Maximising Highway Safety through AI-enabled Detection of Pedestrians and Animals in V2X Environments

Rittick Maity[1,2], Jivthesh M R[1,3], Pralay Sankar Maitra[1], Sanjeevi G[1], Gaushik M R[1,3],
Sai Shibu N B[1], and KA Unnikrishna Menon[1]

[1]Center of Wireless Networks & Applications (WNA), Amrita Vishwa Vidyapeetham, Amritapuri, India
[2]Getinge India Innovation Centre, Bengaluru, India
[3]Tata Elxsi, Trivandrum, India
rittick.maity.58@gmail.com, saishibunb@am.amrita.edu, kaumenon@am.amrita.edu

***Abstract*—Road accidents involving pedestrians and animals can have serious consequences for all parties involved. These types of accidents can occur when pedestrians or animals cross the road and can often be caused by poor visibility, driver distractions, or a lack of appropriate safety measures. In many cases, these accidents can be prevented by implementing measures such as pedestrian crossings and wildlife crossings, as well as by educating drivers about the importance of paying attention to their surroundings and being alert to pedestrians and animals on or near the road. We can work towards a safer and more efficient transportation system for all road users by taking steps to prevent these types of accidents. This paper proposes an AI-enabled pedestrian and animal detection architecture for V2X-enabled highways. The proposed system leverages AI to detect pedestrians and animals on the road and then transmit this information to other road users using the existing V2X Communication protocol available on the highways. The paper also explains the algorithm to detect animals and pedestrians using AI. We envision that the proposed system will reduce road accidents and save lives.**

***Index Terms*—Animal Crossing Detection, Pedestrian Detection, V2X, Machine Learning**

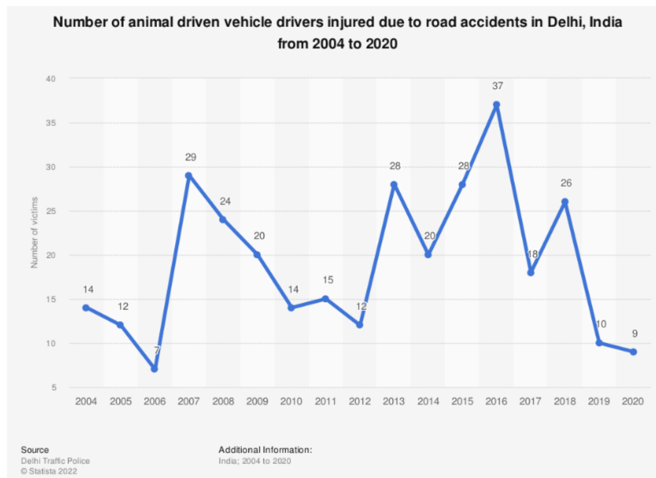

Fig. 1. Number of road accidents in Delhi, India due to animal crossing during the period 2004 to 2020

I. Introduction

Animal and pedestrian crossings can be a significant issue on roads, especially highways where people drive at high speeds. Road accidents [1] due to animals or pedestrians crossing can be dangerous for animals, drivers and passengers. These accidents also cause delays and traffic disruptions [2]. Drivers may be unaware of the presence of animals or pedestrians on the road and may need more time to take evasive action or to slow down to avoid a collision [3]. Figure 1 shows the statistics of road accidents due to animal crossing in Delhi, India, from 2004 to 2020 [4].

The existing systems have several limitations. These systems rely on expensive sensors or cameras that may need to be more practical or cost-effective for widespread deployment. Some systems have difficulty distinguishing between different types of animals or pedestrians or may produce false positives or false negatives, which can lead to confusion or false alarms. We propose using artificial intelligence to detect animals and pedestrians crossing on the roads [5]. Existing roadside and vehicle-mounted cameras capture images, and the proposed system processes these images to identify animals or pedestrians crossing the road [6] and [7]. The system is also V2X enabled. As soon as the system detects animals or pedestrians, it uses the existing V2X communication network to broadcast the message to all vehicles on the road, thereby alerting the drivers. This paper discusses the system architecture and artificial intelligence algorithms to detect animal, and pedestrian crossings [8] and [9].

The paper is organised as follows. Section II discusses state-of-the-art and presents the research gap and key challenges. Section III explains the proposed system architecture for an AI-enabled system to detect animals and pedestrians on the highways. Section IV describes the machine learning models suitable for the proposed system. Section V explains the design of the identified machine learning algorithm to detect pedestrians and animals from the dataset. Section VI describes the implementation of the developed algorithm and discusses the results. The paper is concluded in section VII.

II. Related Work

This section summarises research works that are similar to the proposed system.

Reagan L. Galvez et al. have used Convolutional Neural Networks (CNN) to identify things in the surrounding environment [10]. Both a Single Shot Multi-Box Detector (SSD) with MobileNetV1 and a Faster Region-based Convolutional Neural Network (Faster-RCNN) with InceptionV2 are considered to be state-of-the-art models for object identification. These models were evaluated and contrasted by the authors. According to their findings, one of the models is ideally suited for real-time applications due to its speed. In contrast, the other model can be utilised for more accurate object detection.

Xuefeng Liu et al. used marine animal photos from an underwater robot with an implanted gadget [11]. The authors then created a MobileNetV2 model using a CNN and marine animal photos. This model is real-time. Authors improved classification using transfer learning. After that, they trained the model using the marine animal photographs they acquired to transmit the trained model to the embedded device and categorise underwater marine animal photos in real time. They also assessed the identification accuracy rate and average classification time using the InceptionV3 and MobilenetV1 models from the experiments to evaluate the suggested strategy. Their research suggests that the MobilenetV2 model with transfer learning may be better at real-time marine animal photo classification than the other two models.

Alexander Gomez Villa et al. demonstrated that once certain challenges are solved, deep neural networks can correctly cope with the challenge of automated species classification [12]. The most prevalent 26 out of 48 species from the Snapshot Serengeti dataset were chosen as a case study, and the potential of the Very Deep Convolutional neural networks framework for the species identification job was investigated. This was done to understand better how the framework may be used. In the worst case, the suggested approach achieved an accuracy of 35.4% in the Top-1 and 60.4% in the top 5. In the best-case scenario, the accuracy reached 88.9% in the Top-1 category and 98.1% in the Top-5 category, respectively.

Rakhsith et al. surveyed various object detection algorithms mentioned in 26 publications [13]. The authors reviewed the methods utilised and analysed the recommended approaches in these papers. The authors also highlighted the findings and observations in these papers. The authors conclude that the YOLO model is the efficient algorithm for object detection and classification as this model offers better accuracy than other existing models. Although the YOLO model is efficient, they consume more computation power, which is this model's major disadvantage.

M.Hemaanand et al. proposed a cutting-edge method based on deep learning and the Internet of Things [14]. The suggested method can monitor the driver as well as the surroundings of the vehicle, which results in a measurable improvement in the car's level of safety and security. The sensors are installed in the vehicle's steering to monitor the driver's heartbeat. These sensors can also detect and classify other vehicles, bicycles, and pedestrians. They can alert the user if there is any abnormality in the driver's heartbeat through deep learning techniques and internet technology.

A. *Research gaps identified from the literature review*

- Animals are more difficult to recognise than pedestrians because they come in various sizes, forms, positions, and colours.
- When compared to other animals, humans have a fundamental size and shape that is standard and average. This is not the case with other species.
- While current algorithms can accurately detect and classify pedestrians and animals in many situations, there are still cases where they may fail or produce false positives.
- Pedestrian and animal detection in complex environments, such as dense urban areas or forests, can be challenging. Algorithms should be the one that can accurately detect and classify objects in these environments.
- Pedestrian and animal detection algorithms must handle rare events, such as animals behaving in unexpected ways or pedestrians wearing unusual clothing. So there is a need for algorithms to be developed that can handle these edge cases.

B. *Key Challanges*

- One of the biggest challenges in developing is to obtain high-quality training data. This is especially challenging in pedestrian and animal detection, as it is difficult to capture enough examples of these objects in various lighting and weather conditions.
- It is important to minimise false positives (detecting an object where none exists) and false negatives (failing to detect an object that is present) in pedestrian and animal detection systems. This isn't easy due to these objects' wide variety of shapes, sizes, and appearances.
- Pedestrians and animals can exhibit a wide range of appearances and behaviour, making it difficult to detect and classify them accurately. For example, pedestrians may wear clothing or carry objects that obscure their shape, and animals may move unexpectedly.
- Pedestrians and animals may be partially or fully occluded by other objects, such as vehicles or vegetation, making it difficult for the detection system to detect and classify them accurately.

III. System Architecture for AI-enabled Animal and Pedestrian Detection on Highways

The proposed system consists of V2X entities, modules for communication, and cloud storage, as shown in figure 2. The proposed system is designed to improve road safety by detecting pedestrians and animals on the road and generating real-time warning or alert messages. It uses V2X technology to collect data from cameras and sensors and applies a machine-learning algorithm to accurately identify pedestrians and animals in the data. The messages generated by the system are then verified and validated by the message verification process and broadcast to other entities in the V2X network if they are valid. By providing timely and accurate information to other entities in the network, the system has the potential to significantly reduce the risk of accidents and collisions on the road. The overall functioning of the proposed system is explained in detail section the Methodology.

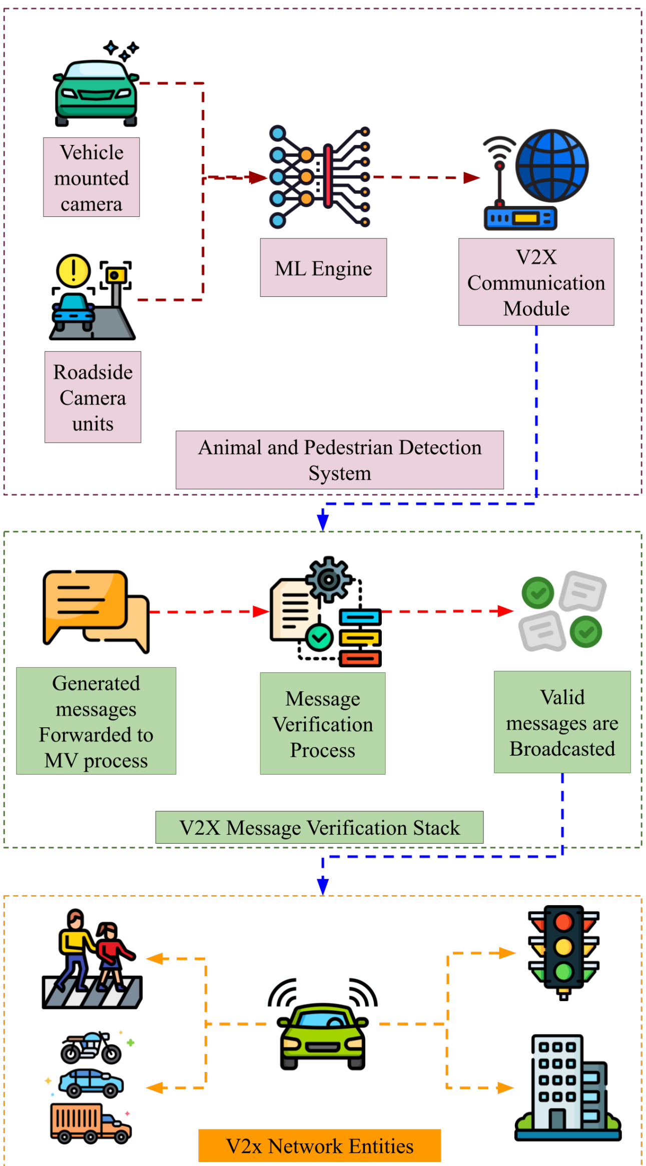

Fig. 2. Proposed System Architecture for AI-enabled pedestrian and animal detection system for V2X networks

The system architecture as shown in figure 2, helps road users to prevent accidents caused by animals & pedestrians crossing the road. The architecture consists of three main components:

1) A: Animal and pedestrian detection system: The proposed system uses a camera that captures images in a 360-degree field of view to detect animals and pedestrians on the road. The camera is mounted on a vehicle and functions similarly to a traffic camera, taking pictures or recording video as the vehicle travels. We assume that the camera mounted on a vehicle has sufficient coverage to capture the animals and pedestrians crossing the road. When the camera captures an image, the system applies a machine learning algorithm to analyse it and determine whether there are any animals or pedestrians. The system generates a warning or alert message if an animal or pedestrian is detected. This message may include information about the animal or pedestrian's location, the travel's direction, and other relevant details such as distance and movement. The warning or alert message is then transmitted to the next step using a V2X (Vehicle-to-Everything) communication module. The V2X communication module is responsible for sending and receiving V2X messages, which are used to exchange information between different entities in the system. The V2X communication module transmits the warning or alert message to other system components, such as vehicles or roadside units.

2) B: V2x message verification stack: In this process stage, the system examines the warning or alert messages generated in the previous stage when the camera and machine learning algorithm detect an animal or pedestrian on the road. These messages are passed through a message verification process to determine whether or not they are reliable. The message verification process ensures that only valid and trustworthy messages are broadcast to other entities in the V2X (Vehicle-to-Everything) network. It uses cryptographic techniques, such as digital signatures and message authentication codes, to verify the authenticity and integrity of the messages. If the message verification process determines that a message is valid and reliable, it is broadcast to other entities in the V2X network. These entities may include other vehicles, roadside units, traffic management systems, and other systems. The goal is to provide timely and accurate information to all the network entities so that they can take appropriate action to avoid accidents or other incidents. It is discarded if the message verification process determines whether it is invalid or unreliable. This helps to prevent false or misleading information from being broadcast to the network and potentially causing confusion or chaos.

3) C: V2X network entities: In this process step, the system broadcasts the warning or alert messages determined to be authentic and reliable to the various entities that are part of the V2X (Vehicle-to-Everything) network. The V2X network is a network of technologies that enables vehicles to exchange information with their surroundings, including other vehicles, the road infrastructure, and pedestrians. By broadcasting the warning or alert messages in real time, the system allows other entities in the network to receive timely and accurate information about potential hazards or incidents on the road. This can help to improve safety and reduce the risk of accidents. For example, if a vehicle receives a warning about an animal crossing the road ahead, it can take action to avoid hitting the animal. This might include reducing its speed, switching lanes, or taking other evasive action.

Similarly, if a pedestrian is detected crossing the road, the vehicle can slow down or take other measures to avoid hitting the pedestrian. Overall, this process step is designed to provide real-time information to the various entities in the V2X network to improve safety and reduce the risk of accidents. By broadcasting the warning or alert messages to the network, the system enables other entities to take appropriate action in response to potential hazards or incidents on the road.

A. Validation of proposed system

The proposed method for preventing accidents caused by animals crossing the road has the potential to improve road safety and reduce the frequency of accidents and collisions on roads. There are several reasons why this might be the case:

- Improved warning and alert systems: By using optical sensors, a machine learning algorithm, and a V2X (Vehicle-to-Everything) communication system, the proposed method can detect animals approaching the road and generate a warning or alert messages in real-time. These warnings can be transmitted to other vehicles or roadside units and can help to give drivers sufficient time to take evasive action or to slow down to avoid hitting the animal. This can significantly reduce the risk of accidents and collisions on the road.
- Increased efficiency: By reducing the risk of accidents and collisions, the proposed method can also help increase transportation systems' efficiency. For example, avoiding delays caused by accidents can help reduce fuel usage and pollution and save time and money for drivers and other road users.
- Improved safety: Ultimately, the primary goal of the proposed method is to improve safety on the road. By helping to prevent accidents caused by animals crossing the road, the method has the potential to significantly reduce the risk of injury or death for all road users.

B. Dataset

The proposed system for preventing accidents caused by animals crossing the road uses a machine-learning algorithm to analyze data and detect animals on the road. To train and evaluate the model, a data set of images containing various classes of wild animals are used, which is available online. The data set we have chosen contains 13 classes of wild animals, such as tigers, giraffes, bears, lions, elephants, deer, wolves, and so on. The images in the data set have a resolution of 1024x405px. Eighty per cent of the images are used to train the model, while the remaining 20 per cent are used to evaluate the model. We trained the model on a large portion of the data set and then tested it on a smaller portion to see how well it performed. By training and evaluating the model on a data set of images containing various classes of wild animals, the system can accurately detect animals on the road and generate warning or alert messages in real-time.

IV. Identification of suitable Machine Learning Models for Animal and Pedestrian Detection on Highways

A. Regional Convolutional Neural Network

Regional Convolutional Neural Network (R-CNN) detects objects in pictures and videos [15]. The R-CNN approach begins with region proposals or picture object locations. Step one. The method checks each point in the image with a defined-size window. This sliding window method creates these region ideas. The R-CNN technique uses a convolutional neural network (CNN) to classify area suggestions as items or backgrounds. A massive dataset of images and captions teaches CNN to recognise things in new shots. The R-CNN algorithm will then utilise regression to refine the bounding boxes around recognised items to fit the image better. R-CNN can accurately detect objects in photos, even if the objects are hidden or in a scenario with many moving elements. Picture categorisation, object identification, and object tracking are used extensively. The R-CNN approach for image and video recognition has revolutionised computer vision. R-CNN is effective in detecting objects in pictures and videos.

1) Faster RCNN: Faster R-CNN proposes regions and classifies them as objects or backgrounds using a CNN. However, it includes some substantial process-speeding advancements. Faster R-CNN initially uses a similar convolutional feature map for area proposal generation, and object categorisation [16]. This avoids repeating computation, improving efficiency. Second, the Faster R-CNN algorithm generates region suggestions using a Region Proposal Network (RPN) instead of a sliding window. RPN is a lightweight CNN trained to predict item presence at every photo point. Finally, the Faster R-CNN algorithm uses an anchor mechanism to maintain several object scales and aspect ratios, improving region precision. The Faster R-CNN algorithm is a popular computer vision object detection method because it is more efficient and effective. Applications include image classification, object recognition, and object tracking.

2) Mask RCNN: Mask R-CNN extends Faster R-CNN. Mask R-CNN detects objects in an image and creates a segmentation mask for each object. Mask R-CNN can identify the visual pixels corresponding to each identified object, enabling more precise and complete object localization [17]. Mask R-CNN uses a region proposal network to find candidate bounding boxes in the input image. The retrieved characteristics categorise regions as objects or backgrounds. Mask R-CNN uses a fully convolutional network to segment each object. Image and video segmentation, medical image analysis, and robotics use Mask R-CNN. It outperforms many benchmark datasets.

B. YOLO V5

You Only Look Once (YOLOv5) is a real-time object detection model. It is a CNN architecture for rapid and accurate image and video object recognition. YOLOv5 has convolutional layers followed by fully connected layers [18]. Convolutional layers extract characteristics from input data, while fully connected layers identify objects and anticipate their bounding boxes. Anchor boxes are crucial to the YOLOv5 design. Anchor boxes match expected and ground truth bounding boxes during training. This improves the model's accuracy by teaching it about different object types and their shapes and sizes. Multi-scale object detection is another significant feature of YOLOv5. The model analyses input data at numerous scales to detect objects of varied sizes in the same image. Convolutional layers with varying strides downsample input data at different scales. The YOLOv5 architecture is fast and accurate and has achieved state-of-the-art results on several object identification benchmarks.

In this work, we've used YOLOv5 in the context of self-driving cars. It can enable autonomous vehicles to detect and classify objects in their environment, such as approaching vehicles, pedestrians and traffic signs. This is an essential capability for autonomous vehicles as it allows vehicles to make decisions about how to navigate their environment safely. YOLOv5 is commonly used in a variety of applications. YOLOv5 is a very effective and widely applied instrument for identifying real-time objects in various applications. This uses an anchor box to provide better bounding box predictions, which is useful when detecting objects on roads with various sizes and aspect ratios in the same image.

V. Designed Algorithm to Process Data Discovered by V2X Modules Using Deep Learning

The proposed machine learning (ML) algorithm is shown in Figure 3, designed to detect pedestrians and animals on highways using data from sensors and cameras. The algorithm collects data from these sources and processes it to filter out noise and irrelevant information. This process is known as preprocessing. After the data has been preprocessed, the developed ML algorithm analyses the data and detect pedestrians and animals. The algorithm sends a warning to nearby vehicles if a pedestrian or animal is detected through a v2x (vehicle-to-everything) communication module. If no pedestrians or animals are detected, the algorithm continues to monitor the environment for any changes. The process of collecting data, preprocessing it, and analyzing it to detect pedestrians and animals is repeated continuously to ensure real-time detection. This is important because it allows the algorithm to respond to environmental changes as they occur and provide timely warnings to nearby vehicles. Overall, the described ML algorithm is designed to improve safety on highways by detecting pedestrians and animals and alerting nearby vehicles to their presence. It uses a combination of sensors, cameras, and machine learning techniques to collect and analyze data in real-time, providing a valuable tool for improving road safety.

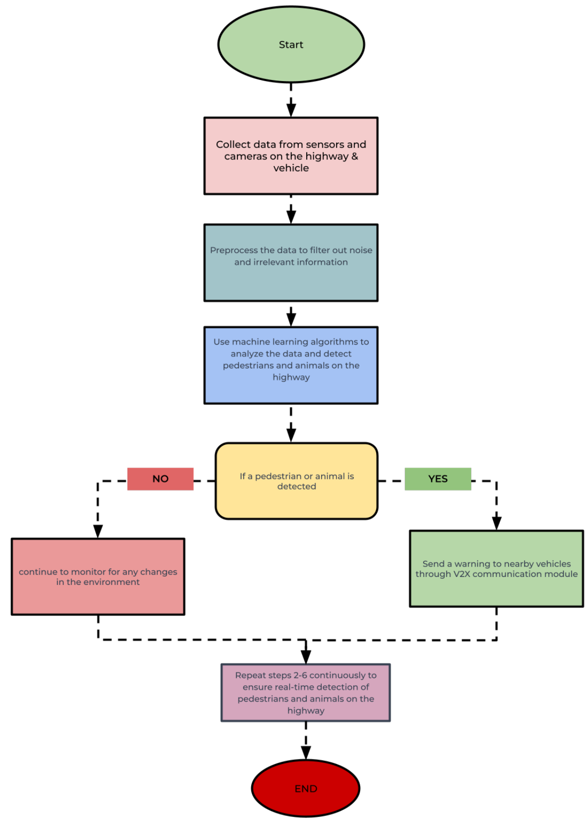

Fig. 3. Flowchart for the proposed system

VI. Implementation and Results

This section describes the output of the experiments carried out on the proposed system, The Yolo V5 model is being trained on a dataset, using 80% of the data for training and reserving 20% as holdout (unseen) data for testing. The optimisation algorithm used for training is SGD (Stochastic Gradient Descent), an iterative method for optimising a differentiable objective function by updating the model's parameters in the direction of the negative gradient of the objective function with respect to the parameters. The loss function being used is cross-entropy loss, which is a commonly used loss function for classification tasks. It measures the distance between the predicted probability distribution over classes and the true probability distribution and is used to train the model to classify correctly.

The model is being implemented in Keras, a high-level Python library for building and training deep learning models, using TensorFlow as the back end. TensorFlow is an open-source software library for machine learning and numerical computation that is used to power many of the largest machine learning models in production today. Google Collaboratory is a free Jupyter notebook environment provided by Google that allows for easy prototyping and testing of machine learning models [19]. The model is being trained using mini-batches of 128 examples, with 100 epochs and a learning rate of 10^{-4} in each epoch. An epoch is a full pass over the training dataset. The learning rate determines the step size at which the optimiser updates the model parameters. A lower learning rate means the optimiser will take smaller steps and require more epochs to find a good solution, while a larger learning rate means the optimiser will take larger steps and potentially find a good solution more quickly, but also with a higher risk of overshooting and not converging.

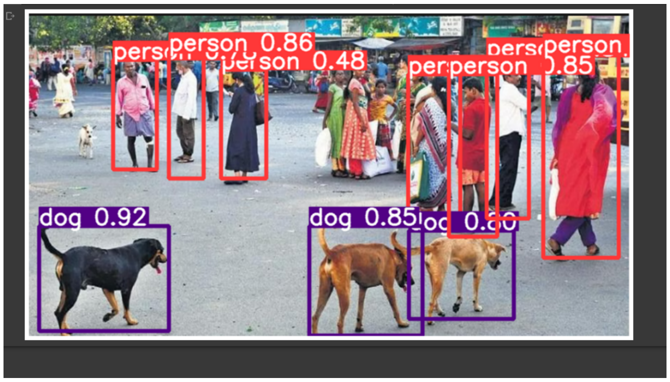

Fig. 4. Output showing pedestrian crossing and detection by the developed machine learning system

Fig. 5. Output showing animal and pedestrian crossings and detection by the developed machine learning system

Figure 4 & 5 shows the output of a proposed object detection model designed to detect animals and pedestrians on roads. Object detection models are machine learning models trained to identify specific objects within an image or video. In this case, the model is specifically designed to detect animals and pedestrians on roads, where the objects of interest (animals and pedestrians) are highlighted with bounding boxes. These bounding boxes are used to identify the detected object's location within the image. The bounding box's colour and thickness can indicate the model's confidence level in the detection. The bounding boxes correctly enclose the objects in the image, indicating that the model has correctly identified their location. The model can also identify animals and pedestrians in different positions on the road, in walking directions and positions.

The ability of the proposed object detection model to accurately detect animals and pedestrians on roads could have important applications in a variety of areas, such as autonomous vehicles and traffic monitoring. For example, an autonomous vehicle equipped with this model could use the information to avoid collisions with animals and pedestrians on the road, making it safer for all involved. Additionally, the model can be used for tracking and monitoring animal movements for conservation and research purposes.

A. *Efficiency matrices for the Machine Learning Model*

1) Precision: is a measure of the accuracy of the model's positive predictions, defined as the number of true positive predictions divided by the total number of positive predictions made by the model [20] A high precision score indicates that the model is good at not labelling a negative example as positive.

2) Recall: is a measure of the model's completeness, defined as the number of true positive predictions divided by the total number of actual positive examples [21]. A high recall score means that the model can identify most of the positive examples in the dataset.

3) F1-Score: is a measure of a model's precision and recall balance, defined as the harmonic mean of precision and recall [22]. It is a useful metric to consider when balancing precision and recall.

4) Accuracy: is a measure of the model's overall accuracy, defined as the number of correct predictions made by the model divided by the total number of predictions made [23].

5) Mean Average Precision: (mAP) is a measure of the model's average precision across all classes, calculated by averaging the precision scores at different recall levels for each class [24]. It is a common metric used for object detection tasks.

The developed object detection algorithm using the Yolo V5 model performs well on the animal detection task, with a precision of 0.91, recall of 0.84, F1-Score of 0.87, the accuracy of 0.87, and mAP of 0.90 as shown in table I.

TABLE I
EFFICIENCY MATRICES OF THE MACHINE LEARNING ALGORITHM

Efficiency Matrix	Score
Accuracy	0.87
Precision	0.91
Recall	0.84
F1-Score	0.87
mAP	0.90

From the table I, it is understood that the model can identify animals with a high level of accuracy and can do so without missing a large number of positive examples (high recall) or making many false positive predictions (high precision). Overall, the model appears well-suited for detecting animals in images or videos from roadside cameras. The animals and pedestrians identified using the YoloV5 algorithm are then passed to the V2X network from the roadside or vehicle-mounted cameras. The V2X stack processes the images, verifies and then disseminates the information to other V2X users.

VII. Conclusion

The YOLO (You Only Look Once) analysed object detection model is a fast and accurate method for detecting objects in images and videos. The YOLO model is trained on a large dataset to recognize various things, including pedestrians and animals. Our proposed algorithm uses the YOLO model to detect pedestrians and animals in V2X (vehicle-to-everything) characterised V2X technology enables the communication between vehicles and other road users (e.g., pedestrians, animals, infrastructure), which can help to improve highway safety by providing real-time information about the location and movements of these road users. By using the YOLO 5 model to detect pedestrians and animals in V2X environments, our algorithm can provide timely and accurate information about the presence and movements of these road users. This can help maximise highway safety by allowing vehicles to avoid collisions with pedestrians and animals and providing early warning of potential hazards. By providing real-time information about the presence and movements of derecognised animals, our AI-enabled object detection algorithm can help to reduce the risk of accidents and improve overall highway safety.

In future, we would like to Integrate V2X (vehicle-to-everything) communication technology to enable real-time communication and coordination between vehicles, pedestrians, and infrastructure. Also, we are looking into the possibility of using unmanned aerial vehicles (UAVs) for monitoring and surveying the highway environment to improve the detection of pedestrians and animals in remote or hard-to-reach areas. Further, test and validate the proposed system in real-world settings, including infield testing on public roads.

Acknowledgment

To our beloved Chancellor and world-renowned humanitarian leader, Sri. Mata Amritanandamayi Devi (AMMA), our Provost and Director, Dr Maneesha Vinodini Ramesh, and faculty and staff members for their support and motivation towards all our endeavours.

References

[1] Q. Cai, "Cause analysis of traffic accidents on urban roads based on an improved association rule mining algorithm," *IEEE Access*, vol. 8, pp. 75 607–75 615, 2020.

[2] M. F. Labib, A. S. Rifat, M. M. Hossain, A. K. Das, and F. Nawrine, "Road accident analysis and prediction of accident severity by using machine learning in bangladesh," in *2019 7th International Conference on Smart Computing & Communications (ICSCC)*, 2019, pp. 1–5.

[3] K. Sabry and M. Emad, "Road traffic accidents detection based on crash estimation," in *2021 17th International Computer Engineering Conference (ICENCO)*, 2021, pp. 63–68.

[4] R. Kasimani, M. P, R. M, and P. Sivaprakash, ""a review on road traffic accident and related factors"," *International Journal of Applied Engineering Research*, vol. 10, pp. 28 177–28 183, 01 2015.

[5] S. Santhanam, S. S. B, S. S. Panigrahi, S. K. Kashyap, and B. K. Duriseti, "Animal detection for road safety using deep learning," pp. 1–5, 2021.

[6] M. Gaushik, M. Jivthesh, S. Siji Rani, N. Sai Shibu, and S. N. Rao, "Architecture design of ai and iot based system to identify covid-19 protocol violators in public places," pp. 1–6, 2021.

[7] P. Ulleri, M. S., S. K., K. Zenith, and S. S. N.B., "Development of contactless employee management system with mask detection and body temperature measurement using tensorflow," pp. 235–240, 2021.

[8] P. Vallimeena, U. Gopalakrishnan, B. B. Nair, and S. N. Rao, "Cnn algorithms for detection of human face attributes – a survey," pp. 576–581, 2019.

[9] P. Vallimeena, B. B. Nair, and S. N. Rao, "Machine vision based flood depth estimation using crowdsourced images of humans," pp. 1–4, 2018.

[10] R. L. Galvez, A. A. Bandala, E. P. Dadios, R. R. P. Vicerra, and J. M. Z. Maningo, "Object detection using convolutional neural networks," in *TENCON 2018 - 2018 IEEE Region 10 Conference*, 2018, pp. 2023–2027.

[11] X. Liu, Z. Jia, X. Hou, M. Fu, L. Ma, and Q. Sun, "Real-time marine animal images classification by embedded system based on mobilenet and transfer learning," in *OCEANS 2019 - Marseille*, 2019, pp. 1–5.

[12] A. Gomez Villa, A. Salazar, and F. Vargas, "Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks," *Ecological Informatics*, vol. 41, pp. 24–32, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574954116302047

[13] R. L. A, A. K. S, K. B. E, A. N. D, and K. K. V, "A survey on object detection methods in deep learning," in *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2021, pp. 1619–1626.

[14] A. Hijaz, W.-Y. G. Louie, and I. Mansour, "Towards a driver monitoring system for estimating driver situational awareness," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2019, pp. 1–6.

[15] "Papers with code - r-cnn explained." [Online]. Available: https://paperswithcode.com/method/r-cnn

[16] R. Girshick, "Fast r-cnn," 2015. [Online]. Available: https://arxiv.org/abs/1504.08083

[17] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," 2017. [Online]. Available: https://arxiv.org/abs/1703.06870

[18] "Pytorch." [Online]. Available: https://www.pytorch.org

[19] E. Bisong, *Google Colaboratory*. Berkeley, CA: Apress, 2019, pp. 59–64. [Online]. Available: $https://doi.org/10.1007/978-1-4842-4470-8_7$

[20] "What is accuracy and precision in machine learning?" Dec 2020. [Online]. Available: https://stackhowto.com/what-is-accuracy-and-precision-in-machine-learning/

[21] K. Ganesan, "What is precision and recall in machine learning?" Aug 2021. [Online]. Available: https://kavita-ganesan.com/precision-and-recall-machine-learning/

[22] "Hasty ai." [Online]. Available: https://hasty.ai/docs/mp-wiki/metrics/f-beta-score

[23] Abhigyan, "Calculating accuracy of an ml model." Jul 2020. [Online]. Available: https://medium.com/analytics-vidhya/calculating-accuracy-of-an-ml-model-8ae7894802e

[24] "Hasty ai." [Online]. Available: https://hasty.ai/docs/mp-wiki/metrics/map-mean-average-precision