AlertNow: A Real-Time, Low Cost, Multi Modal Fusion System to Detect Gun Threats in US Schools

Aditya Khurana
Moorestown High School, 26khuranaa@mtps.us

Abstract – **Gun violence is a dangerous epidemic in the United States. Every year, more than 45,000 people die from guns, and the problem is only growing worse, with a 34% increase in gun related violence in the last decade. Technology can play a pivotal role in combating the gun violence crisis in the US schools by deploying advanced computer vision and deep learning models to detect firearms and gunshots in live video surveillance systems. Published literature demonstrates that the current commercially available systems specialize in either gunshot detection using acoustic signal or firearm detection using visual data only. This research hypothesizes that an integrated system combining both acoustic and visual modalities can enhance the accuracy and robustness of a gun detection system. The purpose of this research is to develop a cost effective, reliable, and multi-modal fusion-based prototype solution, to detect gun threats and notify authorities in real-time. AlertNow is developed by integrating CNN model for gunshot detection and YOLOv8 model to detect firearms. This multi-modal prototype system is trained, validated, and tested using extensive datasets to enhance detection accuracy with an ability to trigger real-time alerts for timely intervention by security staff in critical situations. The results indicate that AlertNow system estimated performance is superior to other approaches used to gun detect threats with an overall precision, recall and F1 score greater than 99%. The integrated system also displayed a remarkable low latency across all notification delivery channels and therefore making it a viable solution for real-time alerts.**

Index Terms - Firearm detection, gunshot detection, gun threat, multi-modal fusion system

Introduction

Gun violence is a dangerous epidemic in the United States. Every year, more than 45,000 people die from guns, and the problem is only growing worse, with a 34% increase in gun related violence in the last decade. During the initial year of the COVID-19 pandemic, when educational institutions were closed, a substantial decrease in gun violence occurrences on school premises was observed. However, subsequent years have witnessed a marked escalation in school shooting incidents, with at least 83 reported cases in 2024 [1]. The consequences of gun violence impact not only the victims, but also affect families, communities, and the nation. About 25% of children witness gun violence in their schools, homes, or neighborhoods, causing trauma that can last a lifetime with devastating consequences; the economic toll of this problem is staggering, and costs the nation $557 billion each year [2].

Technology can play a pivotal role in combating the gun violence crisis in the schools. During the September 2024 Apalachee High School shooting in Georgia, the school's recently installed CrisisAlert system provided crucial time to prepare and protect students and staff before the shooter initiated the attack. This alert system featured a discreetly activated button, which promptly alerted administrators and local law enforcement to the precise location of the ongoing emergency [3]. Other commercially available gun detection technologies such as ZeroEyes and ShotSpotter, can either detect gunshot audio or guns from a live camera feed, but not both. Furthermore, with each unit costing nearly $100,000 USD, these systems cost millions of dollars to scale and still require continuous human monitoring to deploy law enforcement officials to the incident site.

The purpose of this research was to develop a cost effective, scalable, reliable, and an advanced technology-based prototype system, named AlertNow, that can promptly detect gunshot and firearm in real-time and notify security personnel and local law enforcement to prevent gun threats particularly in school environments. Integrating Deep Learning and Computer Vision algorithms into existing surveillance systems like Closed Circuit television (CCTV) presented a potential to overcome the limitations of traditional surveillance camera systems that require manual oversight to monitor multiple video feeds and misses on proactive detection, prevention, and prompt notification. AlertNow was developed by integrating Convolutional Neural Networks (CNN) model for audio analysis of gunshot sounds using 2-Dimensional (2D) spectrograms, and You Only Look Once version 8 (YOLOv8) model for video analysis to detect firearms in live camera feeds from CCTV. This multi-modal prototype system was trained, validated, and tested using extensive datasets that included both gunshot audio and firearm images to enhance detection accuracy with an ability to trigger real-time alerts and to ensure timely intervention by security staff in critical situations.

Literature Review

CCTV surveillance technology is widely utilized by law enforcement agencies and both public and private organizations. This system comprises multiple video cameras connected in a closed circuit, transmitting footage either to a centralized monitor or recording it via a wireless, remote network. Such technology enhances the likelihood of offender apprehension and enables law enforcement or security personnel to respond swiftly to incidents [4]-[5]. However, existing literature indicates that CCTV systems may primarily be effective in addressing drug-related, property, and vehicle crimes [6]. Artificial intelligence has the potential to augment CCTV systems by integrating automated firearm detection algorithms, enabling continuous monitoring with or without human intervention. By equipping CCTV systems with real-time detection capabilities driven by computer vision algorithms, these systems can actively monitor for firearm-related incidents, significantly improving their effectiveness in preventing gun violence related potential injuries and casualties [7]. Computer vision scholars have used artificial intelligence to train algorithms to detect firearms in real-time [8]-[11]. Researchers have highlighted the potential of artificial intelligence to support the criminal justice system by presenting a multidisciplinary approach that integrates criminological frameworks with computer vision techniques. This approach aims to address the gun violence epidemic in the United States effectively [12]. Earlier firearm detection may prompt law enforcement to respond to a potentially dangerous incident through improved response times to neutralize the threat and limit the number of casualties.

Deep learning, a subset of machine learning, leverages neural networks to identify patterns in visual data by analyzing large datasets. It is commonly applied to tasks such as object detection and often employs CNN to process and interpret visual information, such as images [13]. CNN enable object detection algorithms to learn and identify patterns in data, facilitating tasks such as object classification and localization within images or videos. Among state-of-the-art object detection algorithms, YOLO is widely used due to its high speed and accuracy, making it particularly suitable for automated firearm detection applications [14]. Numerous researchers have trained datasets using YOLO and CNN algorithms to evaluate their performance in firearm detection within images and videos. Studies have consistently shown that various versions of YOLO outperform CNN and other algorithms in terms of detection speed and accuracy. This superiority underscores YOLO's reliability and effectiveness for firearm detection in CCTV applications [7]-[8]. Furthermore, the YOLO algorithm demonstrates the capability to detect armed individuals across diverse conditions, including low-light environments, varied backgrounds, and different image qualities, with high precision and recall. This makes it particularly well-suited for integration into CCTV systems monitoring challenging or unfavorable environments [15].

Many researchers have trained their YOLO algorithms on still images and utilized still images for firearm detection in live video feed from CCTV systems. Data was collected from various publicly available sources which includes various images of different firearms (i.e., handguns, rifles, and shotguns), images from surveillance footage, firearms used in various school shootings, publicly available photographs on social media, news media, movies, and TV shows. The dataset is generally split into 80% for training, 10% for validation, and 10% for testing. The model's performance is assessed by evaluating metrics such as precision, recall, F1-score, average precision, and mean average precision specifically for the firearm class [12].

The application of CNN in sound classification has been a topic of increasing interest in the research community for classifying environmental sounds and using CNN to detect speech sounds, highlighting the networks' capability to generalize across diverse acoustic conditions. These studies demonstrate the adaptability and effectiveness of CNN in handling diverse sound classification tasks, paving the way for more advanced audio recognition systems [16]-[17]. The feasibility of using CNN models and sound processing techniques for real-time gunshot detection with minimal cost and computational requirements has been evaluated by the researchers. Both one-dimensional and two-dimensional CNN models were trained on audio data and corresponding spectrograms to identify gunshots. Testing revealed that a majority-vote ensemble combining the one-dimensional and two-dimensional models achieved the best performance, exceeding 99% accuracy on validation data and in differentiating gunshots from fireworks. Other models utilized Precision, Recall, F1-score, Area Under Curve (AUC) and Intersection over Union (IoU) as key evaluation metrics during training and validation sessions [18].

Since existing literature demonstrate that most current systems specialize in either sound-based gunshot detection using acoustic signal (e.g., gunshot spectrograms) or firearm detection using visual data (e.g., images or video frames), the integration of both acoustic and visual modalities into a single deep learning model to identify events in real-time is a significant research opportunity which has not been explored yet. The combination of two modalities in real-time could further improve the accuracy, and robustness of gun detection systems that may curb gun related threats more effectively and promptly.

Materials and Methods

AlertNow is a multi-modal gunshot and firearm detection system designed for real-time threat detection in school environments or indoor settings. The system integrates two key components: CNN model for audio analysis to detect gunshot sounds, and a YOLOv8 model for video analysis to detect firearms. Upon detection, AlertNow promptly sends email and SMS alerts to notify security personnel of critical events. Additionally, it updates a real-time dashboard, providing security staff with a live view of ongoing situations, enabling efficient responses to emerging incidents. **Figure I** provide a high-level overview of the system architecture.

The AlertNow prototype system was developed and tested on a 2022 MacBook Pro with an M2 chip, utilizing the built-in microphone and a Logitech USB microphone with a frequency range of 20 Hz to 20 kHz for audio input, and the MacBook's FaceTime 1080p HD camera operating at 45 fps for video input. The software stack comprised TensorFlow, PyTorch, and Ultralytics for machine learning; Librosa, SciPy, and PyAudio for audio processing; OpenCV for image and video processing; and NumPy for data manipulation. Backend and API functionalities were handled by Django, Django REST Framework, and Gunicorn, while Twilio and smtplib were used for notifications. The methodology consists of five components: dataset selection, model architecture selection, model training and evaluation, system alert notification, and system integration followed by an overview of performance evaluation metrics for each firearm (video) detection and gunshot (audio) algorithms.

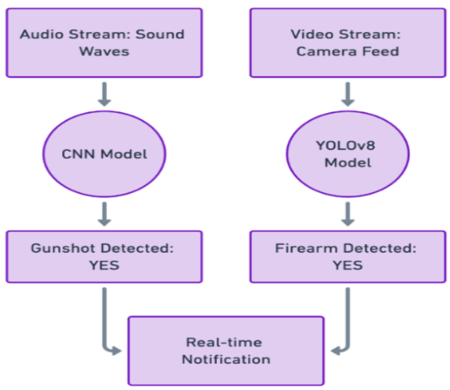

FIGURE I
HIGH-LEVEL OVERVIEW OF ALERTNOW SYSTEM

I. Video Analysis Using YOLOv8

A pre-augmented dataset of 9,897 annotated images from Roboflow and Kaggle was used for training, categorized into 5-gun types with dataset distribution summarized in **Table I**. An additional unaugmented dataset of 5,162 images was used for validation and testing. Handguns were given priority in the dataset distribution, highlighting their significant role in mass shootings across the United States [19].

TABLE I
CATEGORICAL BREAKDOWN OF THE IMAGE DATASET

Category	# of pre-augmented Training Dataset	# of unaugment Validation Dataset	# of unaugment Testing Dataset
Handgun	2968	774	774
Rifle	2475	645	645
Shotgun	1979	516	516
Pistol	1484	387	387
Machine gun	991	259	259

After evaluating several models including EfficientNet and ResNet50, YOLOv8 was chosen for video detection due to its superior performance in real-time applications. YOLOv8 demonstrated high precision, ease of use, speed, and accuracy, making it optimal for the firearm detection task. Video frames were extracted and adjusted to maintain uniform lighting and quality. The images were resized to meet the model's specifications, and pixel values were normalized to a [0, 1] range for standardization [20]-[21]. Annotations were labeled in YOLO format, defining precise bounding boxes for object detection.

The YOLOv8 model was trained using a pre-augmented dataset with various image transformations. Training parameters included a learning rate of 0.001, batch size of 16, and 100 epochs. Validation was performed during training to evaluate model performance and prevent overfitting. The model's final performance was tested on a separate unaugmented dataset **(Table I)**. YOLOv8 detected objects by dividing an image into grid cells and predicting bounding boxes and class probabilities for each cell. The algorithm calculated an intersection over union (IoU) score to determine the accuracy of predicted bounding boxes compared to ground truth data. A confidence score for class probability was also provided for each detected firearm object.

II. Audio Analysis Using CNN

The dataset comprised of 10,385 sound samples (up to 4 seconds each) from Kaggle, with 40% gunshot and 60% non-gunshot sounds. The audio dataset categorization is summarized in **Table II.** The dataset was split into 80% for training, 10% for validation, and 10% for testing.

TABLE II
CATEGORICAL BREAKDOWN OF THE AUDIO DATASET

Category	Training Dataset	Validation Dataset	Testing Dataset
Gunshot	3323 (32%)	415 (4%)	415 (4%)
Non-Gunshot (Fireworks, jackhammer, machinery)	4985 (48%)	623 (6%)	624 (6%)

For audio detection, CNN was selected over alternatives like Long Short-Term Memory (LSTM) networks and SoundNet. CNN showed superior performance in pattern recognition, effectively processing audio signals converted to spectrograms. This approach allowed for more straightforward implementation, faster training times, and better scalability across various hardware platforms. Audio samples were standardized to a 2-second duration and converted into spectrograms for visual representation as shown in **Figure II**. Data augmentation techniques included pitch alteration (75% to 125% range) and background noise addition to simulate real-world conditions. These transformations aimed to improve the model's ability to generalize across different acoustic environments.

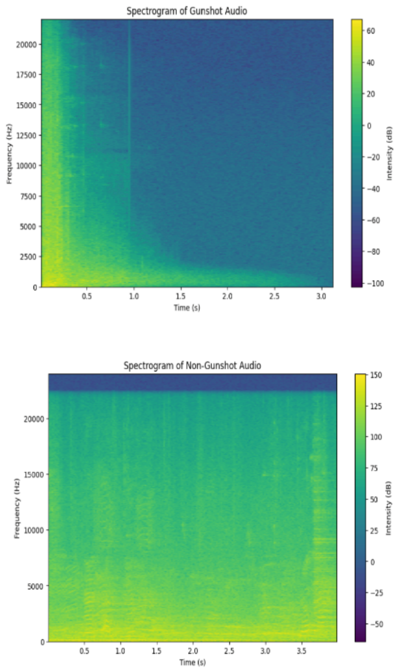

FIGURE II
SPECTOGRAM REPRESENTATION OF GUNSHOT AND NON-GUNSHOT AUDIO

The CNN model was trained on the prepared audio spectrograms. The dataset was split into training, validation, and testing subsets (**Table II**). Early stopping and learning rate scheduling techniques were employed to optimize training and prevent overfitting. The model's final performance was evaluated on the testing dataset to measure its detection ability on unseen data. The CNN model processes audio input as spectrograms, which are 2D representations of the audio signal. Convolutional layers detect local features in the spectrogram, followed by pooling layers that downsample the feature maps. Fully connected layers then learn complex patterns from the extracted features. The trained model outputs a binary classification (gunshot vs. non-gunshot) using a sigmoid activation function.

III. System Alert Notification

The notification system consisted of a detection module, a notification module, and external delivery services. When a detection confidence surpassed the 80% threshold, the system triggered alerts via SMS, email, and dashboard updates. Twilio was used for SMS and Python's smtplib for email alerts. The system maintained logs for traceability and could send notifications to predefined recipients at schools and local law enforcement. To validate the notification system, its reliability, speed, and accuracy were tested under various scenarios, including simulated gunshot detections and network outages. Notification latency was measured, calculating the time from detection to alert delivery. The notification latency rate was used as a key metric to assess delivery performance of the system.

IV. AlertNow System Integration

The individual components of video analysis, audio analysis, and system alert notification, as shown in **Figure III**, are integrated into AlertNow system for comprehensive threat detection.

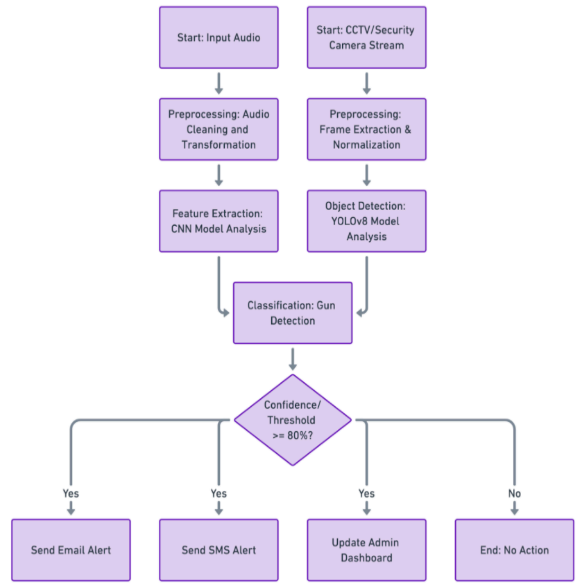

FIGURE III
DETAILED TECHNICAL FLOWCHART OF AUDIO AND VIDEO STREAM AND NOTIFICATION PROCESS

For video detection, OpenCV captured live video streams from CCTV cameras, which were analyzed by the YOLOv8 model implemented with PyTorch and Ultralytics. **Figure IV** illustrates the AlertNow's firearm detection system video stream real-time monitoring process.

FIGURE IV
ALERTNOW FIREARM DETECTION SYSTEM ILLUSTRATION

The audio detection process utilized PyAudio for capturing audio, Librosa for audio processing, and TensorFlow for running the pre-trained CNN model on the captured audio data. **Figure V** illustrates Alert Now's gunshot detection system audio stream processing, analysis, and notification process.

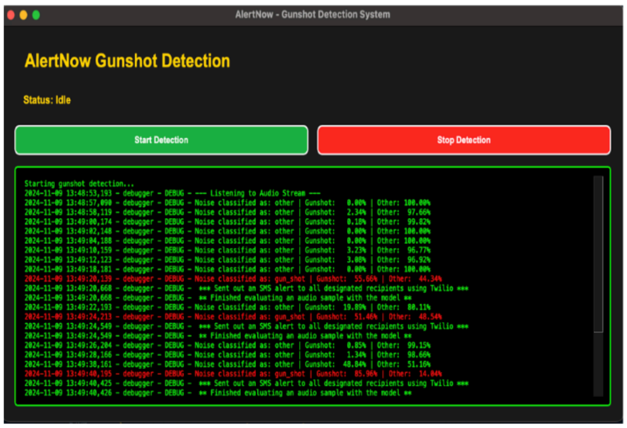

FIGURE V
ALERTNOW GUNSHOT DETECTION SYSTEM ILLUSTRATION

Both detection processes ran concurrently, analyzing incoming data in real-time. When either model detected a threat or met the confidence threshold, the system triggered the alert mechanism as detailed in AlertNow System Integration section. The proposed AlertNow system's integrated performance was derived from the theoretical framework presented by Kittler *et al.* [22] which is a well-regarded work in the field of classifier combination technique, providing a methodology to calculate the combined theoretical performance of the multi-modal system. The proposed integrated model combined estimated (theoretical) probability was calculated using basic probability theorem with following equation:

$$P_{Combined} = 1 - (1-P_{YOLOv8}) * (1-P_{CNN}) \quad (1)$$

where P_{YOLOv8} is the Probability of firearm detection by proposed YOLOv8 and P_{CNN} is the Probability of gunshot detection by proposed CNN.

V. Performance Evaluation Metrics

The performance of YOLO and CNN deep learning models was evaluated using several metrics [23]. Accuracy measured the ratio of correct predictions to total predictions, particularly relevant for CNN's binary classification tasks like gunshot detection. Precision (P) calculated the ratio of true positives to total positive predictions, while Recall (R) determined the ratio of true positives to actual positives. F1-Score provided a balanced measure of precision and recall. For object detection, Intersection over Union (IoU) assessed the overlap between predicted and actual bounding boxes. Average Precision (AP) and Mean Average Precision (mAP) evaluated bounding box accuracy and performance across multiple firearm classes. Inference Speed, measured in Frames Per Second (FPS), to assess YOLOv8 model's real-time processing capability in video streams. The integrated system's Notification Latency Rate measured alert delivery time post-detection. **Table III** summarizes the key performance evaluation metrics used for each of the firearm and gunshot detection models.

TABLE III
KEY PERFORMANCE EVALUATION METRICS FOR GUNSHOT AND FIREARM DETECTION MODELS

Performance Evaluation Metrics	Gunshot Detection Model (CNN)	Firearm Detection Model (YOLOv8)
Accuracy	✓	-
Precision	✓	✓
Recall	✓	✓
F1-Score	✓	✓
Intersection over Union (IoU)	-	✓
Average Precision (AP)	-	✓
Mean Average Precision (mAP)	-	✓
Inference Speed/Frames per Second (FPS)	-	✓
Notification Latency Rate	✓	✓

Results

This section presents the performance results of the YOLOv8 model for firearm detection, the CNN model for gunshot detection, the integrated AlertNow system, and the notification system.

I. YOLOv8 Model Performance

The proposed YOLOv8 model demonstrated a strong performance in firearm detection, achieving a precision of 94.8%, a recall of 90.3%, and an F1-score of 92.5%. The mean average precision (mAP) of 93.4% indicated high accuracy in identifying firearms, with an Intersection over Union (IoU) of 85.8% showcasing effective object localization. These metrics collectively highlight the model's ability to detect firearms accurately with minimal false positives, while the high recall rate reflects fewer missed detections, a critical factor in scenarios requiring rapid threat identification. Furthermore, the model's inference speed of 45 frames per second (FPS) confirmed its suitability for real-time deployment in fast-response environments.

Based on historical data of firearms commonly used in U.S. mass shootings, the model was evaluated across five firearm categories - handgun, rifle, shotgun, pistol, and machine gun. The handgun category achieved the highest average precision (AP) of 95.6%, followed closely by rifles (95.0%) and shotguns (93.4%). Pistols and machine guns showed slightly lower APs of 91.8% and 91.2%, respectively, potentially due to fewer instances in the training dataset as shown in **Table IV**. The mean average precision (mAP) at a 0.5 Intersection over Union (IoU) threshold was consistently 93.4% across all categories. The 0.5 IoU threshold, a standard in object detection, ensures at least 50% overlap between predicted and ground truth bounding boxes, balancing detection accuracy and localization.

TABLE IV
YOLOv8 MODEL AVERAGE PRECISION RESULTS FOR EACH GUN CATEGORY

Gun Category	Number of Images	AP (%)	mAP (%) @ 0.5 IoU
Handgun	774	95.6%	93.4%
Rifle	645	95.0%	
Shotgun	516	93.4%	
Pistol	387	91.8%	
Machine gun	259	91.2%	

Figure VI visually depicts and compares the precision, recall, and F1 score achieved by the YOLOv8 model for each gun category tested. These results highlight the model's robustness in detecting and localizing diverse firearm types effectively under real-world conditions.

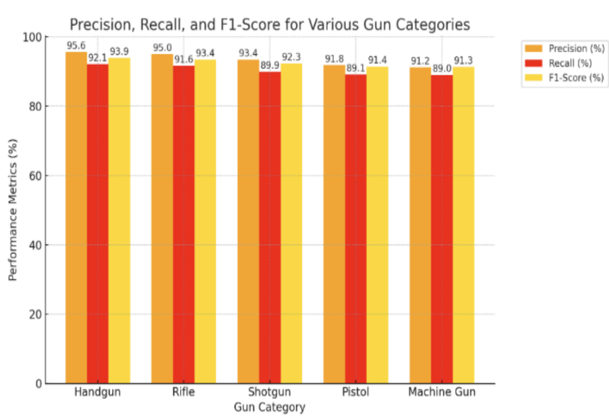

Figure VI
YOLOv8 MODEL PERFORMANCE METRICS COMPARISON FOR EACH GUN CATEGORY

When compared to existing literature utilizing similar methodology and datasets, as shown in **Table V**, our results demonstrated that the YOLOv8 model outperformed previous YOLO model versions in terms of Recall and F1-score for firearm detection. Even though the precision value of YOLOv8 appears to be lower than YOLOv5 model, greater priority was given to a higher recall value since missing an object was unacceptable when it comes to firearm detection. Additionally, to balance both precision and recall in an optimal way, F1-score was also given more weightage while evaluating firearm detection model performance. We leveraged a pre-augmented dataset to train the model and then used pre-processed images during validation and testing phase that led to overall improvement of proposed model performance.

TABLE V
RESULTS COMPARISON OF PROPOSED YOLOv8 MODEL WITH OTHER MODELS IN FIREARM DETECTION APPLICATION

Performance Evaluation Metrics	YOLOv3 Model Results [24]	Scaled YOLOv5 Model Results [8]	YOLOv7 Model Results [12]	Proposed YOLOv8 Model Results
Precision (%)	85	99.5	94.6	94.8
Recall (%)	81	84.6	85.6	90.3
F1-Score (%)	83	91.4	89.9	92.5
mAP@0.5 (%)	-	-	94.7	93.4

II.CNN Model Performance

The proposed CNN model classified gunshots from the testing dataset demonstrated an overall accuracy rate of 98.7%, precision of 97.4%, recall rate of 95.6% and F1-score of 98.2%. Predictions performed on our test data confirmed that the model was able to classify gunshots with relatively high precision. Overall, these metrics highlighted the model's ability to accurately detect gunshot with few false positives in the testing environment which included both gunshot and non-gunshot noises. Incorporation of non-gunshot sounds (i.e., fireworks, jackhammer, machinery) in the training dataset substantially reduced false positive rate by distinguishing gunshots from fireworks and thus leading to improved model performance.

When compared to existing Keras models [18] trained with different training datasets**,** our proposed CNN model performed similarly with minor differences in various metrics, as shown in **Table VI**. The proposed model appears to trade off accuracy and recall for a better overall F1-Score indicating a good balance between precision and recall and suggesting that the proposed model might be more balanced in its predictions for gunshot detection. While the proposed CNN model didn't outperform the Keras models in every metric, it showed competitive performance, particularly in terms of F1-Score. The selection of appropriate models would depend on whether higher accuracy, better recall, or a more balanced performance, as indicated by F1-Score, is deemed important in a particular application.

TABLE VI

RESULTS COMPARISON OF PROPOSED CNN MODEL WITH OTHER MODELS IN GUNSHOT DETECTION APPLICATION

Performance Evaluation Metrics	**Morehead et.al. [18] KERAS Model Results**				**Proposed CNN Model Results**
	1D CNN	**2D CNN (64)**	**2D CNN (128)**	**CNN Ensemble**	
Accuracy (%)	99.4	99.4	99.4	99.5	98.7
Precision (%)	98.0	97.1	97.4	97.9	97.4
Recall (%)	96.6	97.6	97.6	98.0	95.6
F1-Score (%)	97.3	97.4	97.5	97.9	98.2

III. AlertNow Integrated System Performance

The multi-modal fusion approach based AlertNow system performance was assessed by combining the performance of YOLOv8 and CNN models. **Figure VII** visually demonstrates performance evaluation metrics results of both firearm (YOLOv8) and gunshot detection (CNN) models giving a comprehensive view of the capabilities of an integrated system.

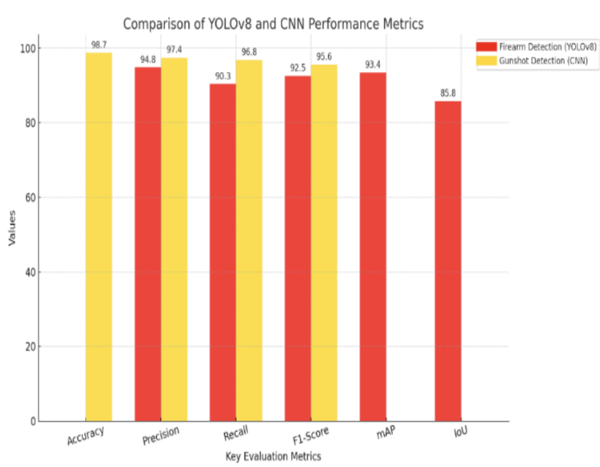

FIGURE VII

YOLOv8 AND CNN PERFORMANCE EVALUATION METRICS RESULTS

The common performance metrics of Precision, Recall, and F1-Score from both models were used to calculate the combined estimated theoretical performance of the integrated AlertNow system, based on basic probability theory referenced in the Methodology section. **Table VII** illustrates performance comparison of the combined integrated system with the individual YOLOv8, and CNN models results. Equal performance weightage was assumed when calculating combined system performance, based on the results achieved for individual models' performance.

TABLE VII

PERFORMANCE COMPARISON OF INTEGRATED SYSTEM WITH INDIVIDUAL MODEL RESULTS

Performance Evaluation Metrics	**Proposed YOLOv8 Model Results** (P_{YOLOv8})	**Proposed CNN Model Results** (P_{CNN})	**Estimated Theoretical Combination Results** $P_{Combined} = 1 - (1-P_{YOLOv8}) * (1-P_{CNN})$
Precision (%)	94.8	97.4	99.8
Recall (%)	90.3	95.6	99.5
F1-Score (%)	92.5	98.2	99.8

The integrated system estimated theoretical results demonstrated a significant performance enhancement, leading to improved overall detection of gunshots and firearms when compared to individual model performance results. However, rigorous empirical testing in the real-world is essential to validate these results, confirm system reliability and assess practical applicability in a school environment.

IV. Notification System Performance

The AlertNow system's notification process was tested using 3,620 samples (2,581 images and 1,039 audio files). The system demonstrated low latency across all delivery channels, with average notification times of 800ms for email, 2000ms for SMS, and 250ms for dashboard updates. Minimum latencies were 500ms, 1200ms, and 100ms, while maximum latencies were 1300ms, 4500ms, and 450ms for email, SMS, and dashboard updates, respectively. Notably, the system achieved 100% delivery rate when the detection confidence level met or exceeded the 80% threshold, with no false alerts triggered below this threshold. These results indicate the system's ability to consistently deliver real-time alerts. However, external factors such as network conditions and server load may influence actual delivery times in real-world deployments.

Conclusions

The proposed AlertNow system demonstrates that a multi-modal fusion based deep learning approach can be effectively integrated into existing school surveillance systems to provide accurate, cost-effective, real-time gun detection and prompt alerts for timely intervention to reduce firearm related harm for youth in the K-12 school communities. By utilizing a CNN model for audio analysis of gunshot sounds and a YOLOv8 model for video analysis, AlertNow effectively detects firearms. The system has been trained, validated, and tested using comprehensive datasets to optimize detection accuracy. Performance comparison of integrated AlertNow model with individual YOLOv8 and CNN models, as well as other weapon detection systems, show that AlertNow outperforms these methods in terms of precision, recall, and F1-score, achieving over 99% in all metrics. Additionally, the integrated system demonstrated a low latency across all notification channels, further establishing its potential as a reliable, real-time solution for security alerts. These results highlight AlertNow's promise as a robust tool in enhancing school safety and can serve as a model for similar applications in other high-risk environments.

AlertNow is developed with the premise of leveraging existing CCTV systems equipped with omnidirectional microphones, extended pickup range, and synchronized audio-video recording. A typical security camera microphone normally captures sound within a 40-foot radius with video/audio performance and coverage is dependent on camera specifications, lighting, microphone sensitivity, and environmental acoustics. For a reliable threat detection system, range limitations of both video and audio components should be carefully assessed.

Future Work

While the AlertNow system has demonstrated promising performance, further evaluation in real-world scenarios is necessary. Integrating the gun threat detection system with live CCTV streams within an operational school surveillance environment will offer a more realistic and comprehensive assessment of its reliability and accuracy. This will help to test the system's detection capabilities in real-life situations and address potential challenges related to detection quality and latency in real-time alert notifications. Such empirical testing will be crucial before moving towards full-scale implementation. If the existing CCTV system's coverage is insufficient to detect gunshots over the desired distance, additional microphone or camera replacements may be warranted.

Acknowledgment

I would like to express my sincere gratitude to Mrs. Charu Khurana, Solutions Architect Manager at Amazon Web Services (AWS), for her invaluable mentorship and guidance throughout my research. My heartfelt thanks also go to my family for their unwavering support, encouragement, and belief in my work toward addressing the gun violence epidemic. Lastly, I would like to honor the courage of survivors and the resilience of families affected by gun violence—their strength continues to inspire my commitment to this cause.

References

[1] Matthews, A., O'Kruk, A., and Choi, A. "School shootings in the US: Fast facts," CNN, Nov. 27, 2024. [Online]. Available: https://www.cnn.com/us/school-shootings-fast-facts-dg/index.html

[2] EveryStat, "Gun Violence in the United States," Oct. 2024. [Online]. Available: https://everystat.org/wpcontent/uploads/2020/03/USA_2024-October.pdf

[3] Bailey, C., Mascarenhas, L., I. Rosales, I., *et al.*, "Panic buttons and phone alerts: How technology helped prevent further bloodshed at Apalachee," CNN, Sep. 8, 2024. [Online]. Available: https://www.cnn.com/2024/09/08/us/apalachee-shooting-alert-system-centegix/index.html

[4] Goold, B.J. February 2004. "CCTV and Policing: Public Area Surveillance and Police Practices in Britain." Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199265145.001.0001

[5] Ratcliffe, J., and J.M. Rosenthal, J.M. 2021. "Video Surveillance of Public Places." *ASU Center for Problem-Oriented Policing*. https://popcenter.asu.edu/content/video-surveillance-public- places-2nd- edition

[6] Piza, E.L., Welsh, B.C., Farrington, D.P., *et. al.* 2019. "CCTV Surveillance for Crime Prevention: A 40-Year Systematic Review with Meta-Analysis." *Criminology and Public Policy* 18 (1), pp. 135– 159. https://doi.org/10.1111/1745-9133.12419

[7] Skogan, W.G. 2019. "The Future of CCTV." *Criminology and Public Policy* 18 (1), pp. 161–166. https://doi. org/10.1111/1745-9133.1242

[8] Ashraf, A.H., Imran, M., Qahtani, A.M., *et al.* 2022. "Weapons Detection for Security and Video Surveillance Using CNN and YOLO-V5s." *Computers, Materials and Continua,* 70 (2), pp. 2716–2775. https://doi.org/10.32604/cmc.2022.018785

[9] Narejo, S., Pandey, B., Vargas, D.E., *et al.* 2021. "Weapon Detection Using YOLO v3 for Smart Surveillance Systems." *Mathematical Problems in Engineering*, volume 2021, Article ID 9975700. https://doi.org/ 10.1155/2021/5808206

[10] Bhatti, M.T., Khan, M.G., Aslam, M., *et al.* 2021. "Weapon Detection in Real-Time CCTV Videos Using Deep Learning." *IEEE Access,* volume 9, pp. 34366–34382.

[11] Ahmed, S., Bhatti, M. T., Khan, M. G., *et al.* May 2022. "Development and Optimization of Deep Learning Models for Weapon Detection in Surveillance Videos." *Applied Sciences*, 12(12), 5772. https://doi.org/10.3390/app12125772

[12] Houser, T., McMillan, A., Dong, B. 2024. "Bridging the Gap Between Criminology and Computer Vision: A Multidisciplinary Approach to Curb Gun Violence." *Security Journal*, volume 37, pp. 1409–1429.

[13] LeCun, Y., Bengio, Y., Hinton, G. 2015. "Deep Learning." *Nature*, 521(7553), pp. 436-444.

[14] Wang, C., Bochkovskiy, A., and Liao, H.M. July 2022. "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors." https://arxiv.org/abs/2207.02696

[15] Rosales, M.P., Rodriguez, C., Pomachagua, Y., *et al.* September 2021. "Firearm-Related Event Detection Using Deep Learning." In *Proceedings* of the 2021 International Conference on Machine Learning and Cybernetics (ICMLC).

[16] Takahashi, N., Michael G., Beat P., and Luc V.G. 2016. "Deep Convolutional Neural Networks and Data Augmentation for Acoustic Event Detection." *arXiv* preprint arXiv:1604.07160

[17] Jaiswal, K., and Patel, D. K. 2018. "Sound Classification Using Convolutional Neural Networks." *IEEE International Conference on Cloud Computing in Emerging Markets (CCEM)*, pp. 1-6.

[18] Morehead, A., Ogden, L., Magee, G., et al. 2019. "Low-Cost Gunshot Detection using Deep Learning on the Raspberry Pi." *IEEE International Conference on Big Data*, pp. 3038–3044. https://doi.org/10.1109/BigData47090.2019.9006456

[19] Statistica Research Department, "Guns used in mass shootings in the U.S. 1982-2024." September 11, 2024. [Online]. Available: https://www.statista.com/statistics/476409/mass-shootings-in-the-us-by-weapon-types-used/

[20] Boesch, G. "YOLOv8: A Complete Guide [2025 Update]." Visio.ai, December 18, 2023. [Online]. Available: https://viso.ai/deep-learning/yolov8-guide/

[21] Key Labs, "Under the Hood: YOLOv8 Architecture Explained," December 20, 2023. [Online]. Available: https://keylabs.ai/blog/under-the-hood-yolov8-architecture-explained/

[22] Kittler, J., Hatef, M., Duin, R. P., *et al.*1998. "On Combining Classifiers." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3), pp. 226-239.
[23] Boesch, G. "Precision vs. Recall – Full Guide to Understanding Model Output," Visio.ai, February 3, 2024. [Online]. Available: https://viso.ai/computer-vision/precision-recall/
[24] Garza, J.I., and Vega, C.A."Gun detection in video frames with YOLOv3," December 2021. [Online]. Available: https://www.researchgate.net/publication/356683717_Gun_Detection_in_Video_Frames_with_YOLOv3