1 Precision

Here, precision is the ratio of true positive samples to the total number of classified positive samples. For our model, we have evaluated the precision of each model, which are GRU, RNN, and LSTM. Assessing the confusion matrix, the precision of the learning set and validation set of the GRU algorithm is 0.91 and 0.96, respectively, whereas, for the RNN algorithm, the learning precision is 0.88, and the validation precision is 0.87. On the other hand, considering the precision of the LSTM algorithm, the learning precision is 0.88, and the validation precision is 0.89. Although to assess the accurate performance of the model, we have considered the validation precision. The figure shows that the GRU algorithm has the highest precision in both the learning set and validation set, whereas the lowest precision was assessed in the RNN algorithm.

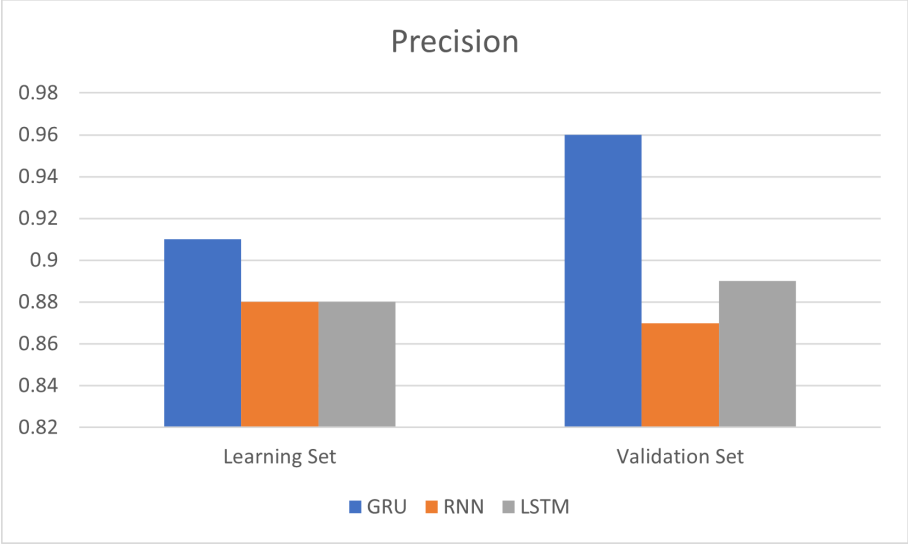

Figure 8: Precision of Algorithms

6.2.2 Recall

The second metric is called recall, which evaluates the ratio of true positive samples to the total number of positive samples. Here, we have also evaluated the recall for all three algorithms: GRU, RNN, and LSTM. The following figure evaluates the recall of each algorithm in a graphical representation. In the GRU algorithm, the learning recall is 0.95, and the validation recall is 0.91, whereas in the RNN algorithm, the learning recall is 0.86, and the validation recall is 0.89. For the LSTM algorithm, the learning recall is 0.89, and the validation recall is 0.88. It is observed that the recall of the GRU algorithm is the highest of all, whereas the recall of the LSTM algorithm is the lowest.

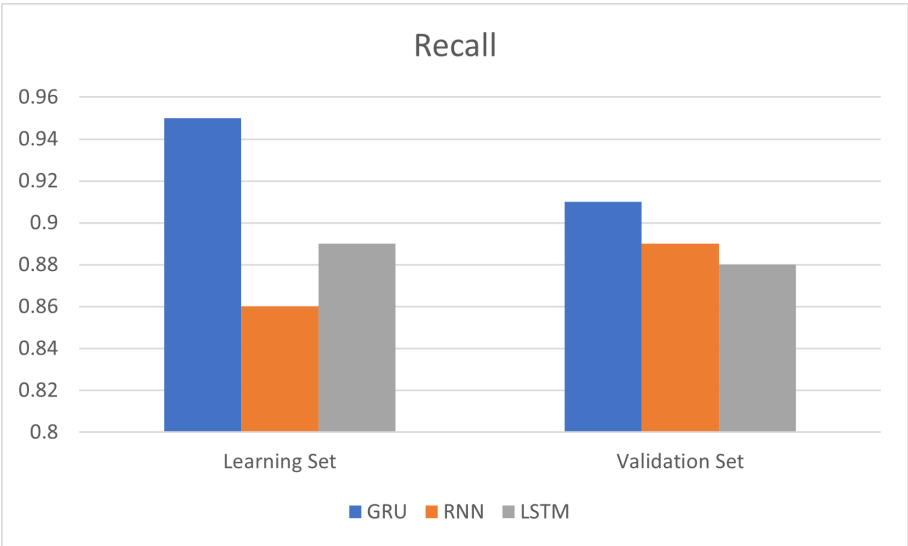

Figure 9: Recall of Algorithms

6.2.3 F-1 Score

Then comes the third metric, the f-1 score, which is the weighted average of the precision and recall of a model. The f-1 score ranges from 0 to 1. By analyzing the figure, the f-1

score of each algorithm can be observed. For the GRU algorithm, the learning f-1 score stood at 0.93, whereas the validation f1-score stood at 0.93. In the RNN algorithm, the learning f-1 score and validation f-1 score are 0.87 and 0.88, respectively, whereas for the LSTM algorithm, the learning f1-score is 0.88, and the validation f1-score is 0.89. As the precision and recall, the validation f-1 score for the GRU algorithm is the highest among all other algorithms. On the other hand, the validation f1-score for the RNN algorithm is the lowest.

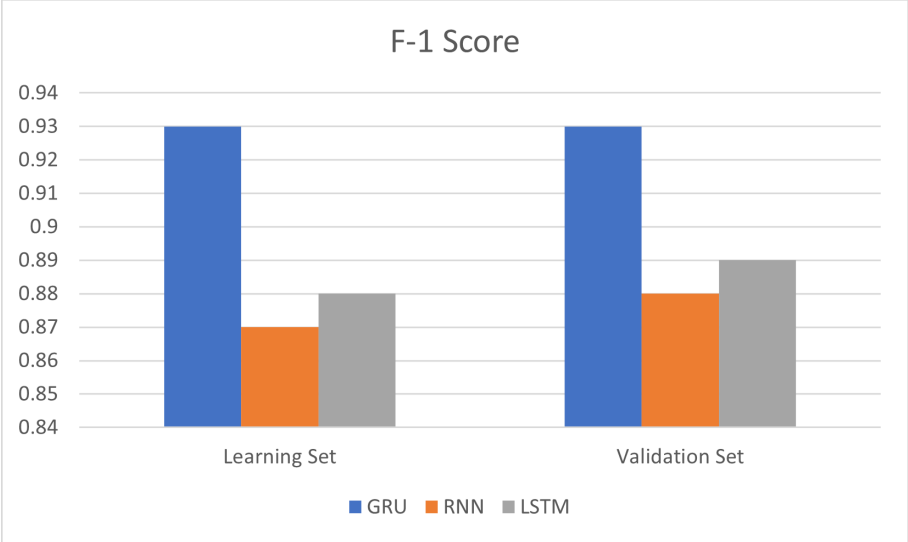

Figure 10: F-1 Score of Algorithms

6.3 Discussion

Our preliminary goal is to attain the best-performing algorithm to implement in the model through these experiments. Our model initially acquires the dataset samples for the implementations and pre-processes them to imbalance the distortion and noise in the samples. During the pre-processing phase, uniform sampling is utilized to acquire specific frames from the video. Also, the attained video sequence is converted to the RGB format. Once done with the pre-processing, the features are extracted from the samples using the VGG-19 framework. This framework implements various approaches such as filtration, convolutional, feature extraction, dimensionality reduction, etc.

Furthermore, the feature vectors are sized to the pixels 4096. Then various DNN algorithms are implemented, GRU, RNN, and LSTM. Multiple metrics are implemented to evaluate the performance, assessing different factors. Metrics considered in our models are accuracy, precision, recall, and f1-score. It was assessed that the GRU algorithm showed higher performance than the RNN and LSTM. The accuracy achieved in the GRU algorithm stood at 93.33% with precision, recall, and f-1 scores at 0.96, 0.91, and 0.93, respectively. On the other hand, the RNN algorithm achieved an accuracy of 87.78% with the precision, recall, and f1-score of 0.87, 0.89, and 0.88. In contrast, the LSTM algorithm achieved an accuracy of 88.89% with precision, recall, an f-1 score of 0.89, 0.88, and 0.89. Here, it is observed that the RNN algorithm's accuracy and LSTM algorithm are somewhat similar. Not only this, but the precision, recall, and f1-score of these algorithms are also fairly similar. Only the GRU algorithm achieved higher accuracy and

other metrics when compared with these algorithms. Once the best-performing algorithm is obtained, the model is stored in the cloud database. We utilized the Google Collab for the user interface as it depends on cloud architecture, and through this, the computational requirement is conveniently derived. The samples are processed in the model during the implementation, and the output is provided in binary form.

7 Conclusion

The violence among people around different geographic has seen a tremendous surge in recent years. Authorities implement various rules and regulations to hinder violence, but these are somewhat inefficacious. Therefore, an alternative and a robust method shall be introduced to counter the challenges. Our study established an advanced deep learning-based solution to counter this challenge. Using the DNN algorithms, our model will detect the violence through surveillance footage and notify the authorities. This will hinder the violence and imply the requirement of a safe and secure environment. Our study utilized three DNN algorithms: GRU, RNN, and LSTM. The best-performing model GRU has been inhibited in the cloud architecture and used for accurately detecting violent behavior. Our current experiment has been performed with a small sample of videos of surveillance cameras. More video samples can be used to train the models to develop a robust with more accuracy. Also, due to technological advancement, many pre-trained and transfer learning models are introduced with pre-defined weights to achieve better accuracy. Those can be explored. This model can be implemented in various social welfare-based applications with better access and advancing technologies. This model could be enhanced using more robust algorithms and effectively tuning the parameters in the future scope of work. The social ecosystem could become safe and secure to interact through such applications.

References

AlBuhairan, F., Abou Abbas, O., Sayed, D., Badri, M., Alshahri, S. and Vries, N. (2017). The relationship of bullying and physical violence to mental health and academic performance: A cross-sectional study among adolescents in saudi arabia, *International Journal of Pediatrics and Adolescent Medicine* **4**.

Baba, M., Gui, V. and Dan, P. (2019). Deep learning approach for violence detection in urban areas, *ITM Web of Conferences* **29**: 03009.

Butts, J., Roman, C., Bostwick, L. and Porter, J. (2016). Cure violence: A public health model to reduce gun violence, *Annual review of public health* **36**.

Gathibandhe, S., Chimantrawar, A., Pusdekar, S. and Dhole, V. (2021). Surveillance violence detection system, *International Journal of Computational and Electronic Aspects in Engineering* **2**.

Hoang, N.-D. (2018). Image processing-based recognition of wall defects using machine learning approaches and steerable filters, *Computational Intelligence and Neuroscience* **2018**: 1–18.

Li, H., Wang, J., Han, J., Zhang, J., Yang, Y. and Zhao, Y. (2020). A novel multi-stream method for violent interaction detection using deep learning, *Measurement and Control* **53**: 002029402090278.

Mandhala, V., Bhattacharyya, D., Bandi, V. and Rao, N. T. (2020). Object detection using machine learning for visually impaired people, *International Journal of Current Research and Review* **12**: 157–167.

of countries by intentional homicide rate, L. (n.d.).
URL: *https://en.wikipedia.org/w/index.php?title=List*$_o f_c ountries_b y_i ntentional_h omicide_r ateoldid =$
1047504212

Patil, A. (2021). Image recognition using machine learning, *Social Science Research Network* .

S., M. and Tamilselvan, L. (2016). A study on object detection, *International Journal of Pharmacy and Technology* **8**: 22875–22885.

Segun Taofeek Aroyehun, A. G. (2018). Aggression detection in social media: Using deep neural networks, data augmentation, and pseudo labeling.

Singh, V., Singh, S. and Gupta, P. (2020). Real-time anomaly recognition through cctv using neural networks, *Procedia Computer Science* **173**: 254–263.

Subramani, S., Wang, H., Vu, H. and Li, G. (2018). Domestic violence crisis identification from facebook posts based on deep learning, *IEEE Access* **PP**: 1–1.

Sumon, S., Goni, R., Hashem, N., Shahria, T. and Rahman, M. (2019). Violence detection by pretrained modules with different deep learning approaches, *Vietnam Journal of Computer Science* **7**.

Szegedy, C., Toshev, A. and Erhan, D. (2016). Deep neural networks for object detection, *Advances in Neural Information Processing Systems* **26**.

U. Dikwatta, T. G. I. F. (2019). Violence detection in social media-review, *Vidyodaya Journal of Science* .

Xin, M. and Wang, Y. (2019). Research on image classification model based on deep convolution neural network, *EURASIP Journal on Image and Video Processing* **2019**.

Yao, H. and Hu, X. (2021). A survey of video violence detection, *Cyber-Physical Systems* pp. 1–24.

Zhao, Z.-Q., Zheng, P., Xu, S.-T. and Wu, X. (2019). Object detection with deep learning: A review, *IEEE Transactions on Neural Networks and Learning Systems* **PP**: 1–21.

Zhou, P., Ding, Q., Luo, H. and Hou, X. (2017). Violent interaction detection in video based on deep learning, *Journal of Physics: Conference Series* **844**: 012044.