Real-Time Monitoring and Detection of Aggressive Behaviors with Alert System

DELCO, Kuh-kai A.
Electronics Engineering Department
Technological University of the Philippines

FERIDO, Angelo
Electronics Engineering Department
Technological University of the Philippines

FRANCO, Kelyn
Electronics Engineering Department
Technological University of the Philippines

MANADONG, Jovert
Electronics Engineering Department
Technological University of the Philippines

SANTIAGO, Brian Jamille
Electronics Engineering Department
Technological University of the Philippines

SALVADOR, Jean Prynce
Electronics Engineering Department
Technological University of the Philippines

ENGR. Mark Melegrito, PECE
Faculty
Electronics Engineering Department
Technological University of the Philippines

I. INTRODUCTION

Injuries contribute to 10% of global mortality and 15% of disability, yet data on injuries in developing countries, where two-thirds of injury deaths occur, are scarce. This report is the first to examine the issue of injuries in the Philippines, a developing country in Southeast Asia. It aims to define the burden of injuries and highlight priority areas for national health research. A review of 35 years of data (1960-1995) reveals a significant increase in injury fatality rates, rising by 196% from 14.3 per 100,000 in 1960 to 42.3 per 100,000 in 1995. One in 11 deaths in the Philippines is due to injuries, with intentional injuries accounting for 48% of all injury deaths. The report emphasizes the need for improved injury surveillance and documentation of non-fatal injury outcomes. It also calls for research into risk factors and interventions to prevent intentional injuries, highlighting the importance of addressing this issue in the Philippines (Consunji & Hyder, 2004).

Video surveillance systems are vital for public security, detecting suspicious behaviors, analyzing crowd dynamics, managing traffic, and tracking vehicles. Manual monitoring of these activities is challenging, leading to research efforts to automate information extraction using machine learning and deep learning. However, the lack of large, annotated video datasets poses a common barrier in this domain. Existing videos are often untrimmed, unannotated, and may contain ambiguous data. This research aims to address these limitations by proposing a machine learning-based transfer learning approach. The goal is to accurately detect violent crowd behavior through surveillance systems (Liyanage & Fernando, 2021).

II. BACKGROUND OF THE STUDY

In surveillance, detecting human abnormal behavior is crucial for public safety, but it's challenging due to lengthy video datasets and occasional occurrences of abnormal activities. This often requires extensive manpower to verify video streams, making the process expensive, inefficient, and time-consuming (Kim et. al., 2021).

The aim of surveillance is to prevent unwanted incidents and respond to them as needed, depending on the situation. An AI-based automated system is necessary to process videos and enhance societal well-being. Recognizing human abnormal behavior through automated surveillance can improve safety for the elderly and patients, reduce criminal activity and theft, decrease workplace harassment and violence (Kim et. al., 2021).

III. STATEMENT OF THE PROBLEM

Maintaining a safe and secure environment in barangays requires vigilance. Aggressive behaviors, from verbal arguments to physical threats, can create fear and disrupt the peace within a community. Traditional methods for addressing such behaviors often rely on resident reports, or security personnel intervention. These methods are reactive that leaves residents feeling vulnerable and may not be sufficient to prevent violence or escalation of tensions. TANAW can identify threatening situations, its real-time detection capability relies on machine learning algorithms that analyzes video feeds from strategically placed cameras. By identifying aggressive behaviors like kicking, strangling, punching, and hair pulling in real-time, TANAW empowers authorities to intervene before situations escalate further.

Early detection is crucial in preventing violence and maintaining a peaceful barangay. Upon detecting aggressive behavior, the alert system of TANAW will be triggered notifying the barangay. This allows for a quick and targeted response before they escalate further. This system and approach to the community promotes a more secure environment for the barangay.

IV. OBJECTIVES

This research aimed to create a Real-Time Detection of Aggressive Behaviors with Alert System

1. To develop a real-time monitoring and detection system for emergency hand gesture recognition with alert system.
2. Implement and improve the real-time detection capability of the system to identify aggressive behaviors and enable authorities for timely intervention in barangays.

V. RELATED STUDIES

According to the study of Hassner & Kliper-Gross (2012), they emphasize the critical need for real-time detection of violent outbreaks in crowded events using surveillance cameras. It introduces a novel approach called ViF (Violent Flows) for efficient crowd violence detection, which outperforms existing techniques by focusing on the magnitudes of optical-flow fields. The proposed method considers how flow-vector magnitudes change over time and uses this information to classify scenes as either violent or non-violent using linear SVM. Additionally, the study provides a unique dataset of real-world surveillance videos and benchmarks to evaluate violent/non-violent classification and real-time detection accuracy. By comparing their method to state-of-the-art techniques, the study demonstrates the effectiveness of their approach in detecting aggressive behaviors in crowded environments, aligning with efforts in real-time detection of aggressive behaviors with alert systems.

The study of Sumon, et al. (2019) focuses on detecting violent crowd flows using deep learning algorithms applied to a small dataset of violent and non-violent videos. The researchers found that a convolutional neural network (CNN) using transfer learning performed better than other CNN and long short-term memory network (LSTM) models. Combining CNN with LSTM improved accuracy but still didn't surpass the transfer learning model. In future studies, they plan to make the model more lightweight through pruning and deploy it on an unmanned aerial vehicle (UAV) for real-time monitoring. They also aim to create an API for easy access to the model via a web server. This research contributes to real-time detection of aggressive behaviors, aligning with efforts in Real-Time Detection of Aggressive Behaviors with Alert Systems.

VI. METHODOLOGY

This section outlines for developing a real-time detection of aggressive behaviors. The methodology includes Dataset collection, Image Annotation and Augmentation, Model Training and Evaluation, and Deployment. The other procedures expound about public dataset gaining, custom dataset creation, image annotation using Roboflow, model training and configuration, evaluation with false positive/negative analysis, image augmentation, addressing class imbalance and hyperparameter tuning, different training epochs, and deployment for real-time video processing.

A. Dataset Collection

Researchers searched for publicly available video datasets containing examples of aggressive behavior, such as hair pulling, kicking, punching, and strangling. It was critical to find datasets with a balanced representation of aggressive and non-aggressive interactions to avoid bias in the model. This balanced representation would help the model distinguish between the two types of behavior more effectively. Additionally, the search prioritized datasets captured by CCTV cameras. This ensured that the model was trained on data closely resembling the real-world scenarios it would be deployed, considering factors like video quality and recording angles. Using CCTV footage, the model could adapt to the specific characteristics of these surveillance systems, improving its effectiveness.

B. Data Processing

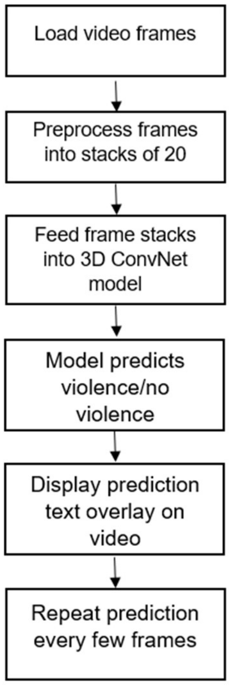

Load video frames
Preprocess frames into stacks of 20
Feed frame stacks into 3D ConvNet model
Model predicts violence/no violence
Display prediction text overlay on video
Repeat prediction every few frames

Figure 1: Program Flow for Aggressive behaviors

This Python code implements a real-time violence detection system using a convolutional neural network model.

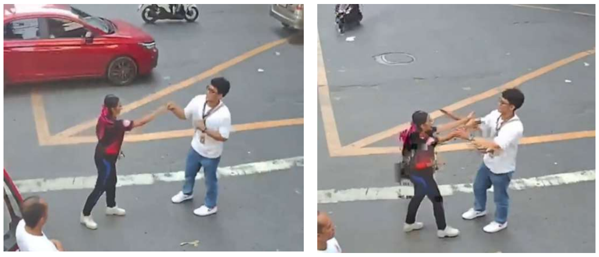
It takes a live video stream from an RTSP URL as input. The video frames are preprocessed by resizing, converting to NumPy arrays, and stacking together groups of frames. These stacked frames are fed into a 3D convolutional neural network model to predict whether violence is present or not.

The model architecture consists of several 3D convolutional layers, max-pooling layers, dropout layers, and dense layers. The SELayer function implements a squeeze-and-excitation block that helps the model pay attention to useful features. The conv3Dnet_and conv3Dnet_b functions define different types of 3D convolutional blocks used in the model.

The key outputs are a prediction variable pred_var that classifies each video segment as either 'VIOLENCE' or 'NO VIOLENCE', and overlay text rendered onto the original video frames displaying this prediction. The prediction variable drives the color and position of the overlay text. It implements an end-to-end real-time violence detection system by using a 3D convolutional neural network on live video.

C. Model Training and Evaluation:

The system relies on TensorFlow to analyze individual frames extracted from video footage. This program comprises three essential parts working together: The first focuses on reducing the amount of information within each video frame, concentrating on the most critical details relevant to identifying aggressive behavior. The second part helps the program become more robust against differences in video quality often encountered in CCTV footage. This is achieved by introducing slight artificial changes to the training data, such as adjusting lighting or camera angles. The final part uses specialized filters and analysis techniques to pinpoint frames containing aggressive behavior within the video footage. Researchers collected videos showcasing aggressive behavior to train this program effectively and then separated them into individual frames. The program fed these frames into three groups: training, testing, and validation. The training allows the program to learn and identify patterns associated with aggressive behavior. The testing evaluates the program's accuracy in real-world scenarios. Finally, the validation plays a crucial role in preventing the program from becoming overly specific to the training data, ensuring it generalizes well to real-world CCTV footage.



Figures 2-5: Dataset Collection

D. Deployment

The trained LSTM model will be deployed for real-time analysis of network traffic data streams. Incoming network traffic will be continuously preprocessed and fed into the model to predict potential anomalies. The system will be configured to trigger alerts based on detected anomalies, allowing network security personnel to investigate and take appropriate actions. The deployed model's performance will be monitored continuously to ensure its effectiveness. Retraining or fine-tuning may be necessary as network traffic patterns evolve.

VII. RESULT AND DISCUSSIONS

a. Modeling Training and Evaluation Results

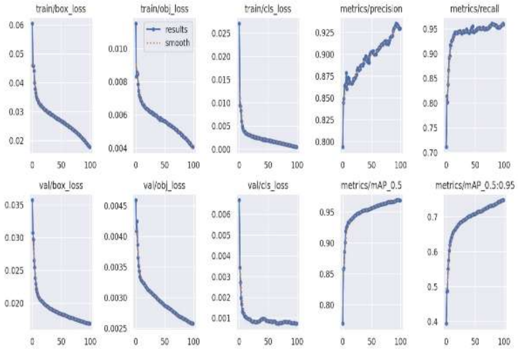
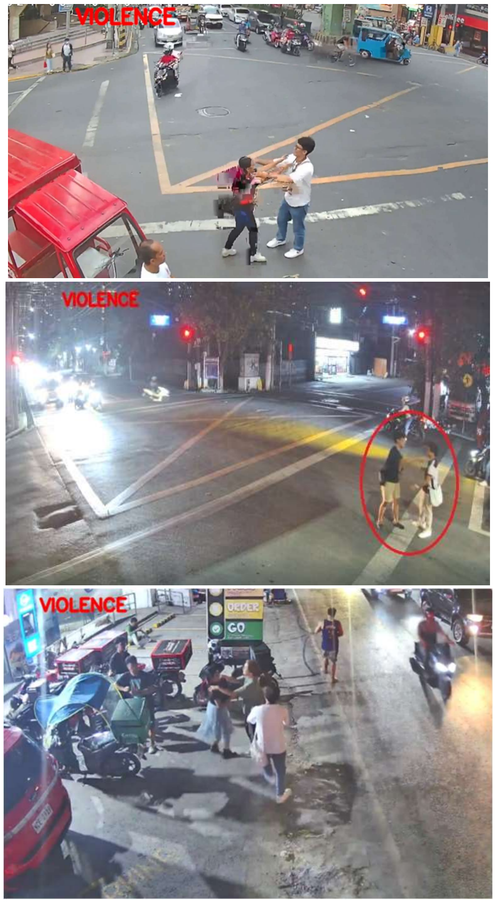


Figure 6: Training and Evaluation Results Chart

Over 100 epochs, the model demonstrated significant improvements in training and validation metrics. Training box loss decreased from 0.06 to 0.02, object loss from 0.010 to 0.004, and classification loss from 0.025 to nearly 0, indicating enhanced accuracy in predicting bounding boxes, detecting objects, and classifying them correctly. Precision improved from 0.800 to 0.925 and recall from 0.70 to 0.95. Validation metrics showed similar trends, with box loss decreasing from 0.035 to 0.020, object loss from 0.0045 to 0.0025, and classification loss from 0.006 to 0.001. The mean average precision (mAP@0.5) increased from 0.80 to 0.95, while mAP@0.5:0.95 rose from 0.4 to 0.7, reflecting ongoing improvements in performance across various IoU thresholds**.**

b. Video and Image Detection



Figures:7-9: Aggressive Behavior Detection

Table 1. Taft Avenue - P. Ocampo St. Camera Detection Trials

		AGGRESSIVE BEHAVIOR
	DISTANCE (m)	STATUS
TRIAL 1	3.00m	DETECTED
TRIAL 2	3.40m	DETECTED
TRIAL 3	3.70m	DETECTED
TRIAL 4	4.00m	DETECTED
TRIAL 5	4.40m	DETECTED
TRIAL 6	4.70m	DETECTED
TRIAL 7	5.00m	DETECTED
TRIAL 8	5.40m	DETECTED
TRIAL 9	5.70m	DETECTED
TRIAL 10	6.00m	DETECTED
TRIAL 11	6.40m	DETECTED
TRIAL 12	6.70m	DETECTED
TRIAL 13	7.00m	DETECTED
TRIAL 14	7.40m	DETECTED
TRIAL 15	7.70m	DETECTED
TRIAL 16	8.00m	DETECTED
TRIAL 17	8.40m	DETECTED
TRIAL 18	8.70m	DETECTED
TRIAL 19	9.00m	DETECTED
TRIAL 20	9.40m	DETECTED
TRIAL 21	9.70m	DETECTED
TRIAL 22	10.00m	DETECTED
TRIAL 23	10.40m	UNDETECTED
TRIAL 24	10.70m	UNDETECTED
TRIAL 25	11.00m	UNDETECTED
TRIAL 26	11.40m	UNDETECTED
TRIAL 27	11.70m	UNDETECTED
TRIAL 28	12.00m	UNDETECTED
TRIAL 29	12.40m	UNDETECTED
TRIAL 30	12.70m	UNDETECTED

Table 2. P. Ocampo St. – A. Mabini St. Camera Detection Trials

		AGGRESSIVE BEHAVIOR
	DISTANCE (m)	STATUS
TRIAL 1	3.00m	DETECTED
TRIAL 2	3.40m	DETECTED
TRIAL 3	3.70m	DETECTED
TRIAL 4	4.00m	DETECTED
TRIAL 5	4.40m	DETECTED
TRIAL 6	4.70m	DETECTED
TRIAL 7	5.00m	DETECTED
TRIAL 8	5.40m	DETECTED
TRIAL 9	5.70m	DETECTED
TRIAL 10	6.00m	DETECTED
TRIAL 11	6.40m	DETECTED
TRIAL 12	6.70m	DETECTED
TRIAL 13	7.00m	DETECTED
TRIAL 14	7.40m	DETECTED
TRIAL 15	7.70m	DETECTED
TRIAL 16	8.00m	DETECTED
TRIAL 17	8.40m	DETECTED
TRIAL 18	8.70m	DETECTED
TRIAL 19	9.00m	DETECTED
TRIAL 20	9.40m	DETECTED
TRIAL 21	9.70m	DETECTED
TRIAL 22	10.00m	DETECTED
TRIAL 23	10.40m	UNDETECTED
TRIAL 24	10.70m	UNDETECTED
TRIAL 25	11.00m	UNDETECTED
TRIAL 26	11.40m	UNDETECTED
TRIAL 27	11.70m	UNDETECTED
TRIAL 28	12.00m	UNDETECTED
TRIAL 29	12.40m	UNDETECTED
TRIAL 30	12.70m	UNDETECTED

Table 3. P. Ocampo St. - F.B. Harrison St. Camera Detection Trials

		AGGRESSIVE BEHAVIOR
	DISTANCE (m)	STATUS
TRIAL 1	4.00m	DETECTED
TRIAL 2	4.25m	DETECTED
TRIAL 3	4.50m	DETECTED
TRIAL 4	4.75m	DETECTED
TRIAL 5	5.00m	DETECTED
TRIAL 6	5.25m	DETECTED
TRIAL 7	5.50m	DETECTED
TRIAL 8	5.75m	DETECTED
TRIAL 9	6.00m	DETECTED
TRIAL 10	6.25m	DETECTED
TRIAL 11	6.50m	DETECTED
TRIAL 12	6.75m	DETECTED
TRIAL 13	7.00m	DETECTED
TRIAL 14	7.25m	DETECTED
TRIAL 15	7.50m	DETECTED
TRIAL 16	7.75m	DETECTED
TRIAL 17	8.00m	DETECTED
TRIAL 18	8.25m	DETECTED
TRIAL 19	8.50m	DETECTED
TRIAL 20	8.75m	DETECTED
TRIAL 21	9.00m	DETECTED

TRIAL 22	9.25m	DETECTED
TRIAL 23	9.50m	DETECTED
TRIAL 24	9.75m	DETECTED
TRIAL 25	10.00m	DETECTED
TRIAL 26	10.25m	UNDETECTED
TRIAL 27	10.50m	UNDETECTED
TRIAL 28	10.75m	UNDETECTED
TRIAL 29	11.00m	UNDETECTED
TRIAL 30	11.25m	UNDETECTED

The figure and tables show that the camera at Taft Avenue and P. Ocampo St. has a detection range of 12.60 meters. The camera at P. Ocampo St. and F.B. Harrison St. offers a maximum detection range of 11.05 meters, while the camera at the intersection of P. Ocampo St. and A. Mabini St. has a maximum detection range of 8.53 meters. These cameras demonstrate advanced capabilities in detecting aggressive behavior both during the day and in low-light nighttime conditions. The figures highlight the system's resilience and accuracy in challenging environments, underscoring its robust performance in real-time incident detection and monitoring. The clear demonstration of the camera's effectiveness in capturing critical details under adverse lighting conditions emphasizes its reliability and effectiveness in enhancing public safety.

c. Alert System Notification and Live Feed

Table 4. Delay in Video Processing at Different Resolutions and Frame Rates

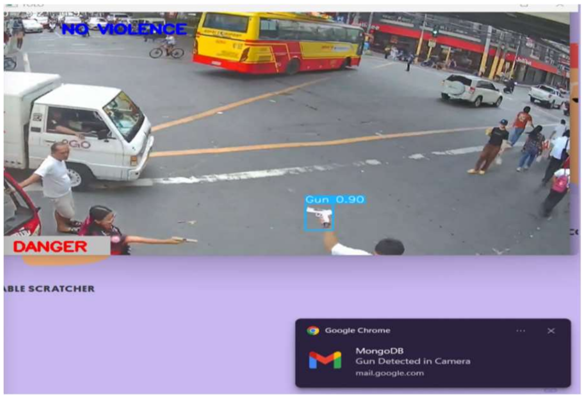
Parameter	**Values**		**Delay (ms)**		
			Video Stream	**ML Output**	**ML Video Stream**
Capture Resolution	2560x1440	15fps	1000	70	7000

The table shows that capturing video at higher resolutions with moderate frame rates causes a 1000 ms delay. Machine learning processing is quicker, with only a 70 ms delay. However, when applied to video streams, machine learning shows a significant 7-second delay due to its complexity.

Table 5. Real-Time Event Processing Latency

Real-time Event	**Data Sent**	
	ML-to- DB	**DB-to-ML**
	40 ms	800 ms

The table presents the latency in milliseconds for different stages of processing real-time events in a system, precisely the period taken for data to be sent from the machine learning (ML) to the database (DB) with 40 ms and then from the database to email notifications with 800 ms.



MongoDB
Gun Detected in Camera
mail.google.com

Figure 10: Visual and Audio Alert using Push Notification

The figure represents the alert system operation through email notifications and push notifications, which include an audio alert accompanying the pop-up notification.

VIII. CONCLUSION

The TANAW system is a real-time monitoring and detection specifically designed for identifying aggressive behavior, utilizing the YOLOv5 algorithm for detection and an alert system for notifications. Hardware components include ESP32, IP Camera C320WS, and a power management system. The YOLOv5 algorithm effectively identifies behaviors such as kicking, punching, hairpulling, and strangling, ensuring prompt notifications to relevant authorities. Monitoring is facilitated through a website hosted on AWS and MongoDB cloud service.

Testing involved collaboration with traffic enforcers and barangay officials, following ISO 9126 standards. Evaluation metrics, including the confusion matrix, F1 confidence curve, precision-recall curve, and training/validation metrics, demonstrate the system's accuracy and improvement over time. The system achieved high precision (0.929), recall (0.959), mAP@50 (0.969), and mAP@50-95 (0.884).

Camera detection ranges were analyzed, with cameras at Taft Avenue and P. Ocampo St., P. Ocampo St. and F.B. Harrison St., and P. Ocampo St. and A. Mabini St. having detection ranges of 12.60 meters, 11.05 meters, and 8.53 meters, respectively. TANAW proves effective for community use, enhancing public safety by accurately detecting aggressive behaviors in both daytime and nighttime conditions.

ACKNOWLEDGEMENT

The proponents extend their deepest gratitude to Engr. Mark P. Melegrito, whose guidance and support throughout the study have been vital in the quality of the study. He has been their wisdom and source of encouragement.

Heartfelt appreciation to the faculty members of the ECE department, particularly to the panel chair, Engr. Jay Fel C. Quijano, and panel members Engr. Cherry G. Pascion and Engr. Gilfred Allen Madrigal. Their valuable comments, suggestions, and advice helped the project to improve.

The Barangay 719 chairman, kagawad, and staff deserve to be thanked for their constant help and support. The proponents are genuinely grateful for the unwavering assistance and cooperation of the barangay during the deployment stage of the study.

Lastly, the proponents extend their deepest appreciation to their families and friends who have shown them emotional and financial support. They motivate them and are simply there to support them throughout the completion of the project study.

Above all, they are grateful for God's divine guidance and blessings. He granted them strength, wisdom, and perseverance throughout this journey.

VIII. REFERENCES

Consunji, R., & Hyder, A. A. (2004). The burden of injuries in the Philippines: implications for national research policy. Accident Analysis & Prevention, 36(6), 1111–1117. https://doi.org/10.1016/j.aap.2004.05.002

Liyanage P., Fernando, P. (2021). Suspicious Human Crowd Behaviour Detection – a transfer learning approach. (2021, December 2). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9774784

Kim, D., Kim, H., Mok, Y., & Paik, J. (2021). Real-Time surveillance system for analyzing abnormal behavior of pedestrians. Applied Sciences, 11(13), 6153. https://doi.org/10.3390/app11136153

Hassner, T., Itcher, Y., & Kliper-Gross, O. (2012). Violent flows: Real-time detection of violent crowd behavior. 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. doi:10.1109/cvprw.2012.6239348

Sumon, S. A., Shahria, M. T., Goni, M. R., Hasan, N., Almarufuzzaman, A. M., & Rahman, R. M. (2019). Violent Crowd Flow Detection Using Deep Learning. Lecture Notes in Computer Science, 613–625. doi:10.1007/978-3-030 14799-0_53