BSc (Hons) in Information Technology Specializing Data Science

Research Project - IT4010 **Data Analysis Report**

Project ID: 24-25J-213

Project Title: ManthraX - Pioneering Precision: The Future of Autonomous Mobility

1. Introduction
    1.1. Background:

The rapid development of autonomous vehicles is highly transforming the transportation industry by guaranteeing much safer, more efficient, and environmentally friendly travel. However, achieving this goal presents complex challenges across multiple domains, including perception, decision making, ethical considerations, and passenger comfort.

Autonomous vehicles must accurately perceive and understand their surroundings to make safe driving decisions. This involves detecting and classifying objects, predicting their movements, and understanding the environment's context. Traditional methods often struggle with real-time processing and handling diverse, dynamic environments. Advanced techniques like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs) offer potential solutions but require further research to enhance their robustness and efficiency.

Once the environment is perceived, the vehicle must make decisions that ensure safety, efficiency, and compliance with traffic laws. Current approaches using rule-based systems or simple algorithms are insufficient for complex, real-world scenarios. Reinforcement Learning (RL) and Monte Carlo Tree Search (MCTS) offer promising frameworks for developing sophisticated decision-making models that can learn from experience and adapt to new situations.

Autonomous vehicles must make decisions that align with ethical standards, especially in scenarios involving potential harm. Developing models that incorporate ethical principles is a significant challenge, as it requires balancing conflicting values and ensuring that decisions are transparent and justifiable. Research in this area explores how to encode ethical considerations into machine learning models and how to align these models with societal values.

For autonomous vehicles to be widely accepted, they must ensure passenger comfort and safety. This involves monitoring in-cabin conditions and passenger states to dynamically adjust settings such as temperature, noise levels, and seat positions. Machine learning models that can interpret sensory data and predict passenger needs are essential for achieving this goal. Research in this area focuses on developing adaptive systems that enhance the overall passenger experience.

1.2. Research Problem:

- Enhancing object detection and motion prediction in crowded environments with occlusions.
- Enabling real-time decision-making in unpredictable, unstructured traffic.
- Designing ethically sound systems that balance cultural norms with universal safety standards.
- Improving in-cabin security through advanced image and voice recognition.

1.3. Objectives:

It Number	Objective	Objective number
IT21160448	Perception and Scene Understanding	1
IT21155048	Decision Making and Collision Avoidance	2
IT21162978	Ethical Decision Making in Autonomous Vehicles.	3
IT21174780	Enhancing In-Cabin Security and time Adaptive algorithms for Comfort.	4

2. Data Exploration

2.1. Data Collection

Objective number	Dataset	How data was Collected
1	Carla Object detection	Roboflow: https://universe.roboflow.com/ds/l8x1Cz24Fr?key=nWDBG8LZjq
	Steer with front cam footage (Carla)	Generated in simulation environment using a custom data generator.
2	Simulator Generated Data	Realtime generated using Carla Simulation for lane detection for Reinforcement Learning. Source: Custom environment generation script.
3	Hazard Detection	Roboflow: adult/child https://universe.roboflow.com/ds/e8it4C3ZsC?key=mvcfFLNtpp Cone https://universe.roboflow.com/ds/qt7KGSdT8Y?key=yfmjOfikIX pothole https://universe.roboflow.com/ds/22aieXdXGj?key=dDqWWspG80

		cat, dog https://universe.roboflow.com/ds/pzbRAMnUNg?key=daA3RqdWVg
4	FER	Kaggle: https://www.kaggle.com/datasets/msambare/fer2013 (FER2013)
	Harmful Status	Kaggle: https://www.kaggle.com/datasets/emrahaydemr/gunshot-audio-dataset (Gunshots) https://www.kaggle.com/datasets/afisarsy/enhanced-audio-of-accident-and-crime-detection (Crime audio events) https://www.kaggle.com/datasets/dmitrybabko/speech-emotion-recognition-en (Neutral human speech data) Downloadable sound effects: https://freesound.org/ https://sound-effects.bbcrewind.co.uk/search
	Harmful Objects	Roboflow: https://universe.roboflow.com/crime-detection/guns_n_knives-h4bky (Knives and guns)

2.2. Dataset Description:

Obj. No.	Dataset	Data source	Description	Resource	Size	Key attributes
1	Carla Object detection	Roboflow and CARLA Simulator	Object detection in Carla simulation environment.	Roboflow: https://universe.roboflow.com/ds/l8x1Cz24Fr?key=nWDBG8LZjq	6,000 images	Includes bounding box labels for objects in simulation.
	Steer with front cam footage (Carla)	CARLA Simulator (Semantic Camera)	Synthetic data for steering behavior analysis.	Generated in simulation environment using a custom data generator.	50,000 cropped images	Cropped to focus on road and lane marking. Color coded using the cityscape palette.

2	Simulator Generated Data	Carla Simulation generated data.	Realtime environment generated for lane detection	Custom environment generation script.		Dynamic real-time scenarios for RL tasks.
3	Hazard Detection	Roboflow: Adult/Child Dataset	Images of adults and children for classification and detection	https://universe.roboflow.com/ds/e8it4C3ZsC?key=mvcfFLNtpp	1.1GB	Contains labeled images of adults and children
		Roboflow: Cone Dataset	Images of traffic cones for detection and classification Adult/Child Dataset	https://universe.roboflow.com/ds/qt7KGSdT8Y?key=yfmjOfikIX	500MB	Includes annotated images of traffic cones for training object detection models.
		Roboflow: pothole Dataset	Images of potholes for detection in road conditions	https://universe.roboflow.com/ds/22aieXdXGj?key=dDqWWspG80	500MB	Features labeled images of potholes to support road condition assessment and detection tasks.
		Roboflow: cat, dog Dataset	Images of cats and dogs for object classification	https://universe.roboflow.com/ds/pzbRAMnUNg?key=daA3RqdWVg	1.1GB	Includes images of cats and dogs with labels, supporting non-relevant object filtering.
4	FER	FER2013 Dataset	Facial emotion recognition dataset.	Kaggle: https://www.kaggle.com/datasets/msambare/fer2013 (FER2013)		Grayscale 48x48 images, categorized into emotions.
	Harmful Status	Gunshot Audio Dataset	Audio dataset for detecting gunshots.	Kaggle: https://www.kaggle.com/datasets		Various gunshot

				/emrahaydemr/gunshot-audio-dataset (Gunshots)		audio samples.
		Crime Audio Events	Audio of accidents and crimes.	https://www.kaggle.com/datasets/afisarsy/enhanced-audio-of-accident-and-crime-detection (Crime audio events)		Enhanced audio for crime detection.
		Speech Emotion Recognition	Neutral human speech data for emotion recognition.	https://www.kaggle.com/datasets/dmitrybabko/speech-emotion-recognition-en (Neutral human speech data) Downloadable sound effects: https://freesound.org/ https://sound-effects.bbcrewind.co.uk/search		English speech with labeled emotions.
	Harmful Objects	Knives and Guns Detection	Object detection for knives and guns.	Roboflow: https://universe.roboflow.com/crime-detection/guns_n_knives-h4bky (Knives and guns)		Includes labeled images of weapons.

2.3. Suitability Analysis

2.3.1. Relevance to Individual Research Objectives:

	1	2	3	4
Roboflow and CARLA Simulator	X			
Steer with front cam footage (Generated in simulation environment using a custom data generator.)	X			
Carla Simulation generated data.		X		
Adult/Child Dataset - Roboflow			X	
Cone Dataset - Roboflow			X	
Pothole Dataset - Roboflow			X	
Cat/Dog Dataset - Roboflow			X	
FER2013 Dataset				X
Gunshot Audio Dataset				X
Crime Audio Events				X
Speech Emotion Recognition				X
Knives and Guns Detection				X

Objective	Dataset	Alignment Explanation
F1: Object Detection	Carla Object Detection (Roboflow)	Highly aligned. Provides annotated data suitable for training object detection models in a simulation environment.
	Steer with front cam footage (Custom)	Perfectly aligned. Directly supports the steering function by providing data generated specifically for simulation.
F2: Simulator Generated Data	Custom environment creation method	Fully aligned. The dataset is tailored to train reinforcement learning models for real time dynamic environments.

F3: Hazard Detection	Hazard Detection (Roboflow)	Well-aligned. Includes various hazards relevant for improving safety detection capabilities in simulations.
F4: FER	FER2013 Dataset (Kaggle)	Highly aligned. FER2013 provides labeled emotion data necessary for facial emotion recognition models.
F4: Harmful Status	Harmful Status (Audio datasets)	Well-aligned. Gunshot and crime audio datasets provide scenarios critical for harmful event detection tasks.
	Speech Emotion Recognition (Kaggle)	Aligned. Supports emotion recognition and distinguishes neutral speech from harmful scenarios.
F4: Harmful Objects	Harmful Objects (Knives and Guns Dataset)	Well-aligned. Provides labeled data for detecting potentially harmful objects, enhancing the safety detection task.

3. Methodology

3.1. Data Preprocessing:

Ex:

Data Cleaning, Data Normalization, Data Standardization, Data Encoding (e.g., One-Hot Encoding, Label Encoding), Handling Missing Data (e.g., Imputation or Removal), Data Aggregation, Feature Engineering, Outlier Detection and Handling, Data Scaling, Data Discretization, Dimensionality Reduction (e.g., PCA), Date/Time Transformation, Data Integration (Merging or Joining), Data Mapping, Data Type Conversion.

Objective 01 - Perception and Scene Understanding

	Transformation technique				
Data Source	Data Cleaning	Data Encoding (e.g., One-Hot Encoding, Label Encoding)	Handling Missing Data (e.g., Imputation or Removal)	Feature Engineering	Data Mapping
Roboflow and CARLA Simulator	X	X	X	X	X
CARLA Simulator (Semantic Camera)	X		X	X	

Objective 02 – Decision Making and Collision Avoidance

	Transformation technique	
	Data Cleaning	Dimensionality Reduction
Simulator Generated Data	Cropped unwanted parts of images and resized images.	Reduced the feature space for computational efficiency.

Objective 03 - Ethical Decision Making in Autonomous Vehicles.

	Transformation technique							
	Data Aggregation	Data Cleaning	Data Mapping and Reannotation	Outlier Detection and Handling	Data Scaling (Image Resize)	Manual Annotation Checks	Label Standardization	Class Balance Adjustment
Adult/Child Dataset - Roboflow	X	X	X	X		X	X	
Cone Dataset - Roboflow	X	X		X		X	X	
Pothole Dataset - Roboflow	X	X		X	X	X	X	X
Cat/Dog Dataset - Roboflow	X	X		X	X			

Objective 04 - Enhancing In-Cabin Security and time Adaptive algorithms for Comfort.

	Transformation technique							
Data Source	Data Cleaning	Data Augmentation	Normalization	Noise Reduction	Data Integration	Data Standardization		
FER2013 Dataset	X	X						
Gunshot Audio Dataset	X		X					
Crime Audio Events	X	X		X	X			

Speech Emotion Recogniti on	X		X					
Knives and Guns Detection	X	X				X		

3.2. Scalability

Objective 01-

Roboflow and CARLA Simulator –

The Object Detection Dataset contains ~6,000 YOLOv5-annotated images for vehicles, pedestrians, lane markings, and traffic signs. While sufficient for initial testing, its size limits generalizability for large-scale, real-world applications. The dataset's diversity from simulated environments supports controlled experiments but struggles with real-world edge cases. Scalability can be achieved through additional CARLA simulations under varying conditions (e.g., weather, road layouts) and augmentation techniques like rotation and flipping to enhance robustness without extra data collection.

CARLA Simulator (Semantic Camera) -

The Semantic Data Collection Dataset contains ~50,000 cropped images labeled with general direction angles (-1, 0, 1) and predicted steering angles, making it ideal for steering angle prediction tasks. Its size and detailed annotations ensure robustness and variability for training models capable of handling complex driving scenarios. However, its reliance on CARLA simulations limits real-world applicability due to potential unrealistic scenarios. Scalability can be improved by integrating real-world driving data with simulated data. Optimizing storage and processing pipelines is essential to manage the computational demands of the dataset's size.

Objective 02 –

The dataset is sufficient for training RL model's lane understanding under diverse conditions.

Objective 03 -

Adult/Child Dataset: At 1.1 GB with 22,000 images, it is sufficient for preliminary detection tasks but lacks representation for rare scenarios. Scalability can be improved by augmenting underrepresented cases or integrating diverse datasets.

Cone Dataset: The 500 MB dataset is adequate for basic detection and classification but limited in handling varied conditions or multi-class scenarios. Adding diverse imagery from different environments would enhance scalability.

Pothole Dataset: At 500 MB, it provides general pothole data but lacks diversity in road conditions and environments, potentially limiting performance. Combining it with datasets from varied climates and road types would improve scalability.

Cat/Dog Dataset: With 1.1 GB, it is suitable for filtering tasks but limited to niche applications. Expanding it to include more categories of animals or objects would increase its scalability and versatility.

Objective 04-

FER2013 Dataset: The FER2013 dataset is a publicly available facial expression recognition dataset containing grayscale images of faces categorized into seven emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral.

Gunshot Audio Dataset: This dataset contains audio clips of gunshots, sourced from both real-world recordings and simulations. It is used to train models for detecting gunfire in surveillance or crime detection systems.

Crime Audio Events Dataset: A dataset comprising audio samples of crime-related sounds, such as breaking glass, alarms, screams, and gunshots. Used for identifying specific events in surveillance system

Speech Emotion Recognition Dataset: A dataset containing speech samples labeled with emotional states (e.g., happy, sad, angry, neutral). Commonly used for emotion recognition tasks in human-computer interaction systems.

Glass Breaking Sound Dataset: A specialized audio dataset focused on the sound of glass breaking, often collected from sound effects libraries like Freesound or BBC Sound Effects. Used in intrusion detection and surveillance systems.

Knives and Guns Detection Dataset: A collection of labeled images containing knives, guns, and other potentially harmful objects. Used for training object detection models in security and surveillance systems.

3.3 Feature extraction

Objective 01:
For object detection and steering angle prediction, explicit manual feature extraction was unnecessary due to the inherent capabilities of modern models like YOLO and CNNs:
Object Detection:
- Models automatically extract spatial features (e.g., edges, textures) through convolutional layers.
- Multi-scale feature extraction ensures accurate detection of objects of varying sizes.
- Bounding box regression and anchor box optimization handle localization and classification.

Steering Angle Prediction:
- Semantic segmentation features (e.g., lane markings, road edges) were extracted for road understanding.
- Models captured spatial relationships and directional patterns for smooth navigation.
- Temporal correlations between frames enhanced steering accuracy.

Objective 02:
CNN model works as a future extractor for RL model by extracting relevant features from raw image.

Objective 03:
Feature extraction relied on automated processes inherent in object detection models like YOLO, supported by preprocessing steps:
Preprocessing: Images were resized, normalized, and augmented (e.g., flipping, rotation) to enhance feature diversity and consistency.
Automated Feature Extraction: YOLO’s architecture handled feature extraction, identifying object patterns, spatial positions, and multi-scale features.
Handling Challenges: Unified class labels, removed outliers, and augmented data for rare scenarios to ensure robust feature representation.
This streamlined approach effectively captured relevant features for hazard detection and classification in autonomous vehicles.

Objective 04:
Audio Processing part have custom feature extraction method.

4. Modelling and Results

Objective 01:

Training and Testing

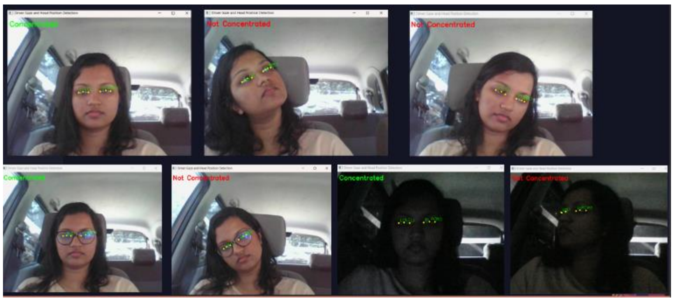

```
tensor([[137.12079, 195.06799, 290.26932, 298.42981,   0.94626,   9.00000],
        [455.11258, 171.43365, 467.20508, 214.65485,   0.90649,   5.00000],
        [246.85228, 237.17397, 543.95447, 446.50391,   0.85170,   9.00000]])
```

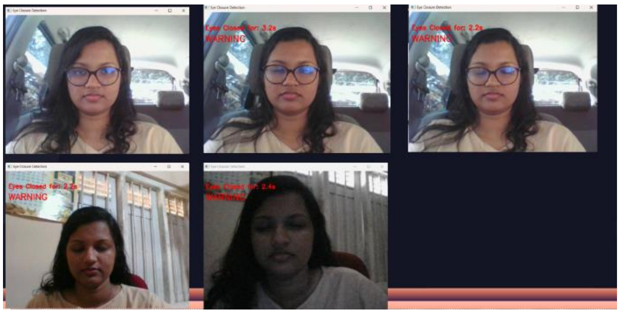
BSc (Hons) in Information Technology Specializing Data Science

Research Project - IT4010 **Data Analysis Report**

Objective 03:

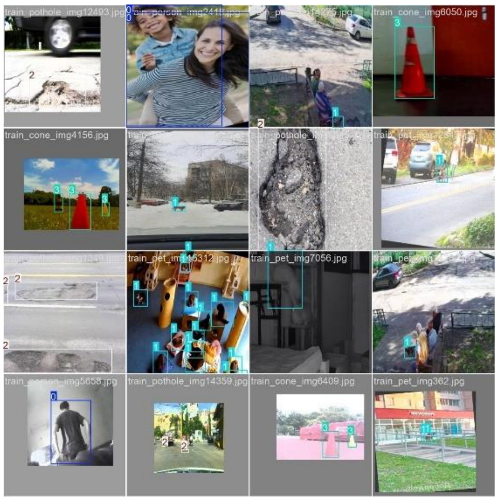
Eyeball Tracking System for Attention Monitoring





Hazard detection

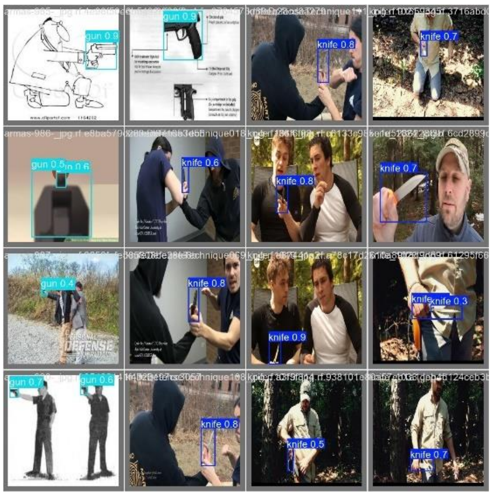
```
Model summary (fused): 168 layers, 11,127,132 parameters, 0 gradients, 28.4 GFLOPs
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100%|██████████| 147/147 [01:09<00:00,  2.10it/s]
                   all       9366      25193      0.832      0.733      0.784      0.533
                Person       2747      10458       0.87       0.86      0.916      0.586
                   Pet       1943       2110      0.671      0.596      0.637      0.339
               Pothole       3438      10940      0.794      0.481      0.589        0.3
                  Cone       1119       1685      0.994      0.995      0.995      0.907
Speed: 0.1ms preprocess, 0.9ms inference, 0.0ms loss, 1.4ms postprocess per image
```



Objective 04:

Weapon and harmful objects detection

```
Validating runs\detect\train4\weights\best.pt...
Ultralytics 8.3.21  Python-3.8.20 torch-2.4.1 CUDA:0 (NVIDIA GeForce RTX 3070 Laptop GPU, 8192MiB)
YOLOv5s summary (fused): 193 layers, 9,112,310 parameters, 0 gradients, 23.8 GFLOPs
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100%|██████████| 16/16 [00:08<00:00,  1.86it/s]
                   all       1011       1132      0.917      0.867      0.926      0.652
                 knife        424        439      0.934      0.913      0.949      0.625
                   gun        587        693        0.9      0.821      0.903       0.68
Speed: 0.1ms preprocess, 0.9ms inference, 0.0ms loss, 1.8ms postprocess per image
Results saved to runs\detect\train4
```



4.1. Key Insights:

Objective 01:

1. Object Detection:

The dataset is dominated by vehicles and pedestrians, whereas smaller objects like signs are not as well-represented.
The bounding box annotation required manual review to confirm the accuracy.

2. Semantic Data Collection:

Cropping images reduces irrelevant information, focusing on critical features like lanes. Random yaw adjustments improve model generalization to varied road conditions.

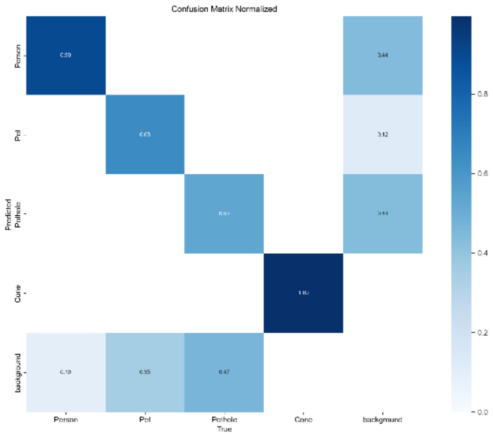
SLIIT
Discover Your Future

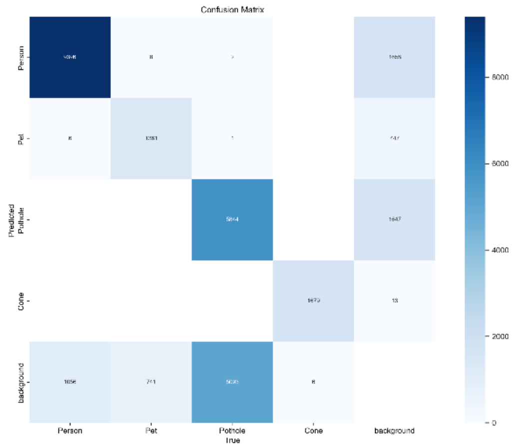
BSc (Hons) in Information Technology Specializing Data Science

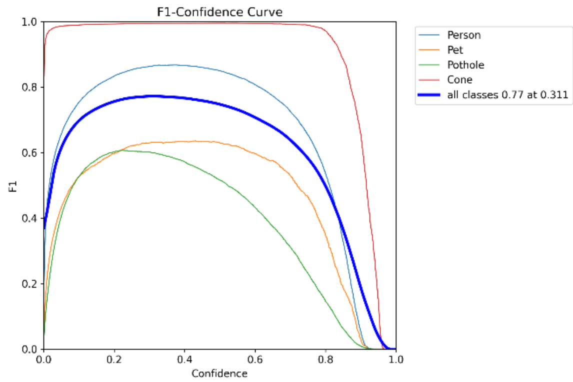
Research Project - IT4010 **Data Analysis Report**

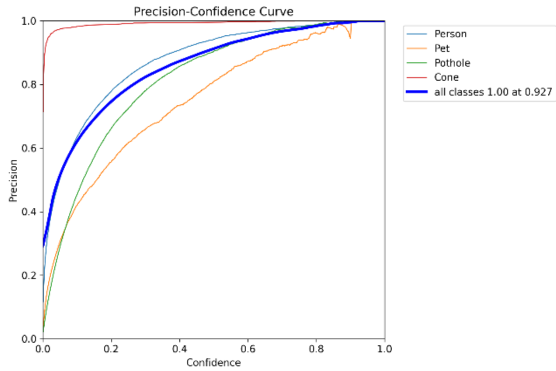
Objective 03:

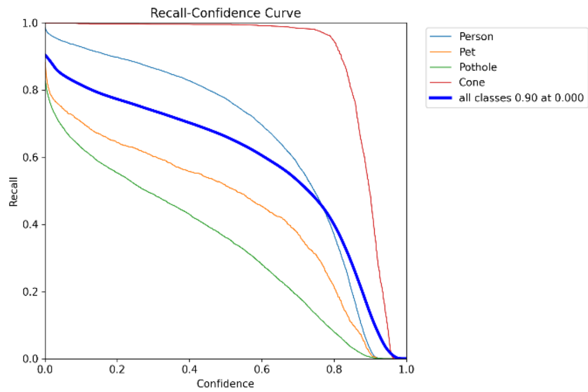
Hazard Detection

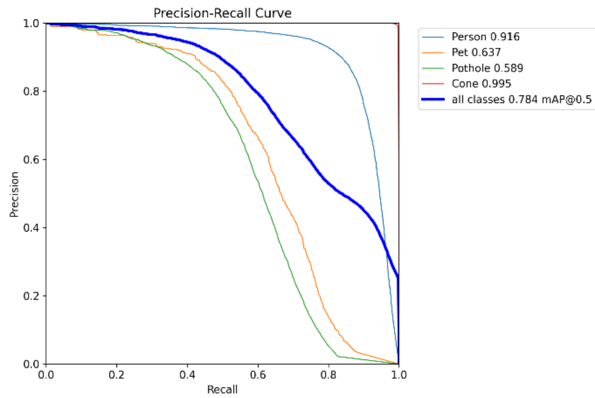


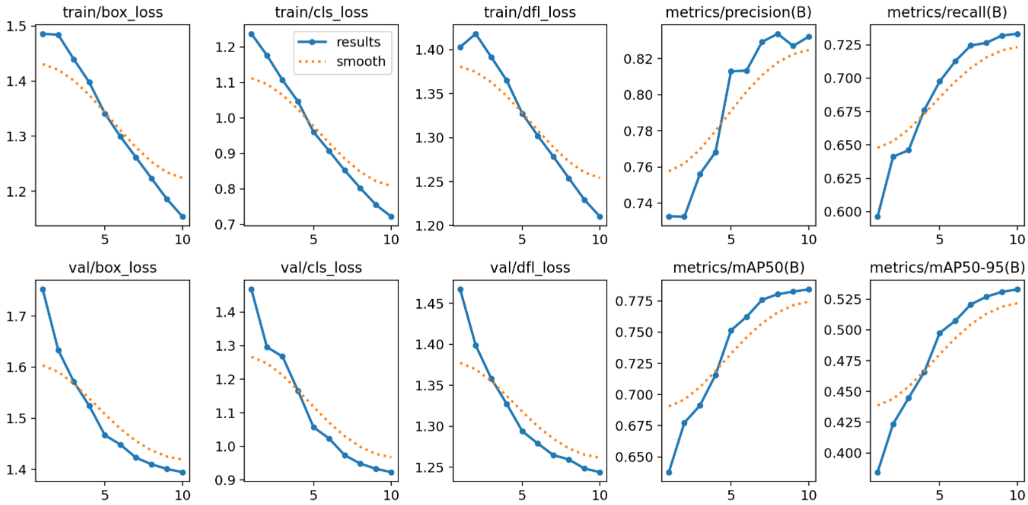












The overall dataset analysis highlighted several key insights:

- Class Imbalance: Some datasets showed an uneven distribution of classes, which could impact model performance. Data augmentation is necessary to balance the dataset.
- Dataset Diversity: Variations in environmental conditions (lighting, weather, clutter) affected accuracy, indicating a need for more diverse data to improve model robustness.
- Feature Representation: Features like size, shape, and texture were well captured, but rare or occluded objects were harder to detect, suggesting the need for more varied examples.
- Labeling Inconsistencies: Labeling issues were identified, requiring manual correction to ensure consistency across datasets.

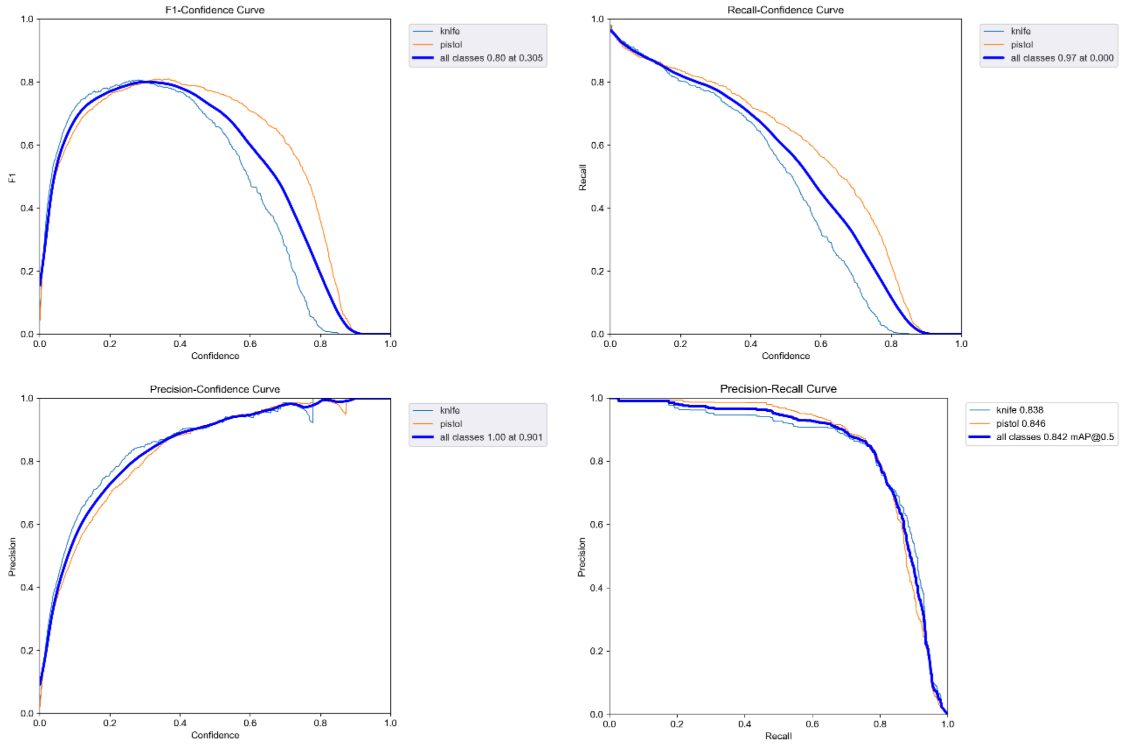
Eyeball Tracking System for Attention Monitoring

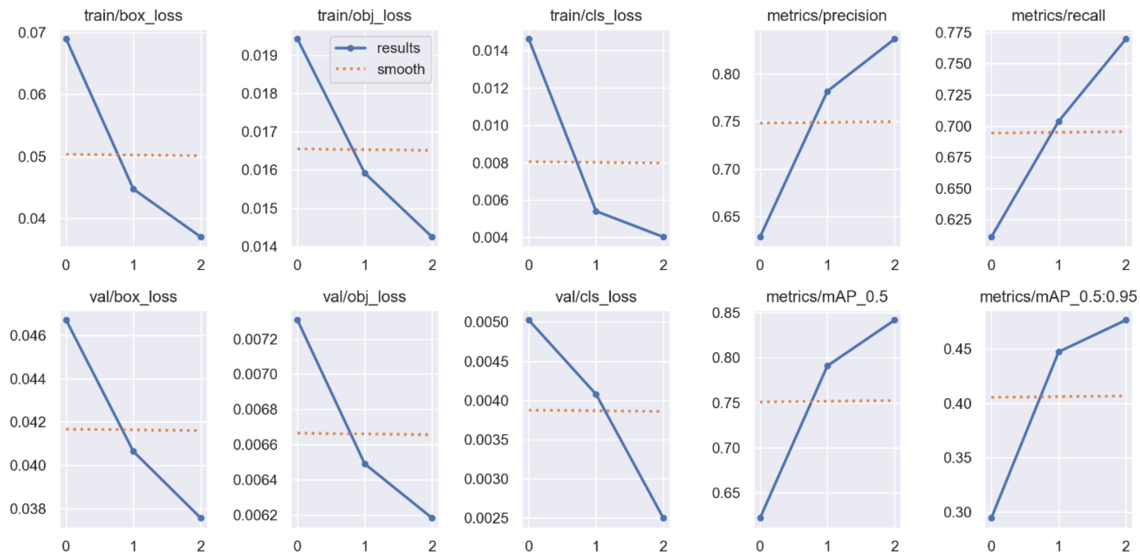
```
# Initialize MediaPipe Face Mesh
mp_face_mesh = mp.solutions.face_mesh
face_mesh = mp_face_mesh.FaceMesh(
    max_num_faces=1,
    refine_landmarks=True,
    static_image_mode=False,
    min_detection_confidence=0.5,
    min_tracking_confidence=0.5
)

# Indices for key eye landmarks
left_eye_top = 159
left_eye_bottom = 145
left_eye_left = 133
left_eye_right = 33

right_eye_top = 386
right_eye_bottom = 374
right_eye_left = 362
right_eye_right = 263
```

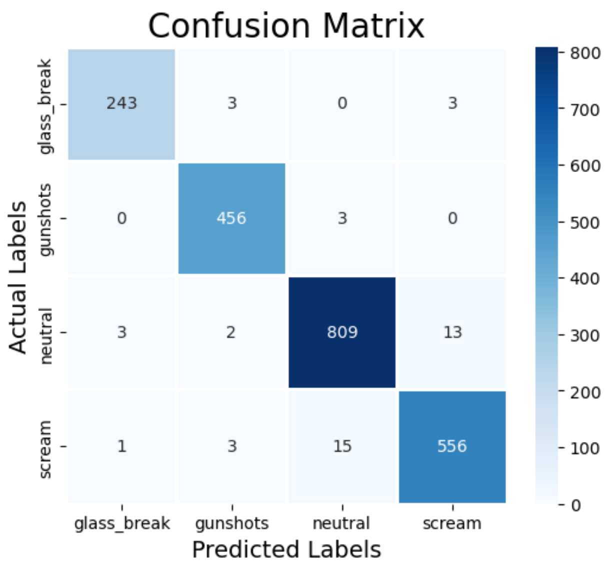
```
# Initialize MediaPipe Face Mesh
mp_face_mesh = mp.solutions.face_mesh
face_mesh = mp_face_mesh.FaceMesh(
    max_num_faces=1,
    refine_landmarks=True,
    static_image_mode=False,
    min_detection_confidence=0.5,
    min_tracking_confidence=0.5
)

# Indices for key landmarks on the iris and eye corners
left_iris_index = 468
left_eye_inner = 133
left_eye_outer = 33
left_eye_top = 159
left_eye_bottom = 145

right_iris_index = 473
right_eye_inner = 362
right_eye_outer = 263
right_eye_top = 386
right_eye_bottom = 374
```

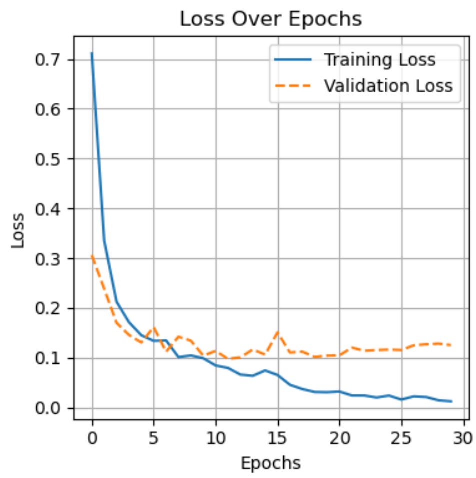
Objective 04:

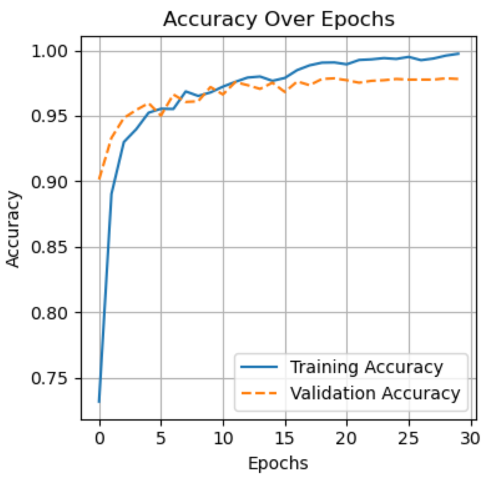
Weapon and harmful objects detection






Confusion Matrix
Actual Labels
Predicted Labels
glass_break
gunshots
neutral
scream
243
3
0
3
0
456
3
0
3
2
809
13
1
3
15
556
800
700
600
500
400
300
200
100
0

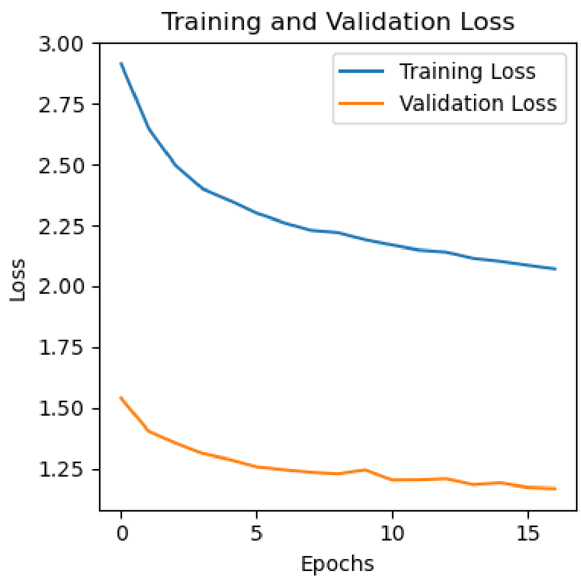

Loss Over Epochs
Training Loss
Validation Loss
Loss
Epochs

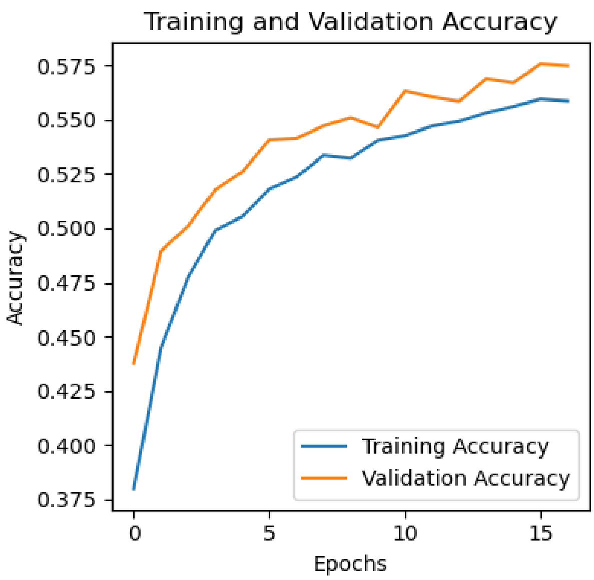

Accuracy Over Epochs
Training Accuracy
Validation Accuracy
Accuracy
Epochs

Facial emotion detection

```
Epoch 18: val_accuracy improved from -inf to 0.63876, saving model to mobilenet_face_ft.h5
444/444 [==============================] - 79s 172ms/step - loss: 1.8238 - accuracy: 0.6208 - val_loss: 0.9844 - val_accuracy: 0.6388
Epoch 19/27
444/444 [==============================] - ETA: 0s - loss: 1.5385 - accuracy: 0.6685
Epoch 19: val_accuracy improved from 0.63876 to 0.65783, saving model to mobilenet_face_ft.h5
444/444 [==============================] - 76s 170ms/step - loss: 1.5385 - accuracy: 0.6685 - val_loss: 0.9507 - val_accuracy: 0.6578
Epoch 20/27
444/444 [==============================] - ETA: 0s - loss: 1.3974 - accuracy: 0.6931
Epoch 20: val_accuracy improved from 0.65783 to 0.66829, saving model to mobilenet_face_ft.h5
444/444 [==============================] - 75s 170ms/step - loss: 1.3974 - accuracy: 0.6931 - val_loss: 0.9238 - val_accuracy: 0.6683
Epoch 21/27
444/444 [==============================] - ETA: 0s - loss: 1.2915 - accuracy: 0.7115
Epoch 21: val_accuracy improved from 0.66829 to 0.68750, saving model to mobilenet_face_ft.h5
444/444 [==============================] - 76s 170ms/step - loss: 1.2915 - accuracy: 0.7115 - val_loss: 0.8827 - val_accuracy: 0.6875
Epoch 22/27
444/444 [==============================] - ETA: 0s - loss: 1.1872 - accuracy: 0.7329
Epoch 22: val_accuracy improved from 0.68750 to 0.69108, saving model to mobilenet_face_ft.h5
444/444 [==============================] - 76s 171ms/step - loss: 1.1872 - accuracy: 0.7329 - val_loss: 0.8975 - val_accuracy: 0.6911
Epoch 23/27
444/444 [==============================] - ETA: 0s - loss: 1.0890 - accuracy: 0.7551
Epoch 23: val_accuracy improved from 0.69108 to 0.69538, saving model to mobilenet_face_ft.h5
444/444 [==============================] - 77s 173ms/step - loss: 1.0890 - accuracy: 0.7551 - val_loss: 0.8968 - val_accuracy: 0.6954
Epoch 24/27
...
Epoch 27/27
444/444 [==============================] - ETA: 0s - loss: 0.7487 - accuracy: 0.8323
Epoch 27: val_accuracy improved from 0.69796 to 0.70212, saving model to mobilenet_face_ft.h5
444/444 [==============================] - 77s 174ms/step - loss: 0.7487 - accuracy: 0.8323 - val_loss: 0.9770 - val_accuracy: 0.7021
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...





4.2. Challenges Faced During Data Analysis:

Objective 01: Perception and Scene Understanding

CNN

- *Annotation Quality:*

Had to add random yaw (with maximum 15 degrees) to generalize images with various steer angle. Inconsistent or missing annotations in some images required manual review and corrections. (Sometimes plain black images ware saved due to Carla API inefficient)

- *Data Imbalance:*

Some object classes were underrepresented, leading to difficulty in detecting these objects accurately. (e.g. Images for 0 steer angle were over presented)

Required custom weight values to balance the dataset.

- *Simulation Constraints:*

The simulation environment sometimes generated unrealistic scenes, reducing model generalizability to real world data.

- *Computational Limits:*

Generating a large dataset was computationally intensive and required optimization of environment creation and reset methods.

Data generation was slowed even with moderate gpu power due to Carla API slow interaction time with agents. (3000 images ware generated per hour and final model was trained using around 30k samples)

Objective 2 – Decision Making and Collision Avoidance

- *Dynamic Environment Complexity:*

Designing environments with realistic and challenging scenarios is time consuming. Often, the environments needed fine tuning for proper reward distribution for the RL algorithm.

- *Simulation Stability (Computational Limit):*

Due to computational limitations, real time simulation would sometimes crash or lag, disrupting training sessions and requiring additional debugging.

- *Exploration vs. Exploitation:*

The model struggled to balance exploring new strategies and learning behaviors in complex environments and scenarios. Once the model was trained to navigate in a straight lane, it was motivated to navigate through curved lanes. But enhancing the model to navigate through intersection was a challenge. The reason was that the model seems to be controlled by the steer mainly while looking at lane lines which will end at intersections. Therefore, waypoint were used in those areas instead of relying on lane lines.

Objective 03: Ethical Decision Making in Autonomous Vehicles

Hazard Detection

- *Combining Images from Different Datasets:*

Differences in labeling conventions required mapping and reannotation to standardize the classes. (there is a custom method to label the dataset correctly)

Variations in image quality, resolution, and lighting conditions caused inconsistencies in training data.

Some datasets had overlapping but not identical definitions for certain hazards, leading to ambiguity in class mapping. (manually checked random sample before aggregate)

- *Dataset Size:*

Limited number of hazard for specific images, especially for rare scenarios, reduced the model's ability to generalize.

- *Outliers:*

Some images contained irrelevant or mislabeled content, which required manual filtering. (Some labeling errors ware there even when we are training the model)

- *Scaling Issues:*

Handling high resolution images demanded significant computational resources, leading to memory bottlenecks. (imsize was dropped to 320 from 640 for final model)

Objective 04: Ethical Decision Making in Autonomous Vehicles

FER (Facial Emotion Recognition)

- *FER2013 Challenges:*

The FER2013 dataset is one of the most widely used for facial emotion recognition tasks and is known for its highly generalized and diverse images across classes.

This inherent variability posed a challenge in maintaining its generalizability while preprocessing it for optimizing model performance.

As noted on *paperswithcode.com*, the current best-performing models achieve only 79% accuracy on this dataset, despite reaching over 98% accuracy on less generalized FER datasets. This highlights the complexity of the dataset.

- *Class Imbalance:*

Certain emotions (e.g., 'Happy') dominated the dataset, impacting model accuracy for other classes. (Used class weights to address this)

Augmentation techniques like flipping, rotation, and noise addition were tested to enhance model performance but risked distorting the datasets variability, leading to reduced real world generalizability.

Introducing custom weights to emphasize specific classes during training worked better than traditional augmentation methods. This approach preserved the datasets diverse characteristics while addressing class imbalances.

- *Low Resolution:*

The images (48x48 pixels) were challenging for models to extract detailed facial features. (Resized to 224*224 for better accuracy)

Harmful Status Detection

- *Data Quality:*

Audio clips had varying lengths, necessitating normalization for consistent input. Custom preprocess method was written to address that and audio editing/Wave plot visualization software (ocenaudio) was used to analyze the datasets

- *Limited Real World Data:*

Synthetic or enhanced audio data lacked diversity, reducing real world applicability. (Finding datasets for glass breaking sound was a challenge)

To obtain an audio dataset for the sound of glass breaking, I used existing datasets and downloadable sound effects available online. (https://freesound.org/, https://sound-effects.bbcrewind.co.uk/search)

- *Noise and Variability:*

Background noise in gunshot and crime audio datasets affected feature extraction.

Generalization.

Added human speech voice dataset to generalize the model to identify human voice as a neutral.

Added vehicle engine sounds to neutral class for better performance. (Crime Audio Events dataset had audio samples for this label)

Harmful Objects Detection:

- *Class Ambiguities:*

Some objects, such as tools and weapons, had ambiguous categorization across datasets, requiring careful handling to avoid misclassification.

Image Quality Variability: Differences in resolution, lighting, and image angles across datasets led to inconsistencies in visual features, making it difficult for the model to generalize across classes.

5. References

[1] Z. Wu, "Path planning techniques for autonomous vehicles," *AIP Conference Proceedings*, Jan. 2024. [Online]. Available: https://doi.org/10.1063/5.0215501.

[2] W. Schwarting, J. Alonso-Mora, and D. Rus, "Planning and decision-making for autonomous vehicles," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 187–210, 2018. [Online]. Available: https://doi.org/10.1146/annurev-control-060117-105157.

[3] Q. Yuan *et al.*, "Decision-making and planning methods for autonomous vehicles based on multistate estimations and game theory," *Advanced Intelligent Systems*, vol. 5, no. 11, Sep. 2023. [Online]. Available: https://doi.org/10.1002/aisy.202300177.

[4] R. Nelson de Moura, R. Chatila, K. Evans, S. Chauvier, and E. Dogan, "Ethical decision making for autonomous vehicles," in *IEEE Symposium on Intelligent Vehicles*, Oct. 2020. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03022605.

[5] Z. Wei and L. Jun, "Overview of research on obstacle avoidance path planning for autonomous vehicles," *Automotive Engineer*, pp. 55–58, 2018.

[6] A. Mishra, S. Lee, D. Kim, and S. Kim, "In-cabin monitoring system for autonomous vehicles," *Sensors*, vol. 22, no. 12, p. 4360, Jun. 2022. [Online]. Available: https://doi.org/10.3390/s22124360.

Datasets and Resources:

[7] Roboflow, "Path planning dataset for autonomous vehicles." [Online]. Available: https://universe.roboflow.com/ds/l8x1Cz24Fr?key=nWDBG8LZjq.

[8] Roboflow, "Traffic dataset for machine learning." [Online]. Available: https://universe.roboflow.com/ds/e8it4C3ZsC?key=mvcfFLNtpp.

[9] Roboflow, "Pedestrian detection dataset." [Online]. Available: https://universe.roboflow.com/ds/qt7KGSdT8Y?key=yfmjOfikIX.

[10] Roboflow, "Collision detection dataset." [Online]. Available: https://universe.roboflow.com/ds/22aieXdXGj?key=dDqWWspG80.

[11] Roboflow, "Vehicle dynamics dataset." [Online]. Available: https://universe.roboflow.com/ds/pzbRAMnUNg?key=daA3RqdWVg.

[12] Kaggle, "FER2013 facial expression recognition dataset." [Online]. Available: https://www.kaggle.com/datasets/msambare/fer2013.

[13] Kaggle, "Gunshot audio dataset." [Online]. Available: https://www.kaggle.com/datasets/emrahaydemr/gunshot-audio-dataset.

[14] Kaggle, "Crime audio event detection dataset." [Online]. Available: https://www.kaggle.com/datasets/afisarsy/enhanced-audio-of-accident-and-crime-detection.

[15] Kaggle, "Speech emotion recognition dataset." [Online]. Available: https://www.kaggle.com/datasets/dmitrybabko/speech-emotion-recognition-en.

[16] Freesound, "Sound effects library." [Online]. Available: https://freesound.org/.

[17] BBC Rewind, "Sound effects collection." [Online]. Available: https://soundeffects.bbcrewind.co.uk/search.

[18] Roboflow, "Guns and knives detection dataset." [Online]. Available: https://universe.roboflow.com/crime-detection/guns_n_knives-h4bky.