RSET
RAJAGIRI SCHOOL OF
ENGINEERING & TECHNOLOGY
(AUTONOMOUS)

Project report On

VigilanceX:Violence Detection System in Jails and Mental Asylums

Submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology

in

Information Technology

By

Ashwin Saji Kumar (U2004023)

Cyriac John (U2004031)

Megha Milton (U2004054)

Roshan Xavier (U2004061)

Under the guidance of

Ms. Bency Wilson

Information Technology
Rajagiri School of Engineering & Technology (Autonomous)
(Parent University: APJ Abdul Kalam Technological University)
Rajagiri Valley, Kakkanad, Kochi, 682039
December 2023

CERTIFICATE

This is to certify that the project report entitled ***"VigilanceX:Violence Detection System in Jails and Mental Asylums"*** *is a bonafide record of the work done by* ***Ashwin Saji Kumar (U2004023), Cyriac John (U2004031), Megha Milton (U2004054), Roshan Xavier (U2004061)****, submitted to the Rajagiri School of Engineering & Technology (RSET) (Autonomous) in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology (B. Tech.) in "Information Technology" during the academic year 2023-2024.*

Ms.Bency Wilson	Ms.Jeshmol P J	Dr.Neeba E.A
Assistant Professor	Assistant Professor	Associate Professor
Project Guide	Project Coordinator	Head of the Department
Dept of IT	Dept of IT	Dept of IT
RSET	RSET	RSET

ACKNOWLEDGMENT

We wish to express our sincere gratitude towards **Dr.P.S Sreejith**, Principal of RSET, and **Dr.Neeba E.A**, Head of the Department of Information Technology for providing us with the opportunity to undertake our project, "VigilanceX:Violence Detection System in Jails and Mental Asylums".

We are highly indebted to our project coordinators, **Ms.Jeshmol P J**, Assistant Professor, IT, **Ms.Bency Wilson** , Assistant Professor, IT ,**Ms.Ancy C A**, Assistant Professor,IT, **Prof.Kuttyamma A.J.**, Professor, IT for their valuable support.

It is indeed our pleasure and a moment of satisfaction for us to express our sincere gratitude to our project guide **Ms.Bency Wilson** for her patience and all the priceless advice and wisdom she has shared with us.

Last but not least, we would like to express our sincere gratitude towards all other teachers and friends for their continuous support and constructive ideas.

Ashwin Saji Kumar

Cyriac John

Megha Milton

Roshan Xavier

Abstract

This project explores the application of Long-term Recurrent Convolutional Networks (LRCN) in the context of violence detection systems for enhanced safety and security in institutional environments such as jails and mental asylums. The LRCN model, a combination of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, is leveraged to recognize violent behaviours within video streams. This advanced model allows for the extraction of spatial features using CNN layers and temporal sequence modeling with LSTM layers, enabling the system to learn spatiotemporal patterns crucial for accurate violence detection.

The project focuses on the integration of the LRCN model with a Telegram bot for real-time alerting and response. Upon detecting violent incidents in the video streams, the LRCN model triggers alerts through the Telegram bot, providing instant notifications to relevant authorities. The Telegram bot facilitates seamless communication and coordination among stakeholders, enabling swift action to mitigate potential risks and ensure the safety of occupants within these facilities.

Through rigorous experimentation and evaluation, the effectiveness and reliability of the LRCN-based violence detection system integrated with the Telegram bot are demonstrated. The research contributes to advancing technology-driven solutions for proactive security measures in high-risk environments, fostering safer and more secure institutional settings.

Contents

List of Figures

Chapter 1

Introduction

In the realm of correctional facilities and mental health institutions, where ensuring the safety of inmates and staff is paramount, a groundbreaking initiative known as Vigilance X is underway. This dedicated Violence Detection System is meticulously designed to address the unique challenges within these environments, leveraging advanced technologies such as the Long-term Recurrent Convolutional Network (LRCN) model.

Vigilance X represents a significant leap forward in utilizing state-of-the-art artificial intelligence and video analytics to not only identify but also proactively prevent instances of violence in real-time. Unlike traditional methods, which rely on reactive responses, Vigilance X operates with a proactive approach, empowering security personnel with tools to intervene decisively and prevent the escalation of potentially harmful situations.

At the core of Vigilance X lies the strategic utilization of cutting-edge deep learning techniques, particularly the Long-term Recurrent Convolutional Network (LRCN) model. This model, renowned for its ability to capture both spatial and temporal features in video data, enables Vigilance X to discern intricate patterns within visual data with exceptional accuracy. By incorporating the LRCN model, Vigilance X surpasses the capabilities of conventional methods, navigating the complexities of correctional settings and mental health institutions with unmatched acuity.

The integration of deep learning-based violence detection systems into correctional facilities and mental health institutions marks a pivotal moment in security protocols. Vigilance X not only detects violence but also aims to prevent potential threats through its proactive approach. This initiative reflects a steadfast commitment to leveraging cutting-edge technologies, such as the LRCN model, to redefine security paradigms and create environments that prioritize safety and well-being.

1.1 Problem Statement

In correctional facilities like jails and mental asylums, ensuring the safety of inmates and staff is an ongoing challenge. Incidents of violence within these facilities can have severe consequences, including injuries and property damage. Therefore, there is a critical need for a specialised Violence Detection System designed to proactively identify and respond to violence in real time using surveillance camera technology.

This specialised system would leverage advanced surveillance technology to analyse real-time video feeds, employing sophisticated algorithms to detect patterns indicative of violent behaviour. By promptly identifying potential threats, the Violence Detection System aims to provide an efficient means of intervention, minimising the risks associated with delayed responses. This proactive approach not only enhances the overall safety of the facility but also contributes to a more secure and controlled environment for both inmates and staff.

1.2 Project Objective

The objective of VigilanceX is to establish a robust Violence Detection System tailored for correctional facilities and mental health institutes. The key objectives include:

- Ensure Safety: VigilanceX focuses on minimising harm and injuries resulting from violence within correctional facilities and mental health institutes. The system prioritises the safety of inmates, staff, and patients through advanced threat detection mechanisms, immediate response protocols, and comprehensive emergency preparedness training.

- Prevent Violence: Proactively identifying early signs of potential violence is a key feature of VigilanceX. By utilising early warning systems and intervening to de-escalate situations promptly, the system contributes to the creation of a secure environment, minimising the likelihood of violent incidents and promoting overall well-being.

- Customization and Adaptation: VigilanceX is designed to be adaptable, offering tailored solutions to suit the unique needs of each facility. With flexible detection

criteria and a modular architecture, the system can be customised to accommodate diverse facility dynamics, ensuring scalability and responsiveness to changing requirements.

- Integration with Existing Infrastructure: Seamless integration with current security and healthcare systems is a hallmark of VigilanceX. Leveraging surveillance cameras, alarm systems, and communication networks, the system provides a comprehensive and interconnected approach. The centralised control interface facilitates efficient management and monitoring of security measures.

- Minimise False Positives: VigilanceX is committed to enhancing system accuracy by continuously refining machine learning models. Through iterative improvement processes and user feedback mechanisms, the system minimises unnecessary alerts, ensuring that identified threats are reliable indicators rather than false positives.

- Rehabilitation and Well-being: In addition to addressing security concerns, VigilanceX supports rehabilitation efforts by integrating therapeutic programs and providing access to mental health professionals. Prioritising the overall well-being of individuals, the system plays a role in preventing self-harm or harm to others, aligning with a holistic approach to rehabilitation.

- Public Perception and Accountability: Enhancing public perception and institutional accountability is achieved through transparent communication about VigilanceX's commitment to safety and security. Internal and external oversight mechanisms are in place, demonstrating a dedication to responsible and ethical use of violence detection technology and ensuring accountability at all levels.

1.3 Design and Implementation Constraints

Implementing violence detection systems in correctional facilities and mental health institutes poses several challenges:

- Privacy Concerns:-
  - Balancing Security and Privacy: Implementing violence detection systems involves striking a delicate balance between ensuring the safety and security

of individuals within correctional and mental health facilities while respecting their privacy rights. The challenge lies in designing systems that can effectively detect potential threats without compromising individual privacy.

  - Legal and Ethical Compliance: Adhering to existing legal frameworks and ethical standards is crucial. This includes obtaining explicit consent from individuals being monitored and ensuring that the surveillance methods employed align with established regulations governing privacy and data protection.

- Ethical Considerations:-

  - Informed Consent: Obtaining informed consent from individuals in correctional facilities and mental health institutes is essential. Users must be aware of the surveillance measures in place and their purpose, allowing them to make informed decisions about participating in such systems.

  - Preserving Autonomy and Dignity: Ensuring that surveillance technology respects the autonomy and dignity of individuals is critical. The implementation should avoid unnecessary intrusiveness and strive to maintain a balance between security measures and respecting the rights and dignity of those being monitored.

- False Positives:- Achieving the right balance between system sensitivity and specificity is a significant challenge. A system that is too sensitive may result in numerous false positives, causing unnecessary disruptions, while a system that is too specific may miss genuine threats. Continuous refinement and tuning of algorithms are necessary to minimize false positives.

- Data Handling & Security:- Managing extensive datasets securely is paramount. Encryption, access controls, and regular audits are essential components of a robust data protection strategy. Compliance with data protection regulations, such as GDPR or HIPAA, is crucial, considering the sensitive nature of the data being collected.

- System Customization:- Correctional facilities and mental health institutes vary widely in terms of size, layout, and operational dynamics. Designing systems that

can be customised to adapt to these variations is essential. This may involve modular components or configurable parameters to meet specific requirements of different facilities.

- Resource Constraints:- Operating within limited computational resources is a common challenge. It requires optimising algorithms for efficiency and exploring hardware solutions that can provide the necessary processing power without overwhelming the infrastructure.

- Training & User Acceptance:-

  - Staff Training: Ongoing training efforts for correctional and mental health facility staff are crucial. They need to be proficient in using and interpreting the violence detection system effectively. Training programs should cover both the technical aspects and the ethical considerations surrounding the system's use.

  - User Acceptance: Gaining acceptance from staff is vital for the successful integration of violence detection systems. Addressing concerns, providing transparent communication, and demonstrating the benefits of the system in enhancing overall security can contribute to better user acceptance.

- Patient Sensitivity:- Designing systems with sensitivity to the emotional and mental well-being of individuals in mental health institutes is essential. The technology should be calibrated to minimise distress, and mechanisms should be in place to handle situations where individuals may be particularly vulnerable to the effects of surveillance.

Chapter 2

Literature Survey

Violence Detection in Videos using Deep Recurrent and Convolutional Neural Networks

A. Traoré and M. A. Akhloufi, "Violence Detection in Videos using Deep Recurrent and Convolutional Neural Networks," 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, 2020, pp. 154-159.[1]

In recent years, substantial advancements have been made in the realm of violence detection, with a diverse range of methods emerging. Two primary categories of approaches have surfaced: classical machine learning and deep learning techniques. A notable contribution in this domain involves the integration of recurrent neural networks (RNNs) and 2-dimensional convolutional neural networks (2D CNNs) for violence detection. A crucial innovation introduced in this work is the inclusion of optical flow, a mechanism designed to encode movements within video scenes. Optical flow becomes particularly significant in the context of violence detection, where the understanding of nuanced temporal dynamics is pivotal. This method addresses the limitations of static frames, which may fail to capture the complexity of dynamic visual information. The proposed deep learning architecture takes an end-to-end approach, employing RGB frames and optical flow in conjunction with a CNN-LSTM network. This amalgamation allows for a comprehensive understanding of both spatial and temporal features within video data, thereby enhancing the model's ability to accurately discern violent scenes. To delve into the specifics of the methods employed, the integration of recurrent neural networks (RNNs) with 2-dimensional convolutional neural networks (2D CNNs) is a key architectural choice. RNNs excel in capturing temporal dependencies by incorporating sequential information, which is crucial for understanding the temporal evolution of actions in videos. On the other hand,

2D CNNs are adept at extracting spatial features from static frames, enabling the model to comprehend the visual content of individual frames effectively. This hybrid architecture leverages the complementary strengths of both models, providing a more comprehensive analysis of video content. As with any method, there are inherent advantages and disadvantages. The strength of this hybrid architecture lies in its ability to capture both spatial and temporal features, enabling a holistic analysis of video data. This proves advantageous in discerning violent actions that involve both spatial and temporal intricacies. On the flip side, the complexity of such models may pose challenges in terms of computational requirements and training time. Striking the right balance between model complexity and efficiency becomes a critical consideration for practical deployment. The incorporation of optical flow further enhances the model's capability to interpret motion dynamics, a critical factor in violence detection where rapid and complex movements often characterise violent actions. Optical flow, in this context, proves beneficial by encoding these motion patterns, providing a valuable dimension to the overall feature representation. The use of optical flow introduces computational overhead, as calculating optical flow vectors can be resource-intensive. Additionally, ensuring the alignment of optical flow information with RGB frames requires careful integration, adding another layer of complexity to the model design. The computational challenges associated with optical flow need to be carefully managed to strike a balance between accuracy and computational efficiency. The integration of RNNs, 2D CNNs, and optical flow in violence detection presents a robust and comprehensive approach. While offering notable advantages in discerning both spatial and temporal features, the complexity of the model demands careful consideration of computational requirements and training times, emphasising the importance of finding a pragmatic balance for real-world applications.

Tuna Swarm Algorithm With Deep Learning Enabled Violence Detection in Smart Video Surveillance Systems

Aldehim, Ghadah & Asiri, Mashael & Aljebreen, Mohammed & Mohamed, Abdullah & Assiri, Mohammed & Ibrahim, Sara. (2023). Tuna Swarm Algorithm With Deep Learning Enabled Violence Detection in Smart Video Surveillance Systems. IEEE Access. PP. 1-1. 10.1109/ACCESS.2023.3310885. [2]

In the realm of violence detection models, there exists a recognized need for improved classification performance, particularly emphasising the optimization of the hyperparameter tuning process, as it significantly influences detection outcomes. The author introduces the Tuna Swarm Optimization (TSO) algorithm to address this critical aspect and enhance the overall performance of deep learning models dedicated to violence detection. The authors acknowledge the prevalence of deep learning models in violence detection but highlight a common limitation—the lack of emphasis on hyperparameter tuning. This oversight can impact the effectiveness of these models. To address this gap, the TSO algorithm is introduced, showcasing its potential to systematically fine-tune hyperparameters and elevate the overall accuracy of violence detection models. Within the broader landscape, the study references various existing techniques, including the utilisation of Convolutional Neural Network (CNN) models such as MobileNet, GoogleNet, AlexNet, and VGG-16 for violence detection. Transfer learning and deep representative techniques are also noted, demonstrating the diversity of methodologies applied to tackle the challenges of violence detection. Additionally, the author makes references to methods incorporating spatio-temporal autocorrelations, deep NeuralNet approaches, and statistical feature descriptors, highlighting the multifaceted nature of research in this domain. Central to the author's contribution is the introduction of the TSODL-VD technique, a novel approach that amalgamates distinct components for effective violence detection. The residual-DenseNet model is selected for feature extraction, leveraging residual connections and densely connected blocks to enhance the model's capacity to capture intricate spatial features in video data. Complementing the feature extraction process, the model incorporates a stacked autoencoder (SAE) classifier. SAEs are known for their ability to learn hierarchical representations of data, contributing to the model's understanding of complex patterns and facilitating more precise classification. The integration of SAE adds a layer of sophistication to the architecture, enhancing the discriminative power of the overall system. Integral to the TSODL-VD technique is the incorporation of the Tuna Swarm Optimization (TSO) algorithm for hyperparameter tuning. Inspired by the swarming behaviour of tuna, TSO efficiently explores the hyperparameter space, aiming to find optimal configurations that enhance the model's performance. By addressing the often-neglected aspect of hyperparameter tuning, the TSODL-VD technique seeks to improve both the precision and speed of violence detection outcomes. Advantages of the

TSODL-VD technique lie in its unique combination of the residual-DenseNet model and SAE classifier, contributing to robust feature extraction and hierarchical pattern recognition. The inclusion of the TSO algorithm for hyperparameter tuning adds a layer of optimization that is often lacking in existing models, potentially leading to improved detection performance. Challenges may arise in terms of computational demands, as the combination of sophisticated feature extraction models and optimization algorithms could require significant resources. Additionally, the success of the TSODL-VD technique may depend on the appropriateness of hyperparameter choices, underscoring the importance of careful experimentation during the tuning process. The TSODL-VD technique presents a unique combination of the residual-DenseNet model, SAE classifier, and TSO algorithm, offering advantages in feature extraction, hierarchical pattern recognition, and optimised hyperparameter tuning. However, challenges related to computational demands and hyperparameter choices underscore the importance of a thoughtful and resource-aware approach in implementing this technique for violence detection.

Deep Learning for Automatic Violence Detection: Tests on the AIRTLab Dataset

P. Sernani, N. Falcionelli, S. Tomassini, P. Contardo and A. F. Dragoni, "Deep Learning for Automatic Violence Detection: Tests on the AIRTLab Dataset," in IEEE Access, vol. 9, pp. 160580-160595, 2021, doi: 10.1109/ACCESS.2021.3131315.[3]

Deep learning-based architectures, particularly 3D Convolutional Neural Networks (CNNs), have demonstrated remarkable effectiveness in the domain of violence detection within video content. The distinctive feature of 3D CNNs lies in their ability to capture both spatial and temporal information simultaneously. In the context of violence detection, this capability is crucial as it enables the model to discern not only the spatial characteristics of a scene but also the dynamic temporal patterns associated with violent actions. This is particularly relevant since violent events often unfold over time. A fundamental principle underlying 3D CNNs is their reception of inputs from a set of units located in a small neighbourhood in the preceding layer. This mechanism allows the model to extract spatio-temporal features from videos efficiently. By considering both the spatial and temporal dimensions, 3D CNNs excel at learning intricate patterns and nuances

within video sequences, contributing to their effectiveness in violence detection tasks. The utilisation of the C3D model as a feature extractor further exemplifies the potency of 3D CNNs in violence detection. Trained on the Sports-1M dataset, the C3D model has already learned to capture complex spatio-temporal features from a diverse range of video content. Leveraging pre-trained models such as C3D can significantly enhance the model's ability to generalise and detect violent actions in novel video datasets. In addition to 3D CNNs, Convolutional Long Short-Term Memory networks (ConvLSTMs) are introduced as a key component to represent spatio-temporal features. Combining the strengths of LSTM architecture with convolutional structures, ConvLSTMs are adept at capturing long-range dependencies and sequential patterns in video data. This is particularly valuable in violence detection where understanding the temporal evolution of events is crucial. ConvLSTMs enhance the model's capability to discern not only spatial features but also the evolving dynamics over time. Deep learning models, including those employing 3D CNNs and ConvLSTMs, have demonstrated state-of-the-art performance in violence detection tasks. Notably, datasets such as Hockey Fight and Crowd Violence have served as benchmarks, showcasing the top-level capabilities of these models. The ability to outperform traditional methods on these datasets underscores the effectiveness of deep learning approaches in tackling the complexities inherent in real-world violence detection scenarios. The proposed models presented in this work are built upon a foundation laid by previous research and literature in the domains of violence detection and action recognition. This signifies a progressive evolution in leveraging existing knowledge to develop more sophisticated and accurate models for detecting violent actions in videos. The incorporation of insights from previous studies reflects a nuanced understanding of the challenges posed by violence detection tasks and the need for continuous improvement in model architectures. While the discussed deep learning-based architectures offer significant advantages, they are not without challenges. One notable advantage is their ability to capture complex spatio-temporal features, enabling precise violence detection. However, the computational demands associated with training and deploying deep learning models, especially those involving 3D CNNs and ConvLSTMs, can be substantial. Additionally, the interpretability of these models remains a challenge, making it harder to understand the decision-making processes and potentially limiting their applicability in certain contexts.

Violence Detection Based on Three-Dimensional Convolutional Neural Network with Inception-ResNet

S. Jianjie and Z. Weijun, "Violence Detection Based on Three-Dimensional Convolutional Neural Network with Inception-ResNet," 2020 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS), Shenyang, China, 2020, pp. 145-150, doi: 10.1109/TOCS50858.2020.9339755.[4]

The proposed network architecture in the study is built upon the integration of the C3D model and a fusion of the Inception-Resnet-v2 network's residual Inception module. This architectural amalgamation aims to enhance the model's capacity to extract spatiotemporal features from videos depicting violent behaviour. The C3D model, known for its proficiency in capturing both spatial and temporal information simultaneously, is complemented by the residual Inception module, leveraging the strengths of the Inception-Resnet-v2 network. This fusion is designed to provide a more comprehensive understanding of complex patterns within video sequences. To address potential overfitting concerns, the paper adopts a Fine-tune strategy and utilises pre-trained models to extract shallow basic features. Fine-tuning involves adjusting the parameters of the pre-trained models, ensuring that the network can adapt to the specific characteristics of the dataset at hand. This approach is particularly valuable when dealing with small datasets, as it helps prevent overfitting, where the model may become too specialised to the training data and struggle to generalise to new examples. The experimental evaluation is conducted on two distinct datasets: the UCF-101 dataset and the HockeyFights dataset. The UCF-101 dataset is a widely used benchmark for action recognition, while the HockeyFights dataset specifically focuses on violent and non-violent behaviours, categorising videos into Fight and NoFight classes. The choice of these datasets allows for a robust assessment of the proposed model's performance in different scenarios. The key contribution of the study is the development of an improved Inception-Resnet-C3D model, which demonstrates enhanced accuracy compared to both the C3D and R3D networks. This improvement signifies the efficacy of the proposed architectural fusion in capturing nuanced spatiotemporal features relevant to violent behaviour detection. The model is specifically fine-tuned to excel in accurately annotating and extracting fragments of violent behaviour from lengthy video

sequences, showcasing its potential utility in real-world applications. The author's delves into the impact of sliding window sizes on both accuracy and processing speed. The exploration of various window sizes provides insights into the trade-offs between precision and computational efficiency. This aspect is crucial in optimising the model for real-time or near-real-time applications, where processing speed is a critical consideration. Advantages of the proposed model lie in its ability to leverage the strengths of both the C3D model and the Inception-Resnet-v2 network, enhancing the extraction of spatiotemporal features critical for violence detection. The adoption of Fine-tune and pre-trained models contributes to preventing overfitting, especially in scenarios with limited training data. The model's capability to accurately detect and annotate violent behaviour in long video sequences showcases its potential for applications requiring the identification of specific events within extended temporal contexts. Potential disadvantages may arise in terms of computational requirements, especially if the model demands significant resources for training and inference due to its sophisticated architecture. Additionally, the effectiveness of the proposed fusion may be contingent on the availability and representativeness of the pre-trained models used for feature extraction.

Violence Detection in Videos Based on Fusing Visual and Audio Information

W. -F. Pang, Q. -H. He, Y. -j. Hu and Y. -X. Li, "Violence Detection in Videos Based on Fusing Visual and Audio Information," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 2260-2264, doi: 10.1109/ICASSP39728.2021.9413686.[5]

In the landscape of violence detection studies, this research takes a distinctive path by focusing on the fusion of visual and audio information, recognizing the importance of considering multiple modalities for a more comprehensive understanding of complex scenarios. The motivation behind this approach stems from the limitations observed in early attempts at multimodal violence detection, where the use of small-scale datasets and simplistic hand-crafted features often resulted in models with compromised generalisation capabilities and stability. Recent advancements in violence detection have predominantly leaned towards leveraging deep learning models, with an emphasis on visual information extracted from RGB data and optical flow. However, this author's go beyond the preva-

lent trend by proposing a neural network architecture that integrates both visual and audio information, acknowledging the complementary nature of these modalities in capturing the intricacies of violent behaviour. The neural network architecture introduced in this study features three crucial modules designed to facilitate the fusion of visual and audio information. The attention module serves as the model's selective focus mechanism, enabling it to highlight specific regions or features within the input data that are deemed essential for violence detection. This attentional mechanism enhances the discernment of pertinent information amidst the complexity of multimodal inputs. The fusion module, another integral component, is responsible for combining visual and audio information in a cohesive manner. This fusion process ensures that both modalities contribute synergistically to the violence detection task, creating a holistic representation of the input data. By harnessing the strengths of both visual and audio features, the fusion module enhances the model's overall discriminative power. The mutual learning module introduces a collaborative aspect to the training process, allowing the neural network to iteratively refine its understanding of multimodal inputs. This iterative refinement contributes to the model's adaptability and robustness, ensuring its effectiveness in diverse and dynamic scenarios where violent behaviours may manifest in varied visual and audio patterns. While the proposed neural network offers several advantages, including its comprehensive approach to violence detection and the ability to selectively focus on crucial aspects of input data, potential challenges exist. Careful consideration is required in the design and training of the attention, fusion, and mutual learning modules to strike a balance between effectiveness and computational efficiency. Additionally, the success of the proposed neural network could be contingent on the availability of sufficiently diverse and representative multimodal datasets to ensure robust training and generalisation.

Chapter 3

System Architecture

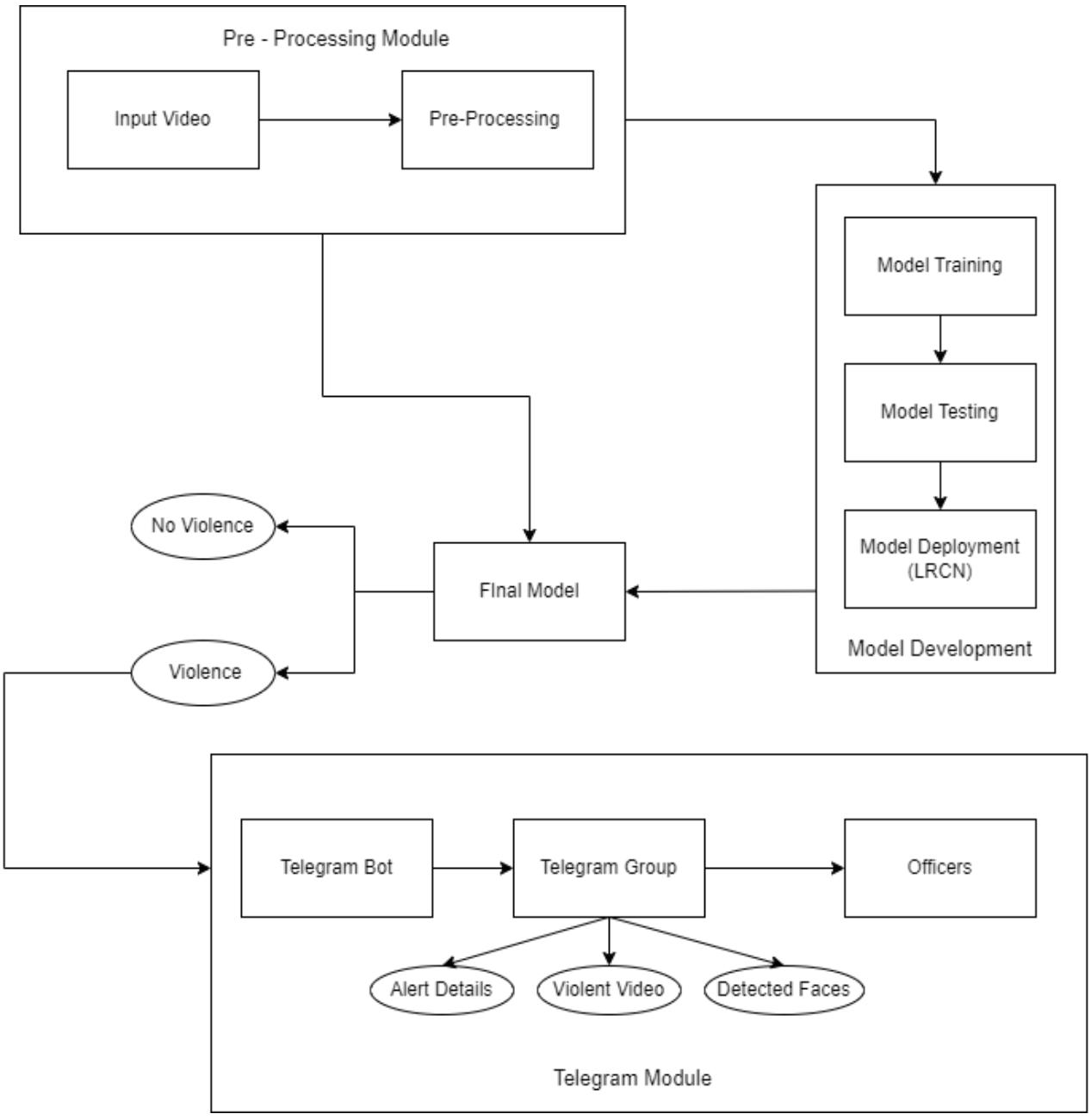


Figure 3.1: System Architecture

Figure 3.1 depicts the system architecture for video-based violence detection which is designed for seamless operation and rapid response. It starts with capturing live footage and preprocessing it is to ensure data quality. The model development phase involves training, testing, and deploying the LRCN model for accurate violence classification. Simultaneously, processed videos are classified as violent or non-violent using the trained model. In case of violence detection, alerts with detailed information and video clips are sent via Telegram, enabling quick actions by authorized personnel. This architecture ensures efficient processing, accurate detection, and swift response for maintaining safety in institutional environments.

3.1 Pre-Processing Module

The Pre-processing module initiates the data refinement process, ensuring that surveillance camera footage is standardized and optimized for subsequent analysis. This module handles formatting, stabilization, noise reduction, and frame optimization tasks, setting the stage for accurate violence detection within the system architecture.

3.1.1 Input Video Component

Figure 3.2: Sample Input Video Screenshot

Figure 3.2 depicts a sample screenshot of the input video stream utilized in this Violence Detection System. The screenshot captures a scene with multiple individuals engaged in various activities within a confined space. The video frame is clear and well-lit, allowing for detailed analysis of facial expressions and body movements.

The Input Video component plays a crucial role as the gateway for video data into the system. It handles a variety of video inputs, ranging from pre-recorded clips to live streams from surveillance cameras placed strategically within correctional facilities and mental health institutions. These cameras capture real-time footage of diverse activities within the monitored areas, such as inmate interactions, patient movements, and general facility operations. Upon receiving these video inputs, the Input Video component ensures their seamless integration into the system's workflow. It manages the transfer and storage of video data, maintaining the integrity and continuity of the streaming process. This component is designed to handle different video formats and resolutions, accommodating the diverse sources of video content encountered in real-world surveillance environments.

Additionally, the Input Video component facilitates the synchronization and coordination of multiple video streams if the system operates with multiple surveillance cameras. It manages the timestamps and metadata associated with each video input, enabling accurate temporal alignment during pre-processing and subsequent analysis stages.

Therefore, the Input Video component serves as the foundational layer for video processing within the system, handling various video inputs from surveillance cameras to ensure a continuous, synchronized, and reliable flow of video data for further processing and analysis.

3.1.2 Pre-Processing Component

The Pre-Processing component is a critical stage in your system architecture, responsible for preparing the video data before it undergoes analysis by the LRCN model for violence detection. This component encompasses several essential steps that ensure the video data is optimized and structured for effective model input and subsequent analysis. Here's a detailed breakdown of these pre-processing steps:

- Frames Extraction and Resizing:
  - Frame Extraction: This initial step involves accessing the video data and extracting individual frames using a Video Capture object. Each frame represents a snapshot of the video at a specific time instance.
  - Total Frame Count: Understanding the total number of frames in the video is crucial for effective frame sampling and sequence creation. It provides insight

into the temporal duration and granularity of the video data.

  - Resizing Operation: Resizing the frames to a standardized dimension of 64x64 pixels ensures uniformity in spatial representation across frames. This uniformity simplifies computational operations and enhances the model's ability to detect patterns consistently.

- Normalization:

  - Pixel Intensity Scaling: Normalization involves scaling down the pixel intensities of the resized frames. By dividing each pixel value by 255, the pixel intensity values are normalized to a range between 0 and 1. This normalization process minimizes the impact of varying pixel intensities across frames and videos.

  - Benefits of Normalization: Normalization not only standardizes the input data but also improves the convergence and stability of the deep learning model during training. It prevents issues such as vanishing or exploding gradients, common challenges in neural network training.

- Frame Sequence Formation:

  - Temporal Context Establishment: The formation of frame sequences is essential for capturing temporal dynamics and motion patterns within the video. By organizing frames into sequences, the model gains a temporal context, allowing it to analyze actions and behaviors over time.

  - Sequence Length Consideration: The chosen sequence length of 20 frames strikes a balance between capturing short-term actions and longer-term behaviors. This consideration ensures that the model receives sufficient temporal information for accurate violence detection.

- Data Integrity and Release:

  - Quality Assurance Checks: Throughout the pre-processing phase, rigorous checks are implemented to maintain data integrity. These checks include verifying successful frame extraction, ensuring accurate resizing and normalization, and validating the temporal consistency of frame sequences.

– Pre-Processed Data Management: The pre-processed data, organized into coherent frame sequences, is meticulously managed and prepared for input into the subsequent Model Development Module. This data management strategy ensures that the input data meets the quality standards required for effective model training and analysis.

3.2 LRCN Model Development

The core of the system lies in a powerful deep learning model trained on a massive dataset of labelled video clips. Each clip is categorised as either violent or non-violent, allowing the model to learn patterns and characteristics that differentiate these categories. The model utilises this knowledge to analyse the pre-processed video feed in real-time, continuously identifying potential occurrences of violence.

3.2.1 Model Training

The Model Training component within the Model Development Module is a pivotal phase in the development of the LRCN (Long-term Recurrent Convolutional Network) model for violence detection in correctional facilities and mental health institutions. This component encapsulates the intricate process of training the model to discern patterns indicative of violent behaviors within video streams, enabling real-time detection and intervention. The training journey commences with the acquisition and curation of a comprehensive dataset comprising video frames depicting various activities within the monitored environments. This dataset is meticulously labeled to distinguish between instances of violent and non-violent behaviors, forming the foundational corpus for model training. The dataset is then partitioned into distinct training and validation sets, facilitating effective learning while ensuring the model's ability to generalize to unseen scenarios.

During the training process, batches of video frames are fed into the LRCN model, which comprises convolutional layers for spatial feature extraction and recurrent layers (e.g., LSTM cells) for capturing temporal dependencies. These architectural components synergize to empower the model with the capability to understand both the content and context of video sequences, crucial for accurate violence detection. Optimization algorithms are employed iteratively to fine-tune the model's parameters, minimizing loss and

maximizing accuracy. Continuous monitoring of key metrics, such as accuracy and loss, guides the training process, ensuring the model learns effectively without overfitting to the training data. Techniques like regularization and dropout may be employed to enhance model generalization and mitigate overfitting risks. Upon achieving satisfactory performance on the validation set, the trained LRCN model emerges as a robust tool for real-time violence detection, poised for deployment within the violence detection system. Regular updates and retraining strategies are integral to maintaining the model's efficacy, adapting it to evolving behavioral patterns and ensuring continual performance optimization in dynamic correctional and mental health environments.

3.2.2 Model Testing

Testing the Long-term Recurrent Convolutional Network (LRCN) model constitutes a crucial phase in validating its efficacy within the violence detection system tailored for correctional facilities and mental health institutions. This testing process is pivotal in assessing the model's accuracy, reliability, and real-world applicability in identifying instances of violence from video streams captured by surveillance cameras. In the testing phase, a diverse and representative dataset is utilized to evaluate the LRCN model's performance across various scenarios encountered in real-world environments. This dataset encompasses a spectrum of lighting conditions, inmate behaviors, and spatial configurations, enabling a comprehensive assessment of the model's robustness and generalization capabilities. Through rigorous testing with diverse inputs, potential limitations and areas for enhancement can be identified and addressed.

The LRCN model undergoes thorough validation and evaluation against a labeled dataset where ground truth labels for violent instances are known. Performance metrics such as precision, recall. Precision assesses the accuracy of positive predictions, recall measures the model's ability to identify all relevant instances of violence. These metrics offer valuable insights into the model's ability to minimize false positives and false negatives, crucial aspects in the context of a violence detection system.

To further bolster the model's reliability, testing extends to unseen or real-time data scenarios. This phase involves deploying the LRCN model within the actual surveillance infrastructure, allowing it to analyze live video feeds and detect violence in a dynamic setting. Continuous monitoring and feedback mechanisms are established to capture

any anomalies, false alarms, or missed detections, enabling iterative refinements and improvements to the model's performance over time. In addition to quantitative metrics, qualitative assessment by domain experts, security personnel, incharge officers, and other stakeholders plays a vital role in the comprehensive evaluation of the LRCN model. Their feedback and insights provide valuable perspectives on the model's practical effectiveness, usability, and reliability within the complex and evolving environment of correctional facilities and mental health institutions. This holistic testing approach ensures that the LRCN model not only meets technical benchmarks but also excels in real-world deployment, contributing significantly to enhanced safety and security measures.

3.2.3 Long-term Recurrent Convolutional Networks (LRCN)

Long-term Recurrent Convolutional Networks (LRCN) represent a sophisticated fusion of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), designed to tackle complex tasks involving sequential data analysis, particularly in the realm of video processing. Unlike traditional CNNs that excel in static image classification, LRCN models are uniquely adept at understanding temporal dependencies and long-term patterns within video sequences. The LRCN architecture inherits the powerful spatial feature extraction capabilities of CNNs, allowing it to dissect visual information in each frame of a video. This is complemented by the temporal understanding provided by recurrent layers, which enable the model to discern the sequential evolution of events over time. By combining these strengths, LRCN models can effectively analyze and interpret dynamic visual content, making them invaluable for applications such as violence detection in correctional facilities and mental health institutions.

The main layers of LRCN are:

- Convolutional Layer
- Recurrent Layers (RNNs/LSTMs)
- Fully Connected Layer

Figure 3.3 illustrates the architecture of the Long-term Recurrent Convolutional Network (LRCN) model for violence detection in correctional facilities and mental health institutions. It showcases the sequential flow of video data processing through various

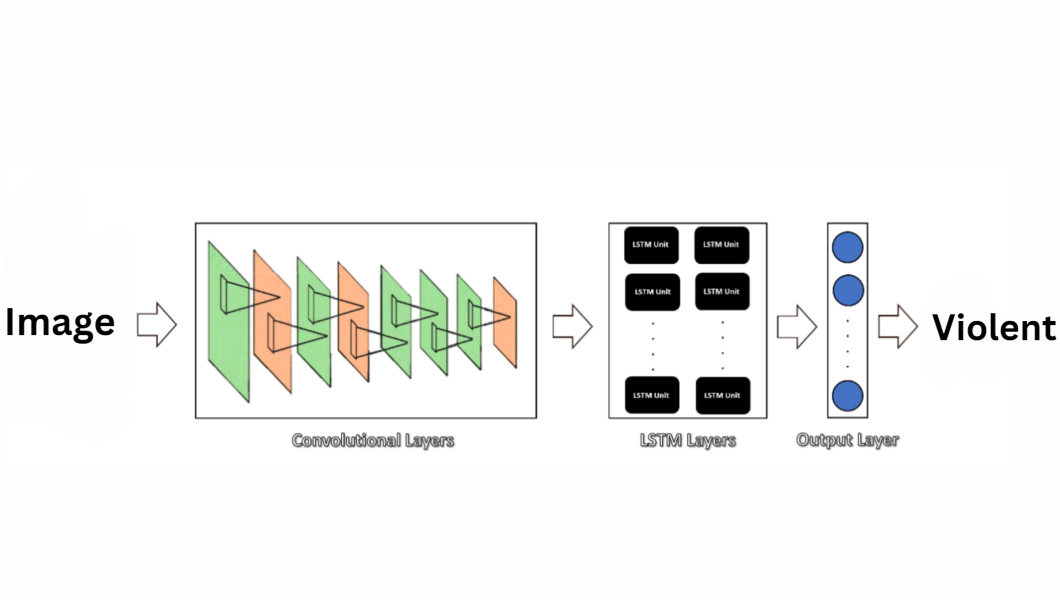


Figure 3.3: Architecture of LRCN

components, including input frames, convolutional layers for feature extraction, recurrent layers for temporal pattern recognition, fully connected layers for integration, and the output layer for violence detection decisions. This visual representation highlights the model's ability to combine spatial and temporal processing, crucial for analyzing video sequences and identifying instances of violence in real-time scenarios.

Convolutional Layer

The Convolutional Layer in a Long-term Recurrent Convolutional Network (LRCN) plays a foundational role in extracting spatial features from individual frames within a video sequence. This layer is responsible for capturing patterns, edges, textures, and other visual details that are crucial for understanding the content of each frame. At its core, the convolutional layer consists of a set of filters or kernels that slide across the input image or frame, performing mathematical operations to extract relevant features. These filters act as feature detectors, highlighting specific aspects of the input data that are essential for subsequent analysis. The process within the convolutional layer involves convolving the input image with the filters to produce feature maps. Each filter focuses on detecting a particular aspect of the image, such as edges, corners, or textures. By applying multiple filters, the convolutional layer generates multiple feature maps, each capturing different aspects of the input.

Additionally, the convolutional layer incorporates activation functions, such as ReLU (Rectified Linear Unit), to introduce non-linearity and enable the network to learn complex representations. This non-linearity enhances the model's ability to capture abstract features and patterns that contribute to higher-level visual understanding.

- The Rectified Linear Unit (ReLU) stands as a widely adopted activation function in the realm of neural networks, frequently employed, including in LRCN. Its mathematical formulation is expressed as:

$$f(x) = \max(0, x)$$

  Here, x represents the input to the function.

  This foundational mathematical function is integral to the operations of neural networks, providing a straightforward yet impactful operational principle. The essence of ReLU lies in its output behaviour—it leaves positive input values unchanged while returning zero for non-positive inputs. This simple mechanism introduces a crucial non-linear element into the network's calculations, thereby enhancing its ability to discern intricate patterns and relationships within the data. The significance of ReLU extends beyond its simplicity. One key challenge in training neural networks is the vanishing gradient problem during backpropagation, where the gradients become exceedingly small, hindering effective learning. The introduction of ReLU addresses this issue by ensuring that positive gradients are maintained, preventing the loss of important information during the learning process. The operational principle of ReLU involves maintaining positive values and setting negative values to zero, a seemingly basic concept that plays a pivotal role in the network's ability to capture non-linear relationships in data. The widespread use of ReLU across neural networks, including LRCNs, underscores its effectiveness in promoting efficient and meaningful learning.

Furthermore, the convolutional layer often includes pooling operations, such as max pooling or average pooling, to down-sample the feature maps. This downsampling reduces computational complexity while retaining essential spatial information, making the network more efficient and robust. Overall, the convolutional layer in an LRCN model acts

as a feature extractor, transforming raw pixel values into meaningful representations that form the basis for subsequent analysis and decision-making within the video processing pipeline.

Recurrent Layers (RNNs/LSTMs)

The Recurrent Layers, specifically Long Short-Term Memory (LSTM) units, in a Long-term Recurrent Convolutional Network (LRCN) contribute significantly to temporal understanding and context preservation within video sequences. Unlike traditional Convolutional Neural Networks (CNNs) that excel in spatial feature extraction, recurrent layers focus on capturing temporal dependencies and long-range patterns across frames in a video. The core functionality of recurrent layers revolves around their ability to retain memory of past information while processing current inputs. This memory retention is crucial for understanding sequential data, such as video frames where the order of frames holds meaningful context.

Within the LSTM units, several key mechanisms operate:

- Cell State: The cell state in an LSTM unit acts as the long-term memory of the network. It runs through the entire sequence of data and allows information to flow across time steps. This continuous flow of information enables the network to retain important context and dependencies over extended periods, making it well-suited for tasks involving sequential data analysis, such as video processing.

- Forget Gate: The forget gate in an LSTM unit plays a crucial role in memory management. It evaluates the relevance of information stored in the cell state from the previous time step and decides which information to discard or forget. By selectively forgetting less relevant information, the network can focus its attention on the most important features for the current context. This mechanism helps prevent the accumulation of irrelevant data in the long-term memory, improving the network's ability to extract meaningful patterns.

- Input Gate: The input gate of an LSTM unit regulates the update of the cell state by incorporating new information from the current input. It evaluates the importance of incoming data and determines how much of it should be added to

the cell state. This gate allows the network to adapt its long-term memory based on the significance of new information, ensuring that relevant features are retained and integrated into the memory for future predictions or analysis.

- Output Gate: The output gate controls the flow of information from the cell state to the output of the LSTM unit. It determines which parts of the cell state should be used to compute the output at the current time step. By selectively choosing relevant information from the long-term memory, the output gate influences the network's predictions or actions based on the current input and context. This gating mechanism enables the network to produce accurate and contextually relevant outputs while considering long-term dependencies.

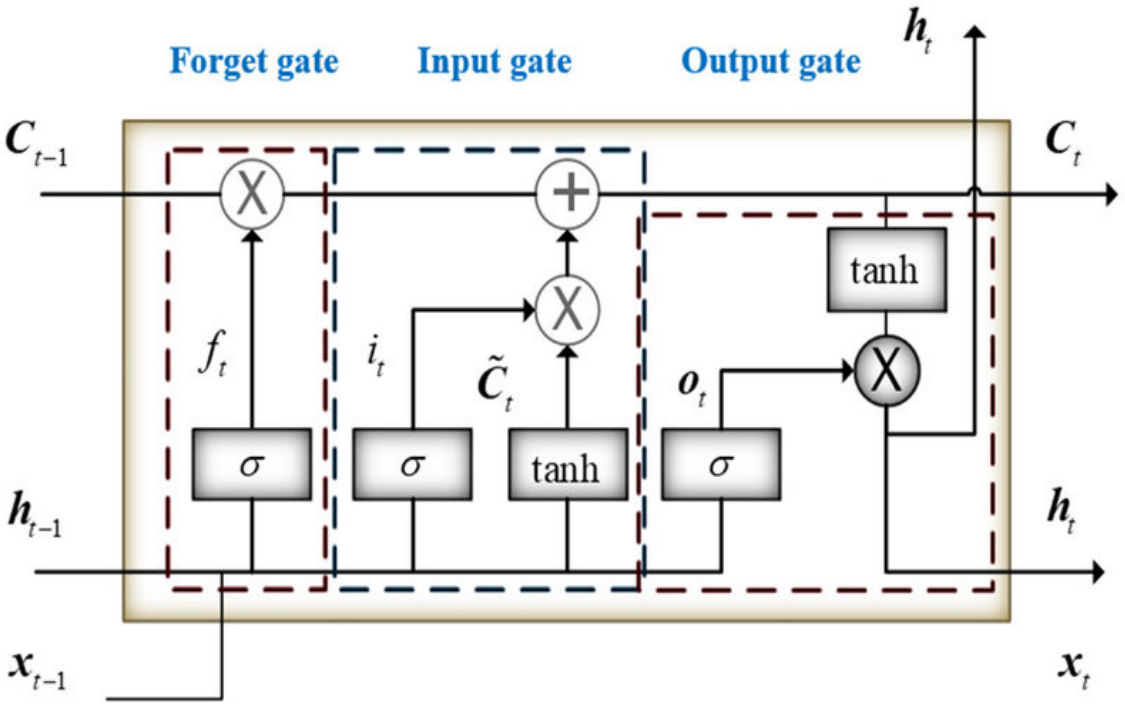


Figure 3.4: LSTM Unit

Figure 3.4 illustrates a Long Short-Term Memory (LSTM) unit, a building block of Recurrent Neural Networks (RNNs) adept at handling sequential data. Unlike standard RNNs, LSTMs overcome the vanishing gradient problem by managing information flow through specialized gates. These gates, including forget gates, input gates, and output gates, control the flow of information into, within, and out of the LSTM unit's cell state, essentially its long-term memory. [6]

These mechanisms collectively enable LSTMs to capture temporal dependencies by learning when to remember, forget, and update information over time.

In the context of LRCN, recurrent layers receive feature maps extracted by the convolutional layers and process them sequentially, considering both spatial and temporal aspects. This integration of CNNs with recurrent layers allows the network to not only capture spatial features within individual frames but also comprehend the dynamics and context of actions across multiple frames.

The recurrent layers, particularly LSTMs, enhance the network's capability to model long-term dependencies, detect temporal patterns, and preserve context throughout the video sequence. This temporal understanding is vital for tasks like action recognition, event detection, and context-aware video analysis, making recurrent layers a fundamental component of LRCN models.

Fully Connected Layer

The Fully Connected Layers in a Long-term Recurrent Convolutional Network (LRCN) play a pivotal role in processing features extracted by convolutional and recurrent layers, ultimately enabling the network to make predictions or classifications based on the input data.

The importance of fully connected layers can be summarized as follows:

- Integration of Features: The Fully Connected Layers serve as a bridge between the extracted features from convolutional and recurrent layers and the final output of the network. These layers consolidate the spatial and temporal information learned from previous layers into a comprehensive feature representation. This integration process is critical for capturing complex patterns and relationships within the data, essential for tasks like violence detection in video streams.

- Dense Connectivity: Unlike convolutional and recurrent layers that operate on localized spatial or temporal regions, Fully Connected Layers establish connections between every neuron in one layer to every neuron in the next layer. This dense connectivity allows the network to consider global information and correlations across all features, enhancing its ability to discern nuanced patterns and make informed decisions.

- Non-linearity and Activation Functions: Each neuron in a Fully Connected Layer applies a non-linear activation function to the weighted sum of inputs received from the previous layer. Common activation functions like ReLU (Rectified Linear Unit) introduce non-linearity into the network, enabling it to learn complex mappings between input and output spaces. This non-linearity is crucial for capturing the intricate relationships present in video data, especially when dealing with dynamic and evolving scenes.

- Parameter Learning: The Fully Connected Layers contain learnable parameters, including weights and biases, which are optimized during the training phase. Through backpropagation and gradient descent algorithms, the network adjusts these parameters to minimize prediction errors and improve its ability to generalize to unseen data. This parameter learning process fine-tunes the feature representation learned by earlier layers, optimizing the network's performance in violence detection and related tasks.

- Output Generation: The final Fully Connected Layer typically produces the network's output, which could be binary (e.g., classifying a video as violent or non-violent) or multi-class (e.g., categorizing different types of violent behaviors). The activation function applied to this layer, such as softmax for multi-class classification or sigmoid for binary classification, determines the format and interpretation of the network's predictions.

  In essence, the Fully Connected Layers in an LRCN model serve as a crucial component for synthesizing complex features, capturing long-term dependencies, introducing non-linearity, learning predictive patterns, and generating meaningful outputs, all of which contribute to the network's efficacy in violence detection and analysis tasks within correctional facilities and mental health institutions.

3.3 Model Deployment and Classification

The deployment of the LRCN model involves a dynamic process that harmonizes the strengths of convolutional and recurrent layers to create a robust violence detection system. As surveillance cameras capture video feeds, these inputs are fed into the LRCN

model, which begins by extracting spatial features through its convolutional layers. These layers excel at identifying patterns and visual cues within frames, such as aggressive gestures or physical altercations. Simultaneously, the recurrent layers of the model come into play, enabling the network to analyze temporal sequences across frames. This temporal analysis is crucial for understanding the progression of events, recognizing patterns of escalation, and distinguishing between normal interactions and potentially violent behaviors. The recurrent layers, including LSTM units, maintain context and memory over time, allowing the model to track the evolution of activities within the monitored environment.

The classification process within the LRCN model involves interpreting the extracted features to make decisions about the nature of the observed behaviors. This process may employ sophisticated algorithms that consider the temporal dynamics of events, the intensity of actions, and contextual cues to determine the likelihood of violence. Classification outcomes are often probabilistic, providing nuanced insights into the model's confidence levels in identifying threats. Upon identifying a potential violent incident, the model triggers a series of response mechanisms. These may include immediate alerts to security personnel through mobile applications or centralized monitoring systems. Additionally, the model may initiate recording of the incident, capturing critical footage for subsequent analysis and evidence gathering.

The deployment of the LRCN model is not static but dynamic, continually learning and adapting to new patterns and evolving threats. Regular updates, retraining with fresh data, and feedback mechanisms ensure that the model remains effective and responsive in real-world scenarios, contributing significantly to the safety and security objectives of correctional facilities and mental health institutions.

3.4 Telegram Module

The Telegram module is a pivotal component of the violence detection system, designed to facilitate swift and effective communication in response to potential threats within correctional facilities and mental health institutions. Its primary function is to ensure that security personnel and relevant stakeholders receive timely alerts and updates regarding detected violent activities or identified individuals. By leveraging the capabilities of the Telegram messaging platform, the module enables real-time communication and collabo-

ration among security teams. It serves as a centralized hub for receiving alerts generated by the system and disseminating them to designated individuals or groups based on the nature and urgency of the alert.

The Telegram module streamlines the response process by providing organized and targeted information, such as alert details, violent video footage, and detected faces, through dedicated channels within the Telegram platform. This structure allows security personnel to focus on specific aspects of the alert and take appropriate actions promptly. The module's integration with Telegram enhances situational awareness, facilitates proactive measures, and supports efficient decision-making in addressing potential security threats. Its seamless communication capabilities and group collaboration features contribute significantly to maintaining a secure and vigilant environment within correctional and mental health facilities.

The Telegram module comprises various integral components, contributing to its robust functionality and seamless integration within the violence detection system:

- Telegram Bot
- Telegram Group
- Officers

3.4.1 Telegram Bot

The Telegram Bot within the violence detection system plays a crucial role as a communication conduit, providing instant alerts and notifications based on live video analysis. It goes beyond simple message delivery, offering intelligent interactions and quick response mechanisms vital for effective security management in correctional facilities and mental health institutions. Fundamentally, the Telegram Bot is designed to swiftly deliver actionable intelligence by relaying information about detected violent incidents directly to designated personnel and response teams. This rapid response capability is essential in situations where immediate action can prevent further escalation and ensure the safety of inmates, patients, and staff.

The Bot operates on predefined triggers and commands, initiating actions when the violence detection system identifies a potential threat. It gathers relevant data, such

as video snippets and incident details, and formats this information into alert messages dispatched to specific Telegram groups or individuals overseeing security. Its workflow follows a structured process, starting with trigger reception, data retrieval, and alert generation, culminating in the prompt delivery of alerts through Telegram's API. Users can interact with the Bot, accessing additional information or acknowledging received alerts, streamlining communication and decision-making processes.

Creating a Telegram Bot involves registration with BotFather, obtaining authentication tokens, and seamless integration into the system's backend. This integration ensures reliable message transmission and user engagement, highlighting the Bot's role in proactive security measures and swift response actions within sensitive environments.

3.4.2 Telegram Group

The Telegram Group in the violence detection system acts as a central hub for disseminating critical information and coordinating response efforts when violence is detected. It serves as a collaborative platform where security personnel, officers in charge, and relevant stakeholders can communicate, share insights, and take immediate action based on real-time alerts.

The Telegram bot transmits the following information to the group:

- Alert Details: When a violent incident is detected by the Telegram Bot, it promptly sends an alert message to the Telegram Group. This alert message contains critical information such as the timestamp of the incident, the specific location within the facility where the violence occurred, and a concise description of the observed violent activity. The purpose of this alert is to swiftly notify all group members about the potential threat and prompt immediate response actions. By providing precise details about the incident, including the time and location, security personnel can quickly assess the situation and initiate appropriate intervention measures.

- Detected Faces: In addition to the alert message, the Telegram Bot uploads images of detected faces involved in the altercation to the Telegram Group. These images provide visual confirmation of the individuals involved in the violent incident, aiding in their identification and subsequent tracking. This feature enhances the security team's ability to respond effectively and take necessary measures to address the

situation. By including images of detected faces, the Telegram Bot enables security personnel to identify potential perpetrators and apply appropriate security protocols to mitigate risks and ensure the safety of inmates, staff, and facilities.

- Violent Video Snippets: Alongside the alert details and detected faces, the Telegram Bot also uploads snippets of the detected violent activity captured from the surveillance footage. These video snippets offer real-time visual context of the incident, allowing group members to assess the severity of the situation accurately. The inclusion of video snippets enhances situational awareness and facilitates informed decision-making among security personnel and supervisory staff. By providing visual evidence of the violent activity, security teams can evaluate the extent of the incident, determine the appropriate response level, and coordinate actions to de-escalate the situation effectively.

The Telegram Group comprises designated members responsible for security management and response actions. These members typically include security officers stationed in the facility, supervisory staff, and administrative personnel overseeing security protocols. Each member receives the alert message and accompanying multimedia content simultaneously, ensuring immediate awareness and coordinated response efforts. Upon receiving the alert in the Telegram Group, members can review the provided details, assess the severity of the situation, and initiate appropriate response procedures. This may involve dispatching security personnel to the location, coordinating with law enforcement if necessary, and documenting the incident for further investigation and reporting.

The Telegram Group serves as a vital communication channel for real-time incident management, enabling rapid decision-making, collaboration among stakeholders, and effective deployment of resources to maintain safety and security within correctional facilities and mental health institutions.

3.4.3 Officers

The officers' role in the Telegram module is pivotal for swift incident response and effective security management within correctional facilities and mental health institutions. When an alert message is received in the Telegram Group, officers immediately assess the situation, verify the alert details, and determine the appropriate response. This rapid

assessment is crucial for preventing the escalation of violence and ensuring the safety of inmates, staff, and facilities.

Within the Telegram Group, officers engage in collaborative decision-making processes. They discuss the incident, share insights, and formulate response strategies collectively. This collaborative environment fosters effective communication, coordination, and decision-making among security personnel, enabling them to respond cohesively and efficiently to security threats. Furthermore, officers utilize the Telegram Group to deploy resources strategically based on the information provided in the alert message, detected faces, and video snippets. This may involve dispatching security teams, initiating lockdown procedures, or requesting additional support as needed. The centralized platform of the Telegram Group facilitates coordinated resource deployment and response efforts.

Documentation and reporting are integral aspects of the officers' role in the Telegram module. They document incident details, actions taken, and outcomes within the Telegram Group. This documentation is vital for post-incident analysis, reporting to higher authorities, and implementing preventive measures. The Telegram module provides a structured framework for documenting incident responses and maintaining an audit trail of security-related activities. Throughout the incident management process, officers provide real-time updates and status reports within the Telegram Group. These updates keep all stakeholders informed about the progress of response efforts, developments in the situation, and incident resolution. Real-time communication enhances transparency, accountability, and coordination among security personnel, contributing to a proactive and coordinated approach to security incident management.

The system architecture designed for violence detection in correctional and mental health settings is a testament to the fusion of cutting-edge technologies and strategic planning. It encompasses sophisticated algorithms, real-time communication channels, and proactive response mechanisms to ensure a comprehensive approach to safety and security. By leveraging artificial intelligence alongside human intervention, the architecture is poised to detect and prevent potential threats effectively, fostering a secure environment for all stakeholders.

Chapter 4

Methodology

4.1 UML Diagrams

4.1.1 Use Case Diagram

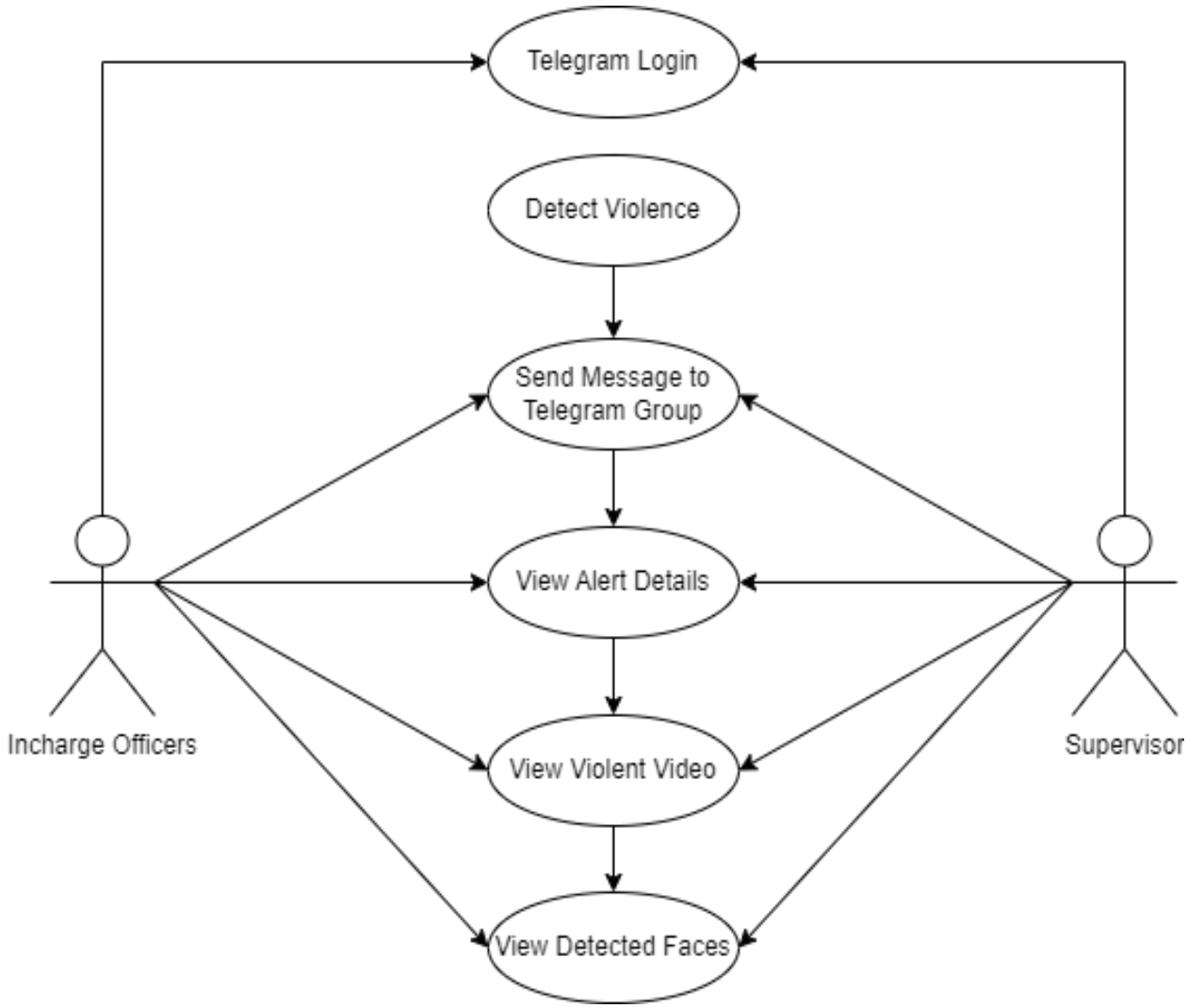


Figure 4.1: Use Case of the System

Figure 4.1 depicts the use case diagram illustrating the interactions between key actors

and the security system, including the "In-Charge Officer" and "Supervisor." Use cases involve Telegram login, alert messages sent to the Telegram group upon violence detection, viewing details of alerts, accessing violent videos and detected faces. In-Charge Officers are responsible for confirming incidents, while Supervisors manage and confirm actions. This diagram highlights a collaborative system designed to address security challenges and ensure swift responses to incidents for victim well-being.

.Here are the main elements of the diagram:

Actors

These entities encompass individuals or systems engaging with the system. In this context, the actors consist of :

- In-Charge Officer: The In-Charge Officer is the central authority responsible for confirming incidents of detected violence within correctional facilities and mental health institutions. This role encompasses critical responsibilities, including evaluating the legitimacy of security threats, guiding immediate actions in response to incidents, and ensuring the overall effectiveness of the security response. The In-Charge Officer's expertise in security protocols, risk assessment, and decision-making is crucial for maintaining a safe and secure environment for inmates, patients, staff, and visitors.

- Supervisor: The Supervisor holds a leadership role in the security system, tasked with managing and overseeing the activities of security officers. This position involves optimising resource allocation, assigning officers to investigate violence incidents, and confirming the validity of reported security threats. Additionally, the Supervisor is responsible for monitoring security protocols, conducting regular assessments to identify vulnerabilities, and implementing strategies to enhance overall security. Their proactive approach contributes significantly to the coordination, efficiency, and continuous improvement of security operations within correctional and mental health settings.

Use Cases

Use cases in a system outline specific scenarios or functionalities, illustrating how actors interact with the system to accomplish particular tasks or objectives.In this case, the use cases are:

- Telegram Login: This use case is crucial for ensuring secure access to the system for both In-Charge Officers and Supervisors. Through Telegram Login, authorised personnel can securely authenticate themselves using their credentials. This authentication process is designed to be robust and reliable, safeguarding sensitive information and ensuring that only designated personnel can access the security system. By logging in via Telegram, officers can seamlessly transition into their roles within the system, enabling them to carry out their responsibilities effectively and securely.

- Detect Violence: The "Detect Violence" use case leverages the system's advanced capabilities to identify instances of violence from video footage captured by surveillance cameras. This functionality is vital for early threat detection and prevention within correctional facilities and mental health institutions. Using sophisticated algorithms and deep learning techniques, the system analyses video feeds in real-time, identifying patterns indicative of violent behavior. This early recognition empowers security personnel to intervene swiftly, mitigating potential security threats and ensuring the safety of inmates, patients, and staff.

- Send Message to Telegram Group: The "Send Message to Telegram Group" use case enables automated alert notifications to be sent to the designated Telegram group when violence is detected. This automated messaging system ensures a rapid and coordinated response to security incidents among relevant personnel. By promptly notifying the Telegram group, officers can mobilise resources, coordinate response actions, and implement necessary security protocols to address the detected violence effectively. This streamlined communication process enhances the overall security response within the facility.

- View Alert Details: In-Charge Officers and Supervisors have access to comprehensive information through the "View Alert Details" use case. This includes specific

details such as the location of the detected violence, timestamp of the incident, and a concise description of the observed violent activity. This detailed information provides valuable context for understanding the nature and severity of the security threat, enabling officers to make informed decisions and take appropriate actions in response to the incident.

- View Violent Video: The "View Violent Video" use case allows In-Charge Officers and Supervisors to access recorded video snippets of detected violent incidents. These video snippets provide a real-time visual context of the incident, offering valuable insights into the nature of the violent behavior. By viewing the video footage, officers can accurately assess the situation, identify individuals involved, and gather additional evidence for investigation and response purposes. This visual information enhances situational awareness and facilitates effective decision-making in managing security incidents.

- View Detected Faces: Through the "View Detected Faces" use case, In-Charge Officers and Supervisors can access images of detected faces involved in the altercation. These images serve as visual confirmation of the individuals implicated in the violent incident, aiding in their identification and subsequent tracking. By viewing the detected faces, officers can gather valuable information about the individuals involved, such as their identities and potential roles in the incident. This capability enhances the investigative process and supports efforts to maintain a secure and controlled environment within the facility.

4.1.2 Sequence Diagram

Figure 4.2 below illustrates a sequence diagram that portrays a streamlined process for violence detection and response within the system architecture. It showcases interactions among key components like the Web Interface for user interaction, the LRCN Model for video analysis, the Telegram Bot for instant alerts, and In-Charge Officers for managing responses. With inputs from surveillance cameras, the LRCN Model detects violence, triggering notifications via the Telegram Bot to alert personnel. In-Charge Officers then access details and video clips to facilitate swift and coordinated actions, ensuring effective security protocols in correctional and mental health settings.

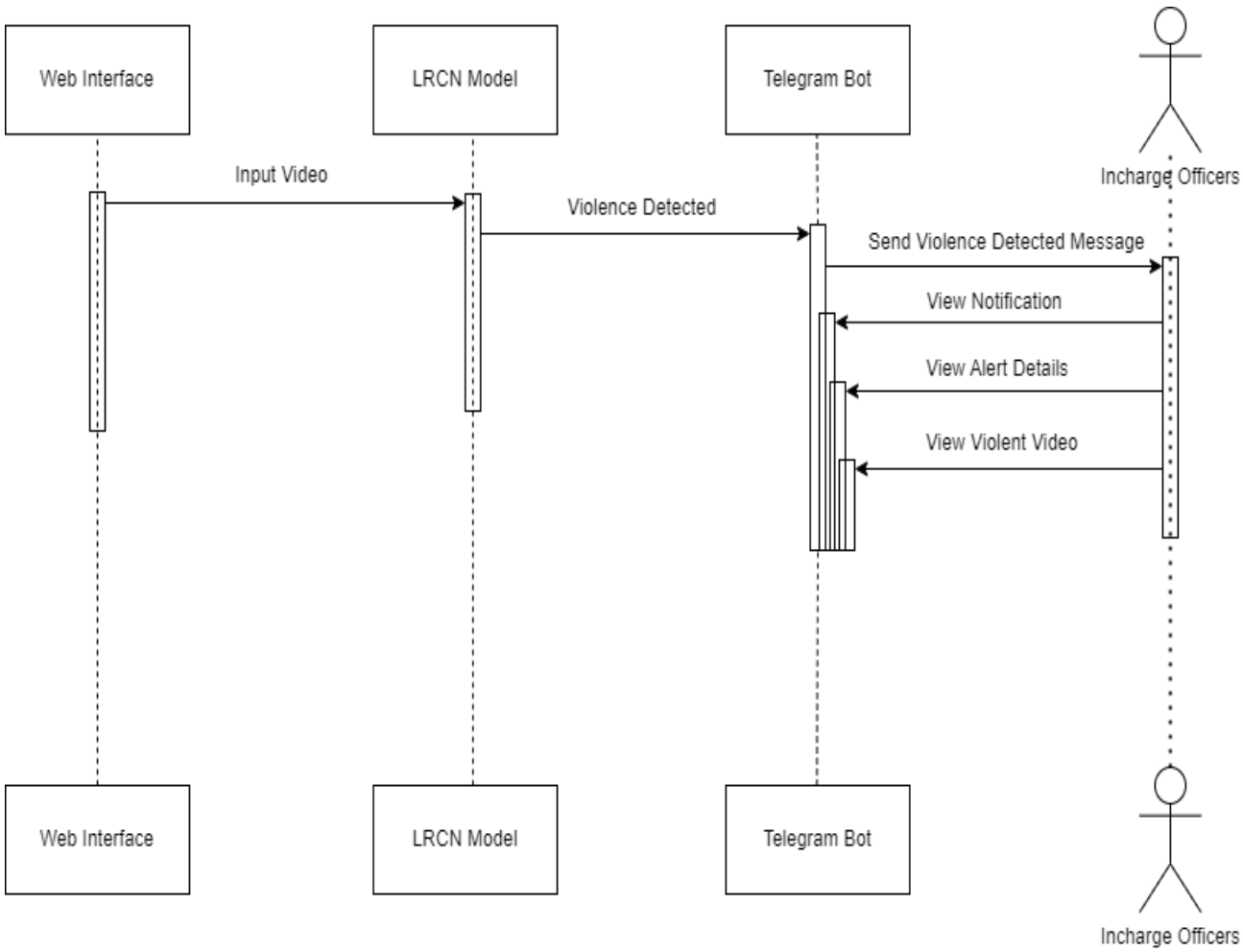


Figure 4.2: Sequence Diagram of the System

The steps involved are:

- Input Video via Web Interface: The system begins when authorized users access the web interface to upload video footage captured within correctional facilities and mental health institutions. This input video serves as the primary data source for the system's analysis.

- Transmission of Input Video to the LRCN Model: The web interface seamlessly transmits the uploaded video footage to the Long Short-Term Recurrent Convolutional Network (LRCN) Model. This transmission is crucial as it initiates the analytical process where the LRCN Model scrutinizes the video content for signs of potential violence.

- Automated Violence Detection by the LRCN Model: Once the LRCN Model receives the video data, it engages in automated violence detection algorithms. These algorithms are meticulously designed to analyze visual cues, body language, and contextual information within the video frames. The model's deep learning capabilities allow it to identify patterns associated with violent behavior.

- Signal Sent to Telegram Bot on Violence Detection: Upon detecting potential violence within the input video, the LRCN Model generates a signal indicating the presence of a security threat. This signal is then transmitted to the designated Telegram Bot, which acts as the communication bridge for alerting relevant personnel.

- Telegram Bot Sends Alert to Telegram Group: The Telegram Bot, upon receiving the violence detection signal, immediately sends an alert message to the predefined Telegram group. This alert message is crafted to convey critical information about the detected security threat, including timestamps, location details, and a brief description of the observed violent activity.

- Notification Reception by In-Charge Officers: In-Charge Officers, being part of the designated Telegram group, promptly receive the alert notification on their devices. This notification mechanism ensures that key personnel are swiftly informed about the security incident, enabling them to initiate response procedures without delay.

- View Alert Details by In-Charge Officers: In-Charge Officers have the capability to access comprehensive alert details through the Telegram interface. These details include timestamps indicating when the violence was detected, precise location information within the facility, and a concise summary of the observed violent behavior.

- View Violent Video by In-Charge Officers: In addition to alert details, In-Charge Officers can also view a specific segment of the video footage that captured the detected violence. This video clip provides visual context and clarity, aiding in the assessment of the severity and nature of the security incident.

- Confirm Violence by In-Charge Officers: In-Charge Officers meticulously review the alert details and the accompanying video clip to confirm the occurrence of violence. This confirmation step is crucial as it validates the automated detection process, ensuring that actionable alerts are based on accurate assessments.

- Take Actions Based on Confirmation: Upon confirming the presence of violence, In-Charge Officers swiftly initiate appropriate response actions. These actions may include deploying security personnel to the location, implementing security protocols, contacting emergency services if necessary, and coordinating with other relevant stakeholders to address the security threat effectively.

4.1.3 Object Diagram

Figure 4.3 below illustrates the key components of the security system within a correctional facility or mental health institution. These include the Video Footage captured by surveillance cameras, the LRCN Model for analyzing video content and detecting violence, the Telegram Bot for alerting authorities, and objects like Alert Details and Detected Violent Video for storing and processing incident data. The diagram showcases the hierarchical relationships and collaborative interactions among these entities, highlighting their roles in ensuring security and facilitating prompt responses to detected incidents.

The diagram shows the different objects, or instances of classes, in the system and how they interact with each other.

- Inmate: Inmates represent individuals within the monitored premises, such as cor-

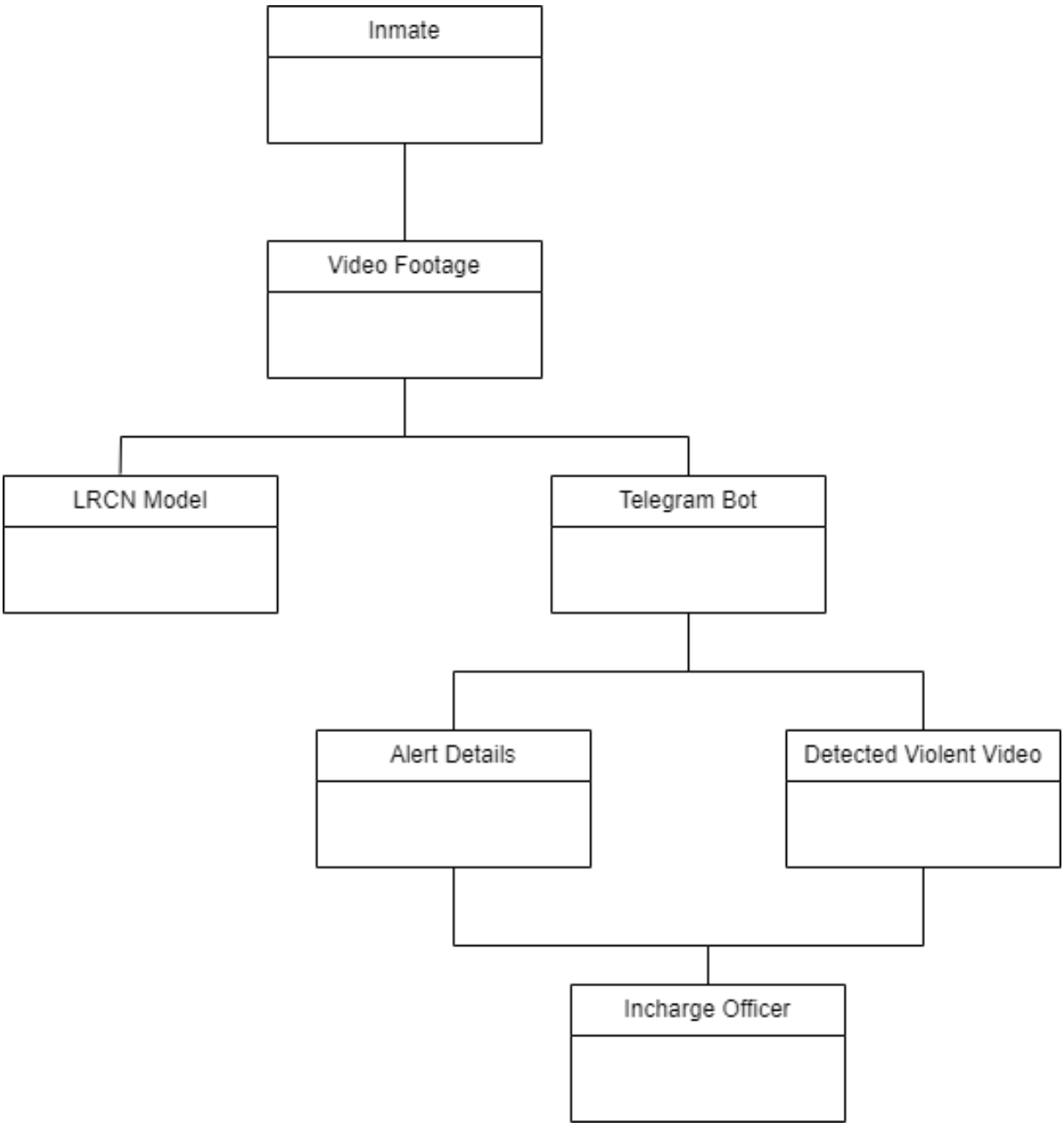


Figure 4.3: Object Diagram of the System

rectional facilities or mental health institutions. These individuals are under constant observation by the surveillance system, generating real-time behavioral data. The Inmate object encapsulates not just the physical presence of individuals but also their interactions, movements, and activities captured by surveillance cameras. This data includes patterns of behavior, social interactions, and potential indicators of disturbances or security threats.

- Video Footage: The Video Footage object embodies the continuous streams of video data captured by surveillance cameras strategically placed within the monitored areas. These streams capture a plethora of visual information, including inmate movements, interactions, group dynamics, and environmental conditions. The Video Footage serves as a comprehensive record of events within the facility, offering insights into daily routines, unusual behaviors, and incidents that may require security intervention.

- LRCN Model: The Long Short-Term Recurrent Convolutional Network (LRCN) Model is a sophisticated deep learning architecture tailored for video analysis and pattern recognition. It operates on the Video Footage, leveraging convolutional layers to extract spatial features and recurrent layers to capture temporal dynamics. The LRCN Model's role is multifaceted, encompassing violence detection, anomaly identification, and behavior analysis. It learns complex patterns, nuances, and deviations from normal activities, enabling it to flag potential security risks accurately.

- Telegram Bot: The Telegram Bot acts as the communication hub between the security system and designated personnel via the Telegram messaging platform. It receives alerts, notifications, and data from various components such as the LRCN Model, Alert Details, and Detected Violent Video. The Telegram Bot ensures seamless and instant dissemination of critical information, facilitating rapid response and decision-making by security personnel.

- Alert Details: The Alert Details object contains a wealth of information related to detected security incidents. This includes timestamps indicating when the incident occurred, specific locations within the facility where the incident took place, descriptions detailing the nature of the observed activity (e.g., physical altercations,

aggressive behavior), and any relevant contextual data (e.g., involved parties, severity level). The richness of this information aids in understanding the context and urgency of the security event.

- Detected Violent Video: Detected Violent Video refers to segmented clips or frames extracted from the Video Footage, specifically highlighting instances of violent or concerning behavior. These video segments provide visual evidence and context, allowing security personnel to assess the situation accurately. The Detected Violent Video object is instrumental in verifying alerts, corroborating information from other sources, and supporting decision-making during security incidents.

- Incharge Officer: The Incharge Officer is a pivotal role within the security hierarchy, responsible for overseeing operations, managing responses to security alerts, and coordinating resources. They receive alerts and detailed reports from the Telegram Bot, including Alert Details and Detected Violent Video. The Incharge Officer's tasks include verifying the legitimacy of alerts, determining appropriate response protocols, deploying security personnel if needed, and liaising with other stakeholders for a cohesive security approach.

4.1.4 Activity Diagram

Figure 4.4 below depicts a streamlined workflow for inmate surveillance and violence detection within the facility's security system. It begins with continuous Inmate Surveillance through cameras, followed by Violence Detection using the LRCN Model. The system then employs the Telegram Bot to promptly send alert messages to designated personnel via the Telegram platform. Upon reception in the Telegram Group, officers can view comprehensive details of the detected violence, including video snippets for visual assessment. A decision point determines whether the incident is confirmed as violence, leading to either alerting the facility and taking necessary actions or concluding with no further action if the violence is not confirmed.

Here are some of the specific steps in the activity diagram:

- Start: The "Start" component signifies the beginning of the security system's operational cycle, indicating its readiness to receive and process data from the inmate

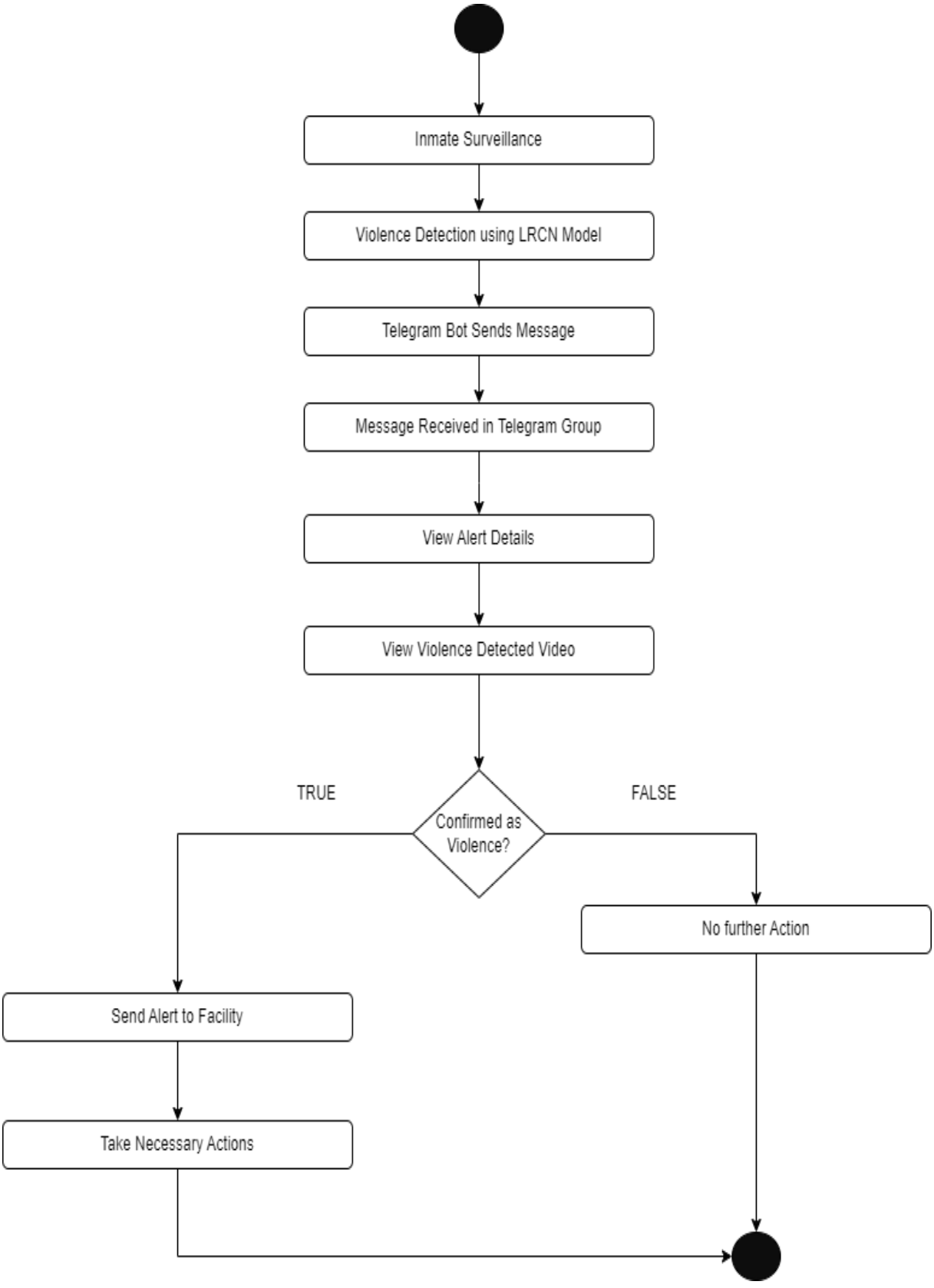


Figure 4.4: Activity Diagram of the System

surveillance system. It involves system initialization and readiness checks to ensure all components are functioning correctly.

- Inmate Surveillance: Inmate Surveillance is a critical phase where surveillance cameras strategically placed within the facility continuously monitor inmate activities. The cameras capture real-time video footage, providing a comprehensive view of events and interactions within the monitored areas.

- Violence Detection using LRCN Model: This phase involves the core functionality of the system, where a Long-term Recurrent Convolutional Network (LRCN) model is employed to detect potential instances of violence within the captured video data. The LRCN model combines convolutional layers for feature extraction and recurrent layers for temporal analysis, enabling it to identify patterns indicative of violent behavior.

- Telegram Bot Sends Message: Upon detecting potential violence, the system activates a Telegram Bot designed to automate communication via the Telegram messaging platform. The bot sends alert messages to designated individuals or groups, ensuring immediate notification of the detected incident.

- Message Received in Telegram Group: In-charge officers and supervisors receive the alert messages within the designated Telegram group. This centralized communication platform allows for swift dissemination of information to relevant personnel, ensuring they are promptly informed about the detected incident.

- View Alert Details: In-charge officers and supervisors have the capability to access detailed information regarding the detected violence within the Telegram group. This includes timestamped data, specific location details within the facility, and a concise description outlining the observed violent activity, providing a comprehensive overview for decision-making.

- View Violent Video: The ability to view recorded video snippets depicting the detected violent incidents allows in-charge officers and supervisors to gain visual context and assess the severity of the situation. This visual information aids in making informed decisions and planning appropriate response actions.

- Confirmed as Violence: Upon reviewing the alert details and video snippets, in-charge officers confirm the presence of violence. This confirmation initiates a series of response actions, ensuring a coordinated and effective response to address the confirmed violent incident.

- Send Alert to Facility: Upon confirmation of violence, the system generates and sends an alert to the facility's security personnel. This alert notifies relevant personnel within the facility, facilitating a coordinated response to the confirmed violent incident.

- Take Necessary Actions: Security personnel, upon receiving the alert, take appropriate response actions based on the confirmed violence. This may include deploying response teams, implementing safety protocols, coordinating medical assistance, or other necessary measures to mitigate the situation and ensure the safety of individuals within the facility.

- No Further Action: If, upon assessment, in-charge officers determine that the detected incident does not require immediate intervention or poses no ongoing threat, no further action is taken. This decision is based on a thorough evaluation of the situation and risk assessment.

- Stop: The endpoint of the activity diagram signifies the completion of all necessary actions or the determination that no further action is required based on the assessment conducted. It represents the closure of the response process regarding the detected violent incident within the facility.

4.1.5 Class Diagram

Figure 4.5 below illustrates a class diagram representing the violence detection system designed for correctional facilities and mental health institutions. Key classes include the Web Interface, facilitating user interaction for alert management and system monitoring; the LRCN Model, responsible for real-time analysis of video footage to detect violence; the Telegram Bot, enabling communication and alert notifications via Telegram; and Staff, representing security personnel who receive alerts and manage incident logs. These

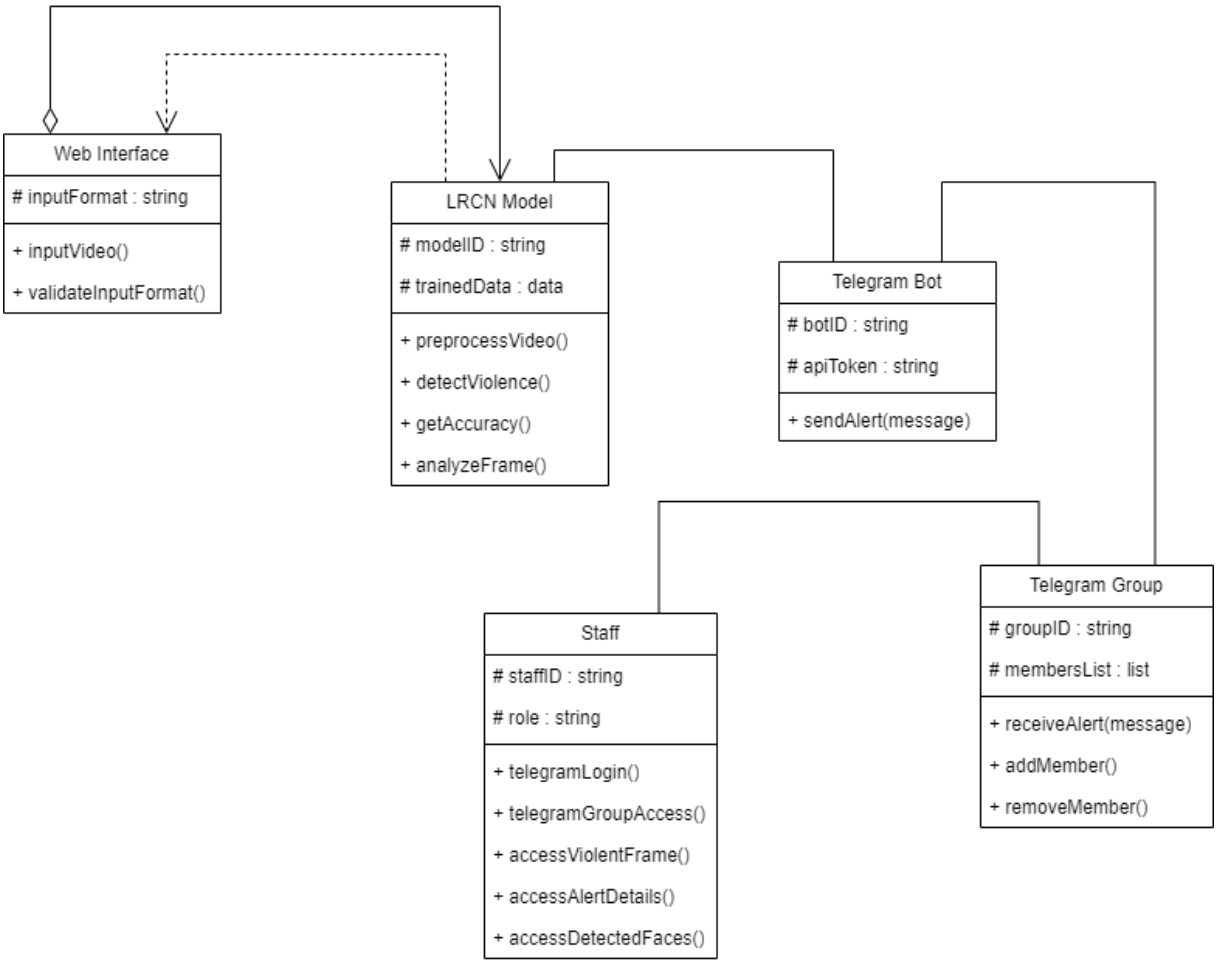


Figure 4.5: Class Diagram of the system

classes and their interactions form a unified system ensuring swift and effective response to potential security threats.

The diagram shows the different classes in the system, their attributes, and the relationships between them. The main components are:

Web Interface Class:

- Attributes:
  - inputFormat: A string representing the format of the input video.
- Methods:
  - inputVideo(): This method serves as a crucial functionality within the Web Interface class, enabling users to seamlessly input video data into the system via the web interface. In the context of your violence detection system, this method plays a vital role in facilitating the upload and integration of video footage captured by surveillance cameras or other monitoring devices. Users, such as security personnel or system administrators, can utilize the inputVideo() method to contribute essential data for analysis and detection of violent incidents. By providing a user-friendly and intuitive means of data input, this method enhances the accessibility and usability of the system, empowering users to actively engage with the surveillance and security processes.
  - validateInputFormat(): The validateInputFormat() method is an integral component embedded within the Web Interface class, designed to ensure the integrity and compatibility of the input video data with the system's processing capabilities.Here where advanced algorithms like the LRCN model are employed for violence detection, the validateInputFormat() method performs crucial validation checks on the format of the uploaded video. This validation process verifies that the input video format aligns with the system's requirements, such as resolution, codec, and encoding standards, to facilitate seamless integration with the detection algorithms. By enforcing compatibility checks, the validateInputFormat() method mitigates potential issues or errors that

may arise from incompatible video formats, ensuring smooth processing and analysis of video data for accurate violence detection.

- Relationships:
  - Composition Relationship with LRCN Model: The Web Interface class's composition relationship with the LRCN Model signifies a fundamental integration where the LRCN Model is an intrinsic part of the Web Interface's architecture. This integration extends beyond mere communication or interaction; it denotes that the LRCN Model's functionalities are seamlessly embedded within the Web Interface, forming a cohesive and unified system. For example, when a user accesses the Web Interface to input video data, the system directly engages the LRCN Model within its framework, allowing for immediate video analysis and violence detection. This composition ensures a streamlined workflow, as the Web Interface can leverage the LRCN Model's capabilities without relying on external services or modules, leading to efficient and autonomous operation.
  - Dependency Relationship with LRCN Model: The Dependency Relationship between the Web Interface class and the LRCN Model underscores the critical reliance of the Web Interface on the LRCN Model's resources and functionalities for video analysis and violence detection tasks. This dependency is not just a technical connection but represents a strategic reliance on the LRCN Model's expertise in processing video data and identifying potential security threats. Without this dependency, the Web Interface would lack the advanced algorithms, trained data, and analytical tools necessary for accurate violence detection. Therefore, the Dependency Relationship ensures that the Web Interface can effectively harness the capabilities of the LRCN Model, enhancing its ability to monitor and respond to security incidents proactively. This interdependence fosters a symbiotic relationship where the strengths of both components synergize to create a robust and reliable security monitoring system.

LRCN Model Class:

- Attributes:
  - modelID: A unique identifier for the LRCN model.
  - trainedData: Data representing the training dataset used for the LRCN model's violence detection.
- Methods:
  - preprocessVideo(): The preprocessVideo() method is a pivotal step in the violence detection system, specifically within the Long Short-Term Recurrent Convolutional Network (LRCN) model. This method is responsible for preparing raw video data for subsequent analysis. One of its primary functions is data normalization, which involves transforming the video data into a standardized format. This normalization step ensures that all video inputs have consistent characteristics, such as scale and range, which are essential for effective analysis by the LRCN model. Additionally, the preprocessVideo() method includes feature extraction processes where relevant features are identified and extracted from the video frames. These features could encompass spatial attributes like object shapes and positions, temporal aspects such as motion patterns, and any other significant visual cues that aid in violence detection. By performing data preprocessing and feature extraction, the preprocessVideo() method sets the stage for accurate and insightful analysis by the LRCN model.
  - detectViolence(): The detectViolence() method represents the core functionality of the violence detection system's LRCN model. This method leverages sophisticated deep learning algorithms embedded within the LRCN architecture to identify instances of violence within video footage. The process begins with the preprocessed video data, where the model analyzes sequential frames to detect patterns indicative of violent actions. The detectViolence() method utilizes convolutional layers to extract spatial features from each frame and recurrent layers to capture temporal dependencies across frames. This combination enables the model to understand complex relationships and context within the video stream, allowing it to differentiate between normal and violent

activities. Through extensive training on labeled datasets, the detectViolence() method learns to recognize subtle cues associated with violence, achieving a high level of accuracy in its predictions. Overall, this method plays a critical role in enhancing security by automating the detection of potential threats in real-time video surveillance.

- getAccuracy(): The getAccuracy() method serves as a vital tool for evaluating the performance of the LRCN model in violence detection. This method calculates the accuracy of the model's predictions by comparing its output against ground truth labels or annotations. The accuracy metric quantifies the model's ability to correctly classify instances of violence, providing insights into its reliability and effectiveness. A high accuracy score indicates that the model can accurately discern violent actions from non-violent ones, demonstrating its robustness in detecting security threats. The getAccuracy() method facilitates continuous monitoring and assessment of the model's performance, allowing for adjustments and optimizations to improve detection capabilities over time. It serves as a key metric for evaluating the overall efficacy of the violence detection system and ensuring its reliability in real-world applications.
- analyzeFrame(): The analyzeFrame() method operates at a granular level, focusing on detailed analysis of individual frames within video footage. This method is designed to enhance the accuracy and sensitivity of violence detection by scrutinizing frame-level information. During analysis, the model examines various visual attributes within each frame, such as motion patterns, object interactions, and spatial configurations. By analyzing frames with precision, the analyzeFrame() method can capture subtle nuances and cues that may indicate potential instances of violence. This level of scrutiny ensures that the model can detect even minor deviations or irregularities that might signify security threats. The analyzeFrame() method complements the overall violence detection process by providing fine-grained insights, contributing to a more reliable and comprehensive approach to identifying and responding to security incidents in real-time video streams.

- Relationships:
  - Composition Relationship with Web Interface: The Composition Relationship between the LRCN Model and the Web Interface class signifies a fundamental integration within the system's architecture. In this relationship, the LRCN Model is composed within the Web Interface, indicating that it is a core component responsible for video analysis capabilities. The composition ensures that the LRCN Model operates seamlessly within the broader context of the Web Interface, facilitating efficient data exchange and analysis processes. By being an integral part of the Web Interface class, the LRCN Model contributes significantly to the system's functionality, particularly in processing video data and generating insights related to violence detection.
  - Dependency Relationship with Web Interface: The Dependency Relationship between the LRCN Model and the Web Interface underscores the model's reliance on the Web Interface for essential functionalities. This dependency ensures that the LRCN Model can receive video data inputs from the Web Interface and initiate analysis processes effectively. The Web Interface acts as a gateway for providing necessary data streams to the LRCN Model, enabling it to perform complex computations and algorithms related to violence detection. The dependency relationship emphasizes the cohesive nature of the system architecture, where the Web Interface plays a crucial role in facilitating communication and data flow to support the LRCN Model's operations seamlessly.
  - Association Relationship with Telegram Bot: The Association Relationship between the LRCN Model and the Telegram Bot establishes a connection that enables communication and alerting functionalities within the system. This association allows the LRCN Model to interact with the Telegram messaging platform through the Telegram Bot, facilitating the dissemination of detection results and alert notifications. The LRCN Model utilizes this association to send alerts and notifications regarding detected instances of violence to relevant stakeholders via the Telegram messaging platform. The association relationship enhances the system's communication capabilities, ensuring that

critical information reaches designated recipients promptly and efficiently. It also showcases the collaborative nature of the system components, leveraging different modules to achieve comprehensive functionality in violence detection and response.

Telegram Bot Class:

- Attributes:
  - botID: A unique identifier for the Telegram Bot.
  - apiToken: The API token required for authentication and interaction with the Telegram messaging platform.
- Methods:
  - The sendAlert(message)‘ function plays a pivotal role in the security system by leveraging the Telegram messaging platform to disseminate crucial alert messages to relevant personnel upon detecting violence or security incidents. This method initiates a rapid communication process, ensuring that key stakeholders receive timely notifications containing vital information such as the nature of the incident, location, and timestamp. By promptly notifying personnel through Telegram, the system facilitates swift response actions, allowing stakeholders to assess the situation, coordinate resources, and implement necessary security measures effectively. This proactive alerting mechanism enhances situational awareness, enables informed decision-making, and contributes to the overall safety and security of the monitored environment.
- Relationships:
  - Association Relationship with LRCN Model: The association between the Telegram Bot class and the LRCN Model plays a pivotal role in the system's functionality. It enables seamless communication and data transfer between these components. When the LRCN Model detects violence or security incidents through its analysis of video footage, it generates detection outcomes. These outcomes, containing vital information about the detected events, are then

transmitted to the Telegram Bot. This association ensures that the Telegram Bot receives real-time updates regarding potential security threats, allowing it to promptly initiate alert messages based on the LRCN Model's findings. This association streamlines the flow of information within the system, enhancing its responsiveness to security incidents.

- Association Relationship with Telegram Group: The association between the Telegram Bot class and the Telegram Group serves as a crucial link in disseminating alert messages to relevant stakeholders. Once the Telegram Bot receives detection outcomes from the LRCN Model, it leverages this information to craft alert messages containing detailed descriptions of the detected violence or security incidents. These alert messages are then sent to the designated Telegram group, where all group members, including security personnel and administrators, have access to the information. This association enables the Telegram Bot to efficiently transmit alerts to the appropriate channels, ensuring that relevant personnel are promptly informed about security threats. The direct association with the Telegram Group streamlines the communication process, fostering quick and coordinated responses to detected violence within the system.

Telegram Group Class:

- Attributes:
  - groupID: A unique identifier for the Telegram group.
  - membersList: A list of members belonging to the Telegram group.
- Methods:
  - receiveAlert(message): The receiveAlert() method within the Telegram Group class plays a crucial role in facilitating communication and alert management within the security system. When an alert message is sent by the Telegram Bot, indicating the detection of violence or security incidents, the receiveAlert() method ensures that the Telegram Group receives and processes these alerts effectively. This functionality is vital for keeping all members of the

Telegram group informed about potential threats in real-time. Upon receiving an alert message, the Telegram Group can initiate appropriate actions, such as notifying security personnel, initiating emergency protocols, or escalating the situation as necessary. By promptly receiving and handling alert messages, the receiveAlert() method contributes to the overall responsiveness and effectiveness of the security system in addressing security challenges.

  - addMembers(): The addMembers() method, part of the Telegram Group class, provides administrators or authorized users with the capability to add new members to the Telegram group. This functionality is essential for maintaining an updated and relevant membership base within the group, ensuring that all relevant stakeholders and personnel are included in group communications and alert mechanisms. When adding new members, administrators can specify the roles and permissions associated with each member, tailoring access levels based on their responsibilities within the security system. The addMembers() method streamlines the process of expanding group participation, fostering collaboration, and enhancing communication channels among security personnel and decision-makers.
  - removeMember(): Conversely, the removeMember() method in the Telegram Group class allows administrators or designated users to remove members from the Telegram group when necessary. This functionality is crucial for managing group dynamics, ensuring that only authorized and active members remain part of the group. The removeMember() method facilitates the maintenance of a streamlined and focused communication environment, reducing clutter and ensuring that alerts and messages reach relevant individuals promptly. Administrators can utilize this method to revoke access for users who are no longer associated with security operations or who have changed roles within the organization. By removing members efficiently, the removeMember() method helps optimize group functionality and ensures that group resources are allocated effectively to active and engaged participants.

- Relationships:
  - Association Relationship with Telegram Bot: The Association Relationship

between the Telegram Group class and the Telegram Bot class is pivotal in establishing seamless communication and information flow within the security system. This association enables the Telegram Group to receive alert messages and notifications from the Telegram Bot regarding detected security threats and incidents. Through this association, the Telegram Bot serves as the communication bridge, relaying critical information to the Telegram Group in real-time. This functionality ensures that members of the Telegram Group are promptly notified about potential security risks, allowing for swift response and action. The Association Relationship with the Telegram Bot enhances the overall coordination and effectiveness of the security system by facilitating immediate communication channels and alert mechanisms.

- Association Relationship with Staff Class: The Association Relationship between the Telegram Group class and the Staff class is integral to defining the membership and composition of the Telegram group within the security system. This association signifies that the Telegram Group includes members who are part of the Staff class, representing security personnel, administrators, and relevant stakeholders. The Association Relationship ensures that the Telegram Group comprises individuals with specific roles and responsibilities related to security monitoring, incident response, and decision-making. By associating with the Staff class, the Telegram Group gains access to a diverse and knowledgeable pool of members, each contributing expertise and insights to the group's discussions and actions. This Association Relationship fosters collaboration, information sharing, and cohesive teamwork within the Telegram group, strengthening the security system's capabilities and responsiveness.

Staff Class:

- Attributes:
  - staffID: A unique identifier for each staff member.
  - role: The role or designation of the staff member within the security system.
- Methods:

- telegramLogin(): The telegramLogin() method is a pivotal functionality within the Staff class, enabling staff members to log in securely to the Telegram messaging platform. This method serves as the initial access point for staff members to engage with group communications, receive alert messages, and participate in real-time discussions related to security incidents. By authenticating users and verifying their credentials, the telegramLogin() method ensures that only authorized personnel can access the designated Telegram group. Upon successful login, staff members gain entry to critical communication channels, enhancing their ability to stay informed, collaborate effectively, and respond promptly to detected security threats. The telegramLogin() method plays a fundamental role in facilitating secure and seamless access to essential communication tools within the security system.
- telegramGroupAccess(): The telegramGroupAccess() method, part of the Staff class functionalities, facilitates staff members' access to the designated Telegram group specifically created for security-related communications and alerts. This method grants staff members entry into a centralized platform where they can receive alert messages, share updates, and collaborate with other team members in real-time. The telegramGroupAccess() method ensures that staff members have direct access to critical information and discussions within the group, enabling them to stay informed about ongoing security incidents, coordinate response actions, and contribute to decision-making processes. By providing streamlined access to the Telegram group, this method enhances communication efficiency, fosters teamwork among security personnel, and promotes a coordinated approach to security management.
- accessViolentFrame(): The accessViolentFrame() method is a specialized function within the Staff class, designed to empower staff members with the ability to view frames or snippets of video footage associated with detected violence incidents. This method allows staff members to gain visual insights into the nature and severity of security threats captured by surveillance cameras. By accessing specific frames depicting violent activities, staff members can assess the situation, gather relevant details, and make informed decisions regarding

response strategies. The accessViolentFrame() method enhances situational awareness among staff members, enabling them to comprehend the context of detected incidents and take appropriate actions promptly. This functionality contributes to a proactive and effective approach to security management, where staff members can leverage visual evidence to mitigate risks, ensure safety, and maintain a secure environment within the monitored area.

  - accessAlertDetails(): The accessAlertDetails() method provides staff members with comprehensive access to detailed information regarding detected security alerts. This functionality allows staff to delve into the specifics of each alert, including critical details such as the exact location of the incident, timestamp, nature of the observed activity, and any additional contextual information. By accessing alert details, staff members gain a deeper understanding of security events, enabling them to assess the severity of incidents, determine appropriate response measures, and coordinate actions effectively. The accessAlertDetails() method plays a crucial role in empowering staff with actionable insights, facilitating informed decision-making, and enhancing the overall responsiveness of the security system.
  - accessDetectedFaces(): The accessDetectedFaces() method equips staff members with the capability to access and review images or visual data depicting detected faces involved in security incidents. This functionality allows staff to visually identify individuals associated with detected activities, aiding in the identification, tracking, and documentation of potential threats. By accessing detected faces, staff members can corroborate information, verify identities, and gather evidence for investigative purposes. The accessDetectedFaces() method enhances staff members' situational awareness by providing visual confirmation of individuals involved in security-related incidents, contributing to a more comprehensive and effective security response.

- Relationships

  - Association Relationship with Telegram Group: The association relationship between the Staff class and the Telegram Group class represents a fundamental link that facilitates seamless communication and collaboration among staff

members within the security framework. This association enables staff members to actively participate in the Telegram group, where crucial communications, alerts, and updates regarding security incidents are shared and discussed. Through this association, staff members can receive real-time alerts, communicate with other group members, and coordinate response actions effectively. The association ensures that staff members remain connected and informed within the Telegram group, fostering a collaborative environment for addressing security challenges and enhancing situational awareness. Staff members leverage the Telegram group's platform to stay updated, share insights, and contribute to the collective effort in maintaining a secure and responsive security system.

4.2 Data Flow Diagrams

4.2.1 Data Flow Diagram - Level 0

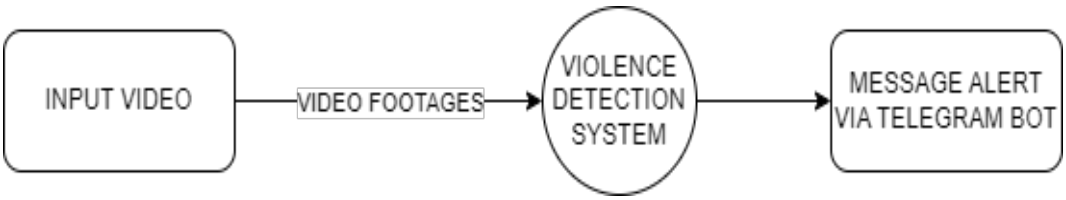


Figure 4.6: Data Flow Diagram (Level 0) of the System

Figure 4.6 illustrates the foundational data flows within the Violence Detection System. The Input Video component serves as the initial source of video footage, delivering data to the Violence Detection System Component. This component is responsible for processing the video streams, leveraging advanced algorithms and deep learning techniques to detect instances of violence in real-time. Upon detecting violence, the system triggers alert notifications through the Message Alert VIA Telegram Bot Component. This seamless flow of data and alerts enables swift and coordinated security incident response, enhancing overall safety and security within correctional facilities and mental health institutions.

A Level 0 Data Flow Diagram (DFD) provides a high-level snapshot of a system's architecture, presenting the major processes and their interactions. It features a central

process box representing the system core, through which external entities, such as users, input data and receive processed information. This macroscopic view establishes the framework for understanding the system's scope and the fundamental flow of data at the most abstract level.

The data flow within the violence detection system involves three key components:

- Input Video Component:
  - Video Footage Submission: The Input Video Component serves as the entry point for users within correctional facilities and mental health institutions to submit video footage. This process is crucial as it captures real-time activities, offering essential visual data for comprehensive security monitoring and analysis. Users can upload footage from various surveillance cameras strategically placed within the monitored areas, ensuring comprehensive coverage and high-quality input for subsequent processing.
  - Continuous Data Transmission: Once video footage is submitted, the Input Video Component ensures continuous data transmission to the Violence Detection System Component. This continuous flow of video data is essential for maintaining an up-to-date understanding of ongoing activities within the monitored environments. It enables the system to remain vigilant and responsive, allowing for timely threat detection and effective response measures.
- Violence Detection System Component:
  - Advanced Video Analysis: Within the Violence Detection System Component, the submitted video footage undergoes rigorous analysis and processing. Advanced algorithms and machine learning models are deployed to conduct sophisticated video analysis. These techniques enable the system to identify patterns, anomalies, and indicators of potential violence accurately. By leveraging cutting-edge technology, the system enhances its ability to discern security threats with precision and reliability.
  - Alert Generation and Prioritisation: Upon detecting potential instances of violence through advanced video analysis, the Violence Detection System Component initiates the generation of alert notifications. These alerts are intelligently

prioritised based on various factors such as the severity of the incident, the context of the situation, and predefined criteria. Prioritisation ensures that critical incidents receive immediate attention and prompt response actions, streamlining the overall security management process.

- Message Alert Via Telegram Bot Component:
  - Real-time Alert Dissemination: The Message Alert Via Telegram Bot Component functions as a vital communication bridge for real-time alert dissemination. Leveraging the Telegram platform's capabilities, the component ensures swift and efficient delivery of alert notifications to relevant stakeholders and security personnel. Real-time dissemination is crucial for enabling quick decision-making and response coordination, reducing potential response delays and enhancing overall security effectiveness.
  - User Interaction and Decision-making Support: In addition to alert notifications, the Telegram Bot interface empowers users with seamless interaction and decision-making support. Users can access detailed alert information, including location details, timestamp, and a concise description of the observed activity. Moreover, the Telegram Bot facilitates direct access to live video feeds, enabling users to assess the situation firsthand and make informed decisions regarding response actions. This interactive feature enhances situational awareness, empowers users with actionable insights, and promotes swift and effective response efforts.

4.2.2 Data Flow Diagram - Level 1

Figure 4.7 below illustrates the core components and data flows within the violence detection system. External entities include the Surveillance Cam, capturing video footage, and the User interacting with the system. The Cloud Storage data store houses raw video data, while processes such as Pre-processing prepare this data for analysis by the LRCN Model. The LRCN Model, a Long Short-Term Recurrent Convolutional Network, analyzes the video footage to detect violence, with Training/Test Data used for model training and evaluation. Detected violence triggers the Telegram Bot process, which sends alert notifications via the Telegram platform. Data flows include Video Data from the

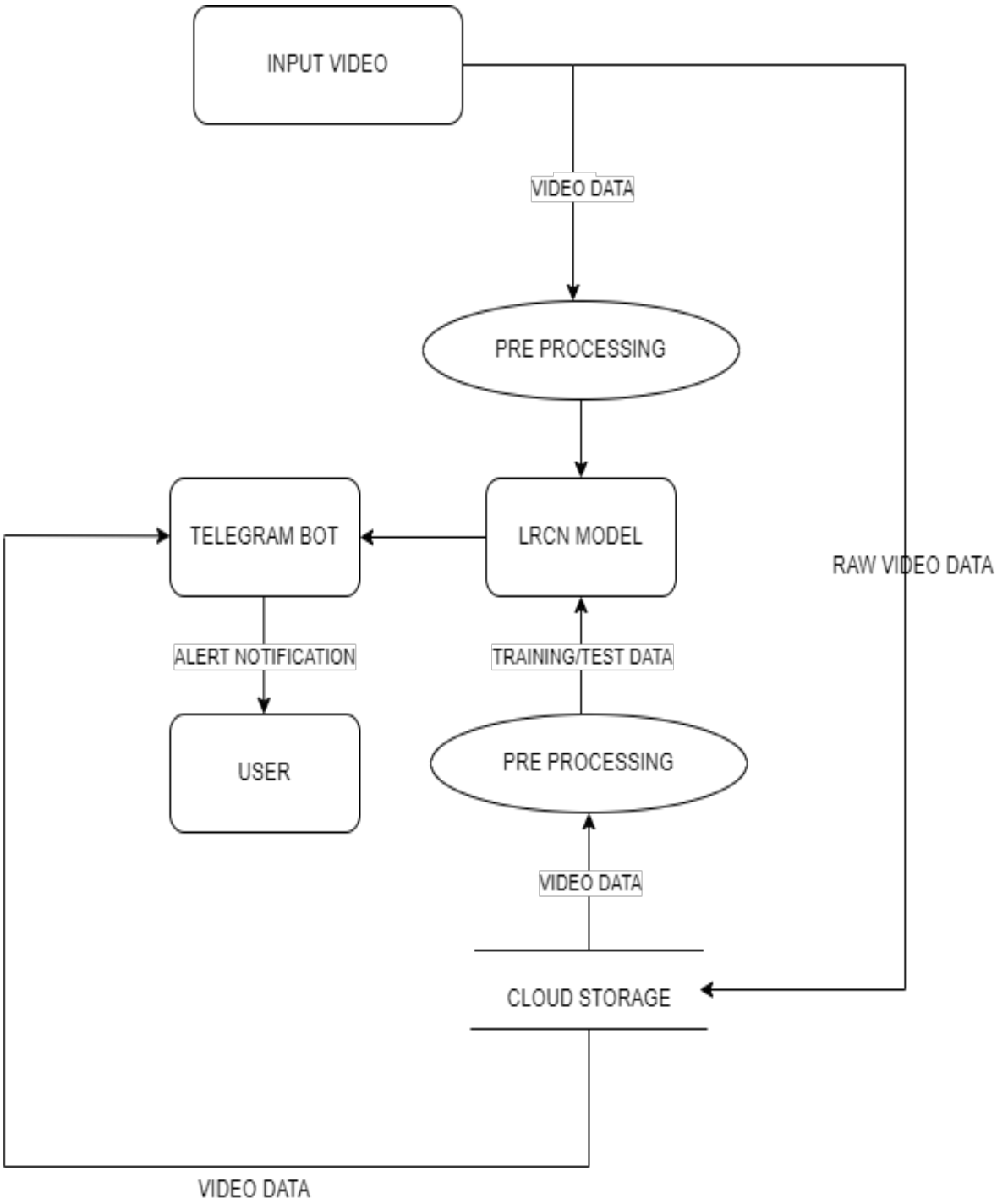


Figure 4.7: Data Flow Diagram (Level 1) of the System

Surveillance Cam, Pre-processed Video Data to the LRCN Model and Cloud Storage, Training/Test Data to the LRCN Model, and Alert Notifications from the LRCN Model to the Telegram Bot. This interconnected system ensures seamless video analysis and timely alerting for effective security management.

A Level 1 Data Flow Diagram (DFD) provides a more detailed view of specific processes identified in the Level 0 DFD. It decomposes the high-level processes into subprocesses, shedding light on the internal functions and data flows within each major process. Typically, a Level 1 DFD delves deeper into the intricacies of the system, breaking down the major components for a more comprehensive understanding of how data moves within and between them. This diagram is instrumental in offering a refined depiction of system functionality and facilitating more detailed analysis.

Components and their Roles:

- External Entities:
  - Surveillance Cam:
    - Role of Surveillance Cam: The Surveillance Camera plays a pivotal role in the system by continuously capturing real-time video footage within monitored areas, such as correctional facilities and mental health institutions. It serves as the primary source of visual data, providing crucial insights into ongoing activities and potential security threats.
    - Interaction with the System: The Surveillance Camera interacts directly with the system by transmitting raw video data for analysis and processing. This interaction is fundamental to the system's functionality, as the raw video data serves as the foundation for detecting and identifying instances of violence or security concerns within the monitored environment.
- Data Stores:
  - Cloud Storage:
    - Role: The Cloud Storage component acts as a central repository within the system, dedicated to storing video data crucial for security monitoring

and analysis. It serves as a reliable and scalable storage solution, capable of handling large volumes of raw and pre-processed video data efficiently.

    * Functionality: The primary functionality of Cloud Storage revolves around storing both raw and pre-processed video data. Raw video data, directly captured from Surveillance Cameras, is securely stored in the cloud, ensuring data availability and integrity. Additionally, the Cloud Storage component accommodates pre-processed video data, which has undergone necessary transformations and enhancements as part of the system's analysis pipeline. This comprehensive storage capability ensures that the system has access to the required video data at all stages of processing, facilitating seamless analysis and decision-making.

- Processes:
  - Pre-processing:
    * Role: The Pre-processing component plays a crucial role in preparing raw video data for in-depth analysis by the LRCN Model. It acts as an intermediary stage, where raw video footage undergoes essential transformations and optimizations before being fed into the LRCN Model for violence detection.
    * Functionality: The functionality of the Pre-processing component encompasses a range of tasks aimed at enhancing the quality and relevance of video data for violence detection. These tasks include noise reduction techniques to eliminate unwanted disturbances from the video, frame rate conversion to standardize the temporal aspect of the footage, and image resizing to ensure uniformity in input dimensions for the LRCN Model. By performing these operations, the Pre-processing component optimizes the data quality and facilitates more accurate analysis by the subsequent LRCN Model, ultimately improving the overall effectiveness of violence detection within the system.
  - LRCN Model:
    * Role: The LRCN Model assumes the critical role of analyzing pre-processed

video data to detect instances of violence within monitored environments. As the core analytical engine of the system, the LRCN Model employs sophisticated algorithms and deep learning techniques to discern complex patterns and anomalies indicative of violent behavior, leveraging insights gained from training/test data to enhance its accuracy and reliability.

* Functionality: The functionality of the LRCN Model revolves around its advanced analytical capabilities tailored specifically for violence detection. By utilizing a combination of convolutional and recurrent neural network architectures, the model is adept at extracting spatial and temporal features from pre-processed video data. These features are then analyzed and compared against learned patterns from the training/test data, allowing the model to identify and flag instances of violence with a high degree of precision. The model's functionality extends beyond mere detection, as it continually learns and adapts to evolving patterns, ensuring robust and proactive security monitoring within correctional and mental health institutions.

– Telegram Bot:

* Role: The Telegram Bot plays a pivotal role in the system by serving as the communication bridge for alert notifications. It acts as an interface between the LRCN Model's detection capabilities and the designated stakeholders, ensuring timely and efficient dissemination of critical information in response to violence detection events.

* Functionality: The primary functionality of the Telegram Bot revolves around its ability to trigger alert notifications upon detecting instances of violence through the LRCN Model's analysis. When the LRCN Model identifies a potential security threat within the monitored video footage, it sends a signal to the Telegram Bot to initiate the alert process. The Telegram Bot then generates and sends alert messages containing essential details such as timestamps, location information, and a concise description of the observed violent activity to designated recipients or groups via the Telegram messaging platform. This functionality enables swift and

coordinated responses to security incidents, empowering stakeholders with actionable information for effective decision-making and intervention.

Data Flows:

- Pre-processed Video Data to Cloud Storage:
  - Flow:After undergoing pre-processing to enhance quality and relevance, video data is transferred from the Pre-processing process to Cloud Storage. This transfer ensures that the refined and optimised video data is securely stored in a centralised repository.
  - Purpose: The purpose of this data flow is to maintain a repository of pre-processed video data in Cloud Storage. This storage strategy facilitates easy access and retrieval of optimised data for subsequent analysis by the LRCN Model. By ensuring that the data is well-organised and readily available, this flow supports efficient data management practices within the system. Moreover, storing pre-processed video data in Cloud Storage enhances the system's ability to conduct timely and accurate violence detection by providing the LRCN Model with a consistent and reliable dataset.
- Training/Test Data to LRCN Model:
  - Flow: Labeled training and test data, which includes information about the presence or absence of violence in video footage, is fed into the LRCN Model for training and validation purposes. This data flow is crucial for enhancing the model's accuracy and reliability in detecting instances of violence.
  - Purpose: The primary purpose of this data flow is to train the LRCN Model effectively. By using labeled training and test data, the model learns to identify patterns and features indicative of violence, improving its overall performance and reliability. The training process allows the model to continuously refine its understanding of violence detection, leading to enhanced accuracy and reduced false positives in identifying security threats.
- Alert Notification from LRCN Model to Telegram Bot:

  - Flow: When the LRCN Model detects violence in video footage, it generates alert notifications that are then sent to the Telegram Bot for dissemination. This flow enables real-time alerting and communication in response to detected security threats.
  - Purpose: The purpose of this data flow is to facilitate swift and effective communication in response to detected violence. By sending alert notifications to the Telegram Bot, relevant stakeholders and security personnel are promptly informed about security incidents. This real-time communication promotes timely response and intervention, contributing to a proactive security approach. The alert notifications enable stakeholders to take immediate actions to address security threats, thus enhancing overall safety and security within the monitored environment.

- Video Data Retrieval from Cloud Storage:
  - Flow: Users have the capability to access stored video data from Cloud Storage for various purposes, including viewing, analysis, and decision-making.
  - Purpose: Enabling users to retrieve video data from Cloud Storage supports their monitoring and analysis activities within the system. This flow enhances user interaction by providing access to historical video footage, allowing for comprehensive examination and informed decision-making based on past events. The availability of stored video data in Cloud Storage promotes transparency, accountability, and effective management of security incidents.

Chapter 5

Results and Discussion

In the realm of violence detection within correctional facilities and mental health institutions, the LRCN model underwent rigorous evaluation to assess its efficacy and reliability. The evaluation encompassed a range of key metrics and analyses, shedding light on the model's performance and its implications for enhancing security protocols.

The initial phase of evaluation involved training the LRCN model using a diverse dataset comprising video frames depicting both violent and non-violent activities within institutional settings. The model demonstrated robust learning capabilities during training, achieving a loss of 0.345 and an accuracy of 0.890, showcasing its ability to discern patterns indicative of violent behavior.

$$Accuracy = \frac{TruePositives + TrueNegatives}{TotalPredictions}$$

To objectively evaluate the model's performance, a stratified shuffle split cross-validation scheme was employed, where 16-frame segments of test videos were labelled as either non-violent (0) or violent (1). This labelling facilitated the calculation of crucial metrics such as

- True Positives (TP): Violent incidents the model accurately identifies.
- False Positives (FP): Non-violent incidents mistakenly labelled as violent.
- True Negatives (TN): Non-violent incidents correctly labelled as such.
- False Negatives (FN): Violent incidents the model fails to detect.

The model's accuracy, precision, recall, and specificity were computed using standard formulas, providing a comprehensive understanding of its predictive capabilities.

- Precision: Precision is a metric that measures the accuracy of positive predictions made by a model. It is calculated using the formula:

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}$$

- Recall : Also known as sensitivity or true positive rate, measures the proportion of true positive predictions out of all actual positive instances in the dataset. It indicates how many of the actual positive instances were correctly identified by the model.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

- Specificity : Also known as true negative rate, measures the proportion of true negative predictions out of all actual negative instances in the dataset. It indicates how many of the actual negative instances were correctly identified as negative by the model.

$$Specificity = \frac{TrueNegatives}{TrueNegatives + FalsePositives}$$

The computed metrics revealed promising results for the LRCN model. Precision, measuring the proportion of correctly identified violent instances among all instances predicted as violent, yielded a value of approximately 0.920. This high precision indicates the model's proficiency in accurately identifying violent activities. Similarly, the recall value of approximately 0.890 highlights the model's ability to detect a significant portion of actual violent instances. Moreover, the specificity value of around 0.962 underscores the model's aptitude for correctly identifying non-violent instances.

Figure 5.1 depicts the construction of a confusion matrix, offering a visual representation of the model's classification performance. The matrix provided insights into the model's ability to classify instances as either non-violent or violent, showcasing its overall effectiveness in discerning between these categories.

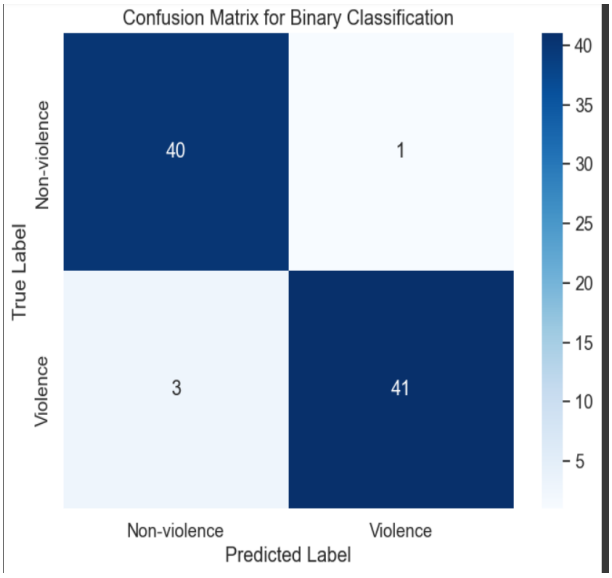


Figure 5.1: Confusion Matrix

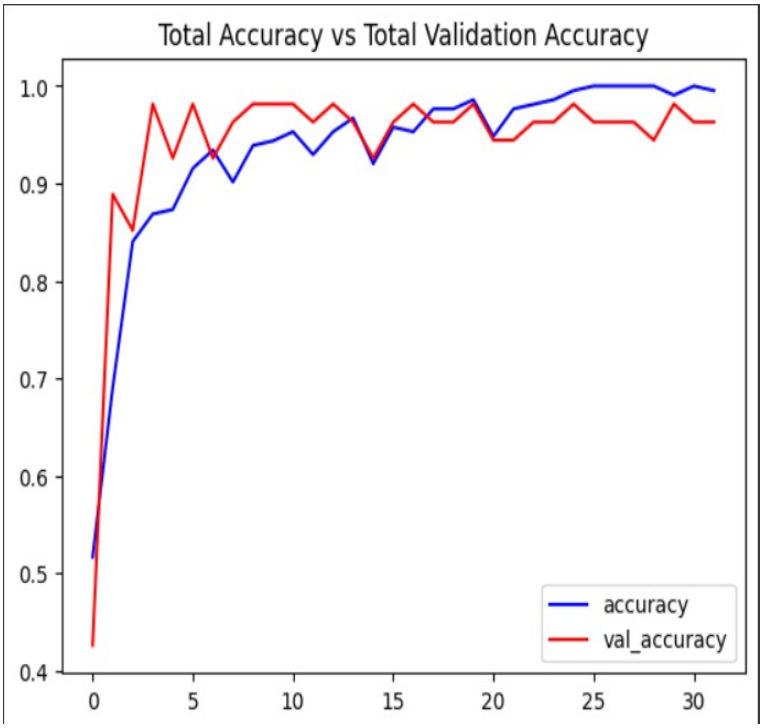


Figure 5.2: Total Accuracy vs Total Validation Accuracy

Figure 5.2 depicts the visual representations in the form of graphs that were generated to depict the model's performance metrics. The Total Accuracy vs Total Validation Accuracy graph illustrated the model's accuracy trends across training and validation phases, offering insights into its learning progress and generalization capabilities. It showcased how the model's accuracy improved during training iterations and its ability to generalize

well to unseen data during validation.

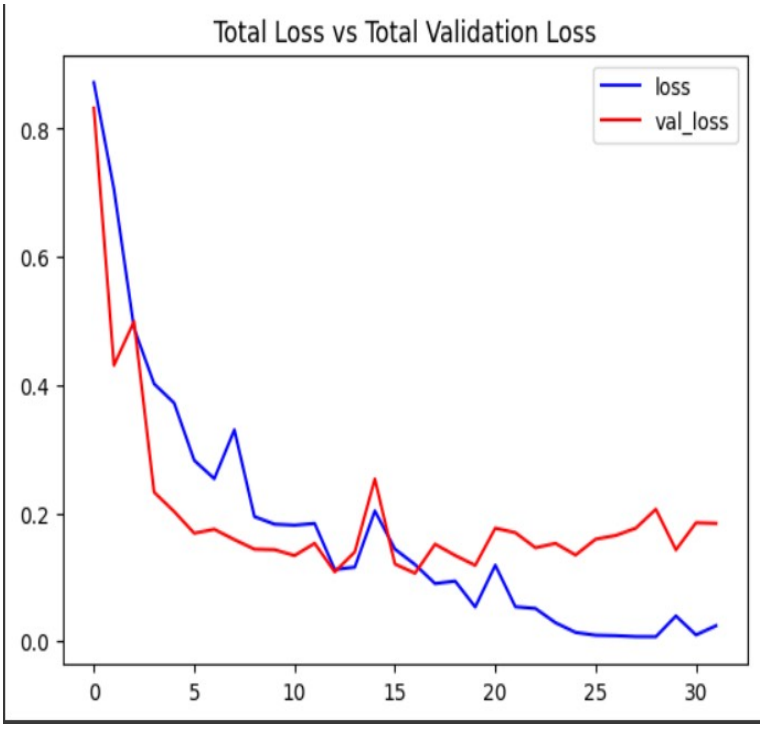


Figure 5.3: Total Loss vs Total Validation Loss

Figure 5.3 depicts the model's loss trends over training and validation epochs. A decreasing trend in loss indicated the model's capability to minimize errors and improve its predictive accuracy as training progressed. This graph provided valuable insights into the model's learning dynamics and convergence towards optimal performance.

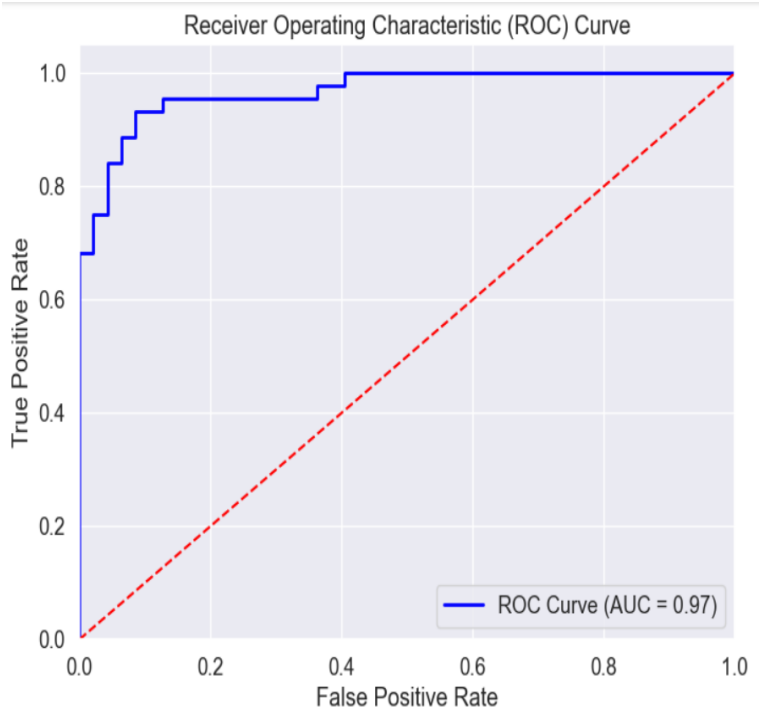


Figure 5.4: ROC Curve

Figure 5.4 depicts the ROC-AUC (Receiver Operating Characteristic - Area Under Curve) curve to evaluate the model's performance across various threshold values for distinguishing between non-violent and violent instances. The curve visually represented the trade-off between true positive rate (sensitivity) and false positive rate (1 - specificity) as the classification threshold varied. A higher area under the ROC curve indicated superior discrimination capability of the model, showcasing its effectiveness in differentiating between violent and non-violent activities.

The results showcased the LRCN model's robustness in detecting violent incidents within correctional facilities and mental health institutions. Its high accuracy, precision, recall, and specificity metrics validate its efficacy in real-time violence detection, providing security personnel with valuable tools for swift intervention and prevention. The integration of cutting-edge technology like LRCN signifies a significant step forward in fortifying security measures and ensuring the safety of individuals within these environments. As advancements in artificial intelligence continue to shape security protocols, the LRCN model stands as a testament to innovation in safeguarding vulnerable populations and fostering secure institutional settings.

Chapter 6

Conclusion

In conclusion, Vigilance X emerges as a pivotal advancement in bolstering safety and security within correctional facilities and mental health institutions. The amalgamation of cutting-edge technologies, including deep learning, recurrent networks, and real-time communication through Telegram, underscores a proactive approach towards violence detection and prevention. By harnessing the power of artificial intelligence, Vigilance X not only identifies potential threats in real-time but also equips security personnel with actionable insights for swift intervention.

Looking ahead, several avenues for future enhancements and expansions present themselves. One promising direction is the implementation of real-time surveillance capabilities, enabling Vigilance X to operate seamlessly without the need for pre-recorded video inputs. This would significantly enhance the system's responsiveness and accuracy, further fortifying its ability to mitigate potential risks effectively. Additionally, the development of a dedicated mobile application tailored to the Vigilance X ecosystem holds immense potential. Such an application could streamline the process of receiving alerts, accessing vital information about detected incidents, and facilitating rapid response actions. Integrating features like live video feeds, interactive mapping of incidents, and real-time communication channels would empower security personnel with enhanced situational awareness and decision-making capabilities. Furthermore, ongoing research and development efforts could focus on refining the underlying algorithms and models, fine-tuning them for the unique challenges and nuances of correctional environments.

In essence, Vigilance X stands as a testament to the transformative potential of technology in safeguarding vulnerable populations and creating safer environments. With a commitment to ongoing improvement and a vision for future advancements, Vigilance X sets a benchmark for intelligent, proactive security solutions in complex operational settings.

References

[1] A. Traoré and M. A. Akhloufi, "Violence detection in videos using deep recurrent and convolutional neural networks," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2020, pp. 154–159.

[2] G. Aldehim, M. Asiri, M. Aljebreen, A. Mohamed, M. Assiri, and S. Ibrahim, "Tuna swarm algorithm with deep learning enabled violence detection in smart video surveillance systems," *IEEE Access*, vol. PP, pp. 1–1, 01 2023.

[3] P. Sernani, N. Falcionelli, S. Tomassini, P. Contardo, and A. F. Dragoni, "Deep learning for automatic violence detection: Tests on the airtlab dataset," *IEEE Access*, vol. 9, pp. 160 580–160 595, 2021.

[4] S. Jianjie and Z. Weijun, "Violence detection based on three-dimensional convolutional neural network with inception-resnet," in *2020 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)*, 2020, pp. 145–150.

[5] W.-F. Pang, Q.-H. He, Y.-j. Hu, and Y.-X. Li, "Violence detection in videos based on fusing visual and audio information," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 2260–2264.

[6] C. Jiang, Y. Chen, S. Chen, Y. Bo, W. Li, T. Wenxin, and J. Guo, "A mixed deep recurrent neural network for mems gyroscope noise suppressing," *Electronics*, vol. 8, p. 181, 02 2019.

Appendix A: Presentation

VigilanceX

Guided By
~Ms. Bency Wilson

Ashwin Saji U2004023
Cyriac John U2004031
Megha Milton U2004054
Roshan Xavier U2004061

OVERVIEW

VigilanceX

INTRODUCTION

VigilanceX

This project:

- Entails a dedicated Violence Detection System customized for correctional facilities.
- Leverages cutting edge artificial intelligence and video analytics to promptly detect and counteract violence as it occurs.
- Integration of deep learning-based violence detection systems within correctional facilities and mental health institutions.
- Stands as a critical stride towards bolstering overall safety and security.

Motivation

VigilanceX

- Saving Lives
- Ensuring Security
- Legal Duty
- Violence Prevention
- Efficient Operations
- Inmate Safety
- Staff Safety
- Prevent Escalation

PROBLEM STATEMENT

VigilanceX

WHY ?

In correctional facilities like jails and mental asylums, ensuring the safety of inmates and staff is an ongoing challenge. Incidents of violence within these facilities can have severe consequences, including injuries and property damage. Therefore, there is a critical need for a specialized Violence Detection System designed to proactively identify and respond to violence in real time using surveillance camera technology.

GOALS AND OBJECTIVES

VigilanceX

1. Ensure Safety:

- Goal: Enhance safety and security within jails and mental health institutes.
- Objectives: Minimize harm and injuries resulting from violence and protect inmates, staff, and patients.

2. Prevent Violence:

- Goal: Proactively prevent violent incidents from occurring.
- Objectives: Identify early signs of potential violence to intervene and de-escalate situations.

3. Customization and Adaptation:

- Goal: Create adaptable systems that suit the unique needs of each facility.
- Objectives: Allow for customization of detection criteria to account for facility-specific dynamics.

4. Integration with Existing Infrastructure:

- Goal: Seamlessly integrate with current security and healthcare systems.
- Objectives: Work in conjunction with existing surveillance cameras, alarm systems, and communication networks.

5. Minimize False Positives:

- Goal: Enhance system accuracy while minimizing unnecessary alerts.
- Objectives: Continuously improve machine learning models to reduce false positives.

6. Rehabilitation and Well-being:

- Goal: Support rehabilitation and the well-being of individuals in mental health institutes.
- Objectives: Prevent self-harm or harm to others among patients.

7. Public Perception and Accountability:

- Goal: Enhance public perception and institutional accountability.
- Objectives: Showcase the commitment to safety and security.

CHALLENGES

Implementing violence detection systems in jails & mental health institutes is crucial for safety but it comes with several challenges:

- **Privacy Concerns**
- **Ethical Considerations**
- **False Positives**
- **Data Handling & Security**
- **System Customization**
- **Resource Constraints**
- **Training & User Acceptance**
- **Patient Sensitivity**

Addressing these challenges requires a careful balance between security, privacy & transparency to successfully implement violence detection systems in jails and mental health institutes.

PROJECT SCOPES

- Real-time monitoring for swift incident identification.
- Customization to suit facility-specific requirements.
- Seamless integration with existing infrastructure.
- Strict adherence to privacy regulations.
- Detailed incident reporting for enhanced safety and security within correctional facilities.

TARGET GROUPS

The targeted groups for violence detection systems in jails & mental asylums encompass :

- Corrections officers and security personnel
- Jail administrators and facility manager

- Inmates and patients
- Healthcare providers, such as psychiatrists and nurses
- Law enforcement and legal authorities

RESOURCES NEEDED

HARDWARE RESOURCES

Input Devices: Surveillance camera

Output Device: Mobile/PC

Hard Disk: 2TB

Processor: Intel Core i7 or above

Server Components: CPU, Memory(RAM)

SOFTWARE RESOURCES

- Keras
- OpenCv
- Mongo DB database
- FFmpeg
- Google Colab
- React

LITERATURE SURVEY

SL NO	PAPER TITLE	YEAR OF PUBLICATION & JOURNAL NAME	METHOD	LIMITATION
1	**Violence Detection in Videos Using Deep Recurrentand Convolutional Neural Networks:** Abdarahmane Traore and Moulay A. Akhloufi, ´ Senior Member IEEE Perception, Robotics, and Intelligent Machines Research Group (PRIME) Department of Computer Science, Universite de Moncton	**2020 IEEE International Conference on Systems**	• **Convolutional Neural Network (CNN):**The paper uses EfficientNet-B0, a deep convolutional neural network (CNN), for extracting spatial features from video frames. • **Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU):**The paper leverages LSTM and GRU, which are types of recurrent neural networks (RNNs), to capture temporal features in video sequences. • **Datasets:**Three datasets are used for evaluation: Hockey dataset, Violent Flow dataset, and Real Life Violence Situations Dataset.	• **Computational Complexity:** The paper mentions the use of deep neural networks, which can be computationally intensive, especially when dealing with large video datasets. • **Lack of Real-time Considerations:** It's important to note that the paper does not specifically address the real-time processing requirements for violence detection in videos. Real-time applications may have stricter constraints on processing time.

SL NO	PAPER TITLE	YEAR OF PUBLICATION & JOURNAL NAME	METHOD	LIMITATION
2	**Tuna Swarm Algorithm With Deep Learning Enabled Violence Detection in Smart Video Surveillance Systems** **1Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University.**	**2023 IEEE Access**	• **Feature Extraction with Deep Learning**: Use a deep neural network, Residual-DenseNet for feature extraction. • **Tuna Swarm Algorithm for Optimization**:TSO model is utilized as a hyperparameter enhancer for the residual-DenseNet model. • **SAE** model is enforced for the classification of events into violence and non-violence events	• **Algorithm Complexity** **The complexity of this optimization process depends on the number of hyperparameters being tuned, the search space, and the number of iterations. Optimization algorithms like TSO typically involve running the model with different hyperparameters multiple times to find the best set, which can be computationally expensive.** • **Data Quality and Diversity** **The effectiveness of deep learning-based models, including Residual-DenseNet, heavily relies on the quality and diversity of the training data.**

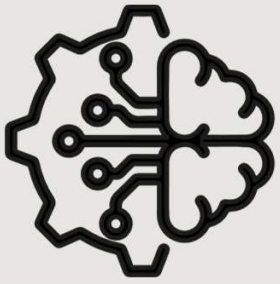
SL NO	PAPER TITLE	YEAR OF PUBLICATION & JOURNAL NAME	METHOD	LIMITATION
3	**Deep Learning for Automatic Violence Detection: Tests on the AIRTLab Dataset:** PAOLO SERNANI 1 , NICOLA FALCIONELLI1 , SELENE TOMASSINI1 , (Student Member, IEEE), PAOLO CONTARDO 1,2, AND ALDO FRANCO DRAGONI 1	**2021 IEEE Access**	• **Datasets:** The research utilizes the AIRTLab dataset, which contains video clips labeled as violent and non-violent. • **Model Types:** Model 1: Uses the C3D architecture as a feature extractor and employs a linear Support Vector Machine (SVM) classifier. Model 2: Also uses C3D as a feature extractor but integrates fully connected layers for end-to-end classification. Model 3: Employs the Convolutional Long Short-Term Memory (ConvLSTM) architecture, trained from scratch, for violence detection.	• **Lack of Real Violence in Dataset** : The dataset comprises videos from non-professional actors and lacks real instances of violence, potentially limiting its representation of real-world violent scenarios. • **Training Complexity:** ConvLSTM networks typically have more complex training procedures compared to traditional CNNs. • **Testing on Short Video Sequences.**

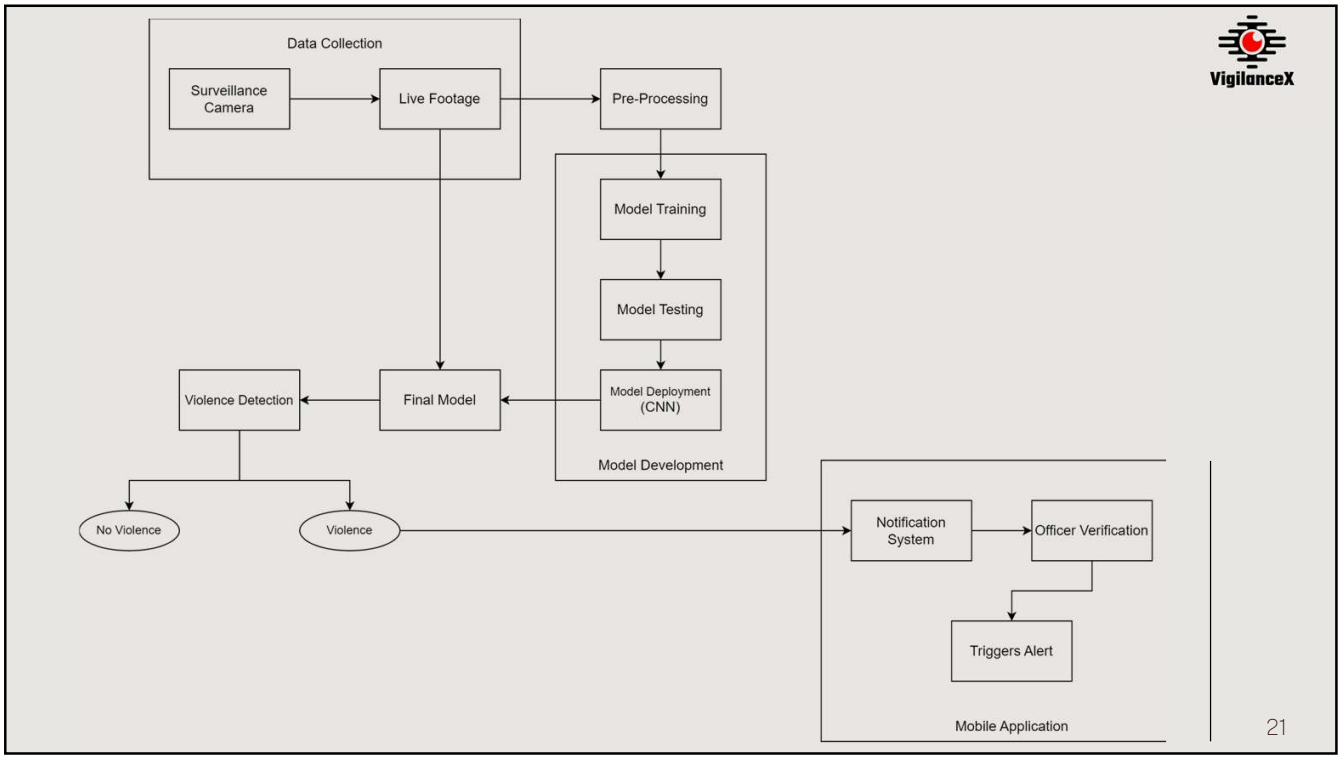
SL NO	PAPER TITLE	YEAR OF PUBLICATION & JOURNAL NAME	METHOD	LIMITATION
4	**Violence Detection Based on Three-Dimensional Convolutional Neural Network with InceptionResNet**	**2020 IEEE Conference on Tele communications, Optics and Computer Science (TOCS)**	• **Three-Dimensional Convolutional Neural Network (3D CNN):** Processes both spatial (width and height of video frames) and temporal (sequence of frames) aspects of a video. • **Inception-ResNet Architecture**: Combines the Inception architecture and the ResNet architecture	• **Overfitting:** • **Data Requirement:** • **Interpretability:** • **Transfer Learning Limitations** • **Optimization Challenges:** • **Memory Usage**

SL NO	PAPER TITLE	YEAR OF PUBLICATION & JOURNAL NAME	METHOD	LIMITATION
5	**VIOLENCE DETECTION IN VIDEOS BASED ON FUSING VISUAL AND AUDIO INFORMATION** Wen-Feng Pang Qian-Hua He * Yong-jian Hu Yan-Xiong Li	**2021 School of Electronic and Information Engineering South China University of Technology, Guangzhou, China**	• **Visual Feature Extraction** Temporal dynamics can be captured using recurrent layers or 3D CNNs. • **Audio Feature Extraction:** Deep learning architectures, such as CNNs or RNNs, can be applied to these representations to capture audio patterns indicative of violence. • **Feature Fusion:** • **Early Fusion:** • **Late Fusion:** • **Hybrid Fusion:**	• **Ambiguity in Interpretation** • **Generalization** • **Computational Overhead:** • **Feature Fusion Challenges** • **Noise**

VigilanceX

SYSTEM ARCHITECTURE





ALGORITHMS IN USE

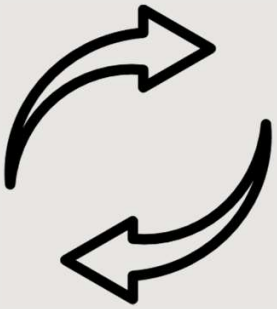
VigilanceX

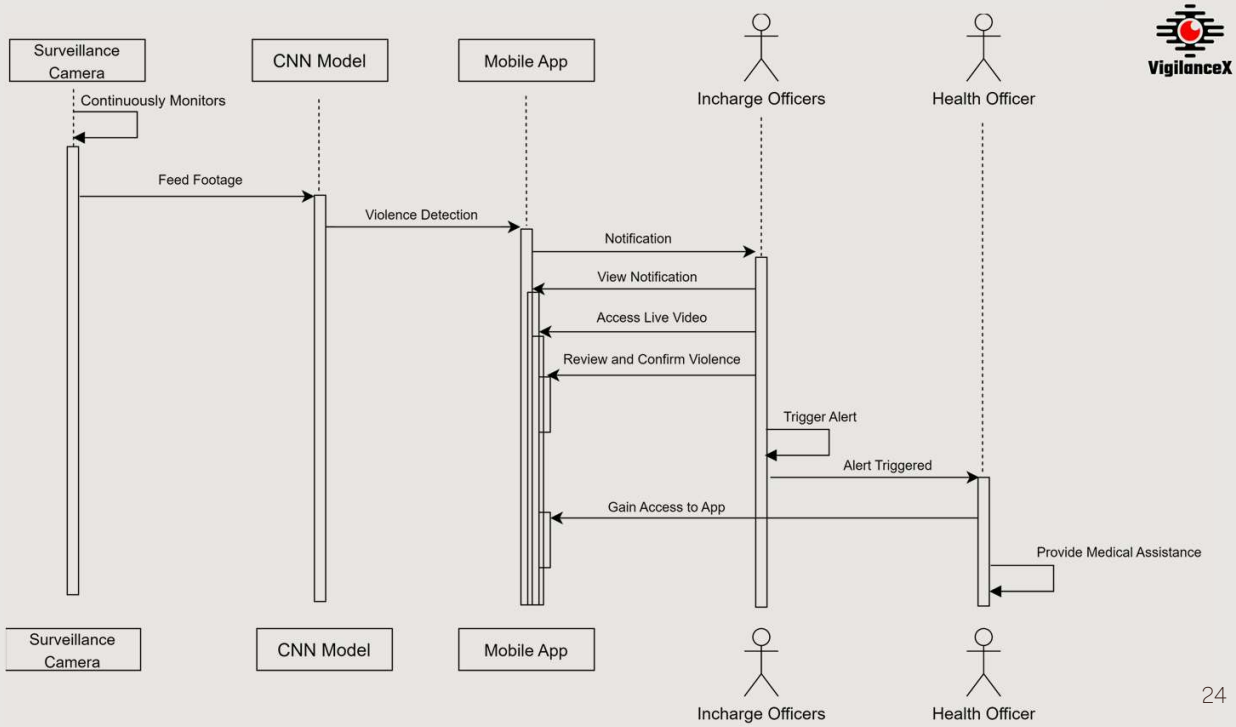
Convolutional Neural Networks (CNNs) are the core algorithm in our project

- They are used for a violence detection system.
- Excel in analyzing image and video data.
- Optimal for real-time identification of violent actions or individuals from camera feeds.
- **Convolutional Layers**: Extract relevant features from each frame.
- **Fully Connected Layers:** Used for classification
- Training: CNNs are trained on labeled datasets.
- Fine-Tuning: Pre-trained models can be further fine-tuned if required.
- Performance: Efficiently distinguishes between violent and non-violent activities in video streams.

VigilanceX

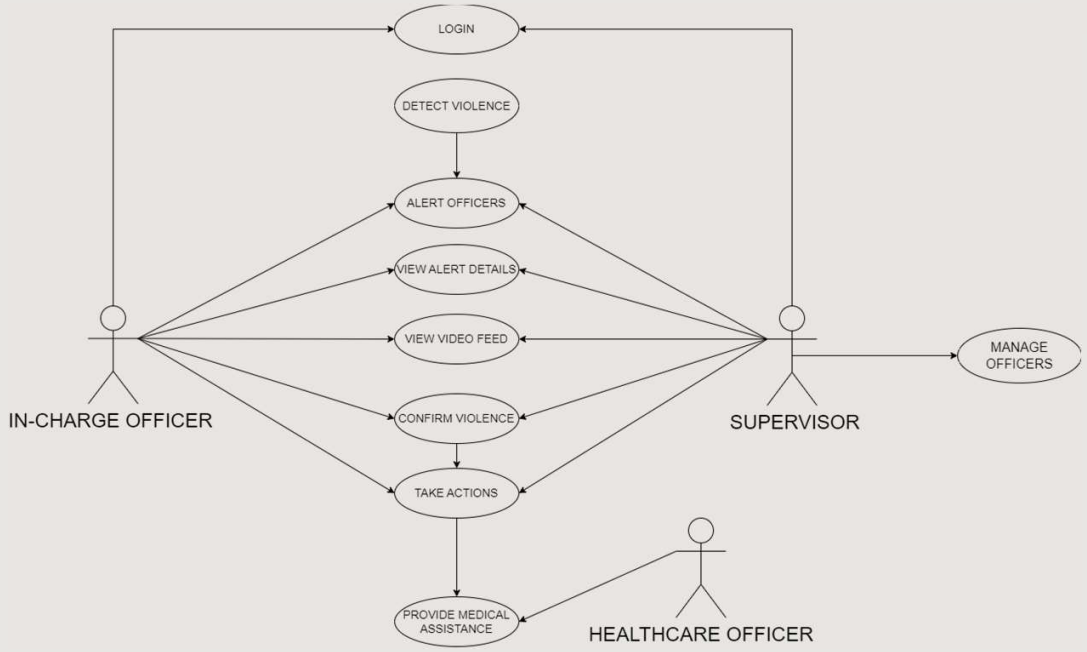
SEQUENCE DIAGRAM





USE CASE DIAGRAM

VigilanceX



VigilanceX

WORK PLAN

	September	October	November	December	January	February	March	April
Project Initiation								
Literature Survey								
Requirement Specification								
Gathering Required Hardware								
Usage of Camera to Collect Data								
Model Training								
Fine Tuning the Model								
App Design								
App Development								
Testing and Debugging								
Project Deployment								
Documentation and Project Report								
Submission								

CONCLUSION

In summary, VigilanceX is a significant advancement in enhancing safety and order. By leveraging cutting-edge technology, it equips staff with proactive tools to address and prevent violence, reducing harm. It aims to create a safer environment, fostering transformative change for inmates and staff.

THANK YOU

Appendix B: Vision, Mission, Programme Outcomes and Course Outcomes

Vision, Mission, Programme Outcomes and Course Outcomes

Institute Vision

To evolve into a premier technological institution, moulding eminent professionals with creative minds, innovative ideas and sound practical skill, and to shape a future where technology works for the enrichment of mankind.

Institute Mission

To impart state-of-the-art knowledge to individuals in various technological disciplines and to inculcate in them a high degree of social consciousness and human values, thereby enabling them to face the challenges of life with courage and conviction.

Department Vision

To evolve into a department of excellence in information technology by the creation and exchange of knowledge through leading-edge research, innovation and services, which will in turn contribute towards solving complex societal problems and thus building a peaceful and prosperous mankind.

Department Mission

To impart high-quality technical education, research training, professionalism and strong ethical values in the young minds for ensuring their productive careers in industry and academia so as to work with a commitment to the betterment of mankind.

Programme Outcomes (PO)

Engineering Graduates will be able to:

1. Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. Conduct investigations of complex problems: Use research-based knowledge including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. Modern Tool Usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. Individual and Team work: Function effectively as an individual, and as a member or leader in teams, and in multidisciplinary settings.

10. Communication: Communicate effectively with the engineering community and with society at large. Be able to comprehend and write effective reports documentation. Make effective presentations, and give and receive clear instructions.

11. Project management and finance: Demonstrate knowledge and understanding of engineering and management principles and apply these to one's own work, as a member and leader in a team. Manage projects in multidisciplinary environments.

12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and lifelong learning in the broadest context of technological change.

Programme Specific Outcomes (PSO)

Information Technology Program Students will be able to:

PSO1: Acquire skills to design, analyse and develop algorithms and implement those using high-level programming languages.

PSO2: Contribute their engineering skills in computing and information engineering domains like network design and administration, database design and knowledge engineering.

PSO3: Develop strong skills in systematic planning, developing, testing, implementing and providing IT solutions for different domains which helps in the betterment of life.

Course Outcomes (CO)

CO1: Model and solve real world problems by applying knowledge across domains (Cognitive knowledge level: Apply).

CO2: Develop products, processes or technologies for sustainable and socially relevant applications (Cognitive knowledge level: Apply).

CO3: Function effectively as an individual and as a leader in diverse teams and to comprehend and execute designated tasks (Cognitive knowledge level: Apply).

CO4: Plan and execute tasks utilizing available resources within timelines, following ethical and professional norms (Cognitive knowledge level: Apply).

CO5: Identify technology/research gaps and propose innovative/creative solutions (Cognitive knowledge level: Analyze).

CO6: Organize and communicate technical and scientific findings effectively in written and oral forms (Cognitive knowledge level: Apply).

Appendix C: CO-PO-PSO Mapping

CO - PO Mapping

CO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
1	3	1	2	1		2	1					2
2	2	1	2	3		1	3	1		2	2	3
3	3	2			2	1		2		3	1	3
4	3			2	2			1	2	3		2
5	3	2	3	3		2	1		2	2	3	2
6	1	1	3				1	2		2	2	2

CO - PSO Mapping

CO	PSO 1	PSO 2	PSO 3
1	1		
2	2		1
3	2		
4			2
5			2
6	2		1