Attention-Based Deep Learning Approach for Pedestrian Detection in Self-Driving Cars

Wael Ahmad AlZoubi[1], Prof. Girish Bhagwant Desale[2], Dr. Sweety Bakyarani E[3],
Dr Uma Kumari C R[4], Dr. Divya Nimma[5], K Swetha[6], Dr B Kiran Bala[7]

Applied Science Department, Ajloun University College, Al-Balqa Applied University, Jordan[1]
HOD, Department of Computer Science & IT, JET'S Z. B. Patil College, Dhule. (M.S.), Jalgoan, India[2]
Department of Computer Science-Faculty of Science and Humanities,
SRM Institute of Science and Technology, Kattankulathur, India[3]
Associate Professor, Department of Electronics and Communication Engineering, KCG College of Technology, Chennai, India[4]
PhD in Computational Science, University of Southern Mississippi, Data Analyst in UMMC, USA[5]
Assistant Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India[6]
Head of the Department-Department of Artificial Intelligence and Data Science,
K. Ramakrishnan College of Engineering, Trichy, India[7]

***Abstract*—Autonomous vehicle safety relies heavily on the ability to accurately detect pedestrians, as this capability is crucial for preventing accidents and saving lives. Pedestrian recognition is particularly challenging in the dynamic and complex environments of urban areas. Effective pedestrian detection is crucial for ensuring road safety in autonomous vehicles. Current pedestrian identification systems often fall short in capturing the nuances of pedestrian behavior and appearance, potentially leading to dangerous situations. These limitations are mainly due to difficulties in various conditions, such as low-light environments, occlusions, and intricate urban settings. This paper proposes a novel solution to these challenges by integrating an attention-based convolutional bi-GRU model with deep learning techniques for pedestrian recognition. This method leverages deep learning to provide a robust solution for pedestrian detection. Convolutional layers are utilized to extract spatial features, attention mechanisms highlight semantic details, and Bidirectional Gated Recurrent Units (Bi-GRU) capture the temporal context in the proposed model. The process begins with data collection to build a comprehensive pedestrian dataset, followed by preprocessing using min-max normalization. The key components of the model work together to enhance pedestrian detection, ensuring a more accurate and comprehensive understanding of dynamic pedestrian scenarios. The implementation of this unique approach was carried out using Python, employing libraries such as TensorFlow, Keras, and OpenCV. The proposed attention-based convolutional bi-GRU model outperforms previous models by an average of 17.1%, achieving an accuracy rate of 99.4%. The model significantly surpasses Random Forest, Faster R-CNN, and SVM in terms of pedestrian recognition accuracy, which is critical for autonomous vehicle safety.**

Keywords*—Pedestrian recognition; autonomous vehicle safety; deep learning; attention mechanism; Bidirectional Gated Recurrent Units

I. INTRODUCTION

One of the most important aspects of road transport, and a key issue in the automotive industry, is vehicle safety. This includes a wide range of techniques and strategies aimed at reducing the likelihood of a collision and, if it occurs, the number of injuries to car occupants and other road users [1]. To help establish safety norms and standards, government agencies and independent agencies conduct extensive safety tests to assess how vehicles perform in collision situations Organizational development these tests include the Insurance Institute for Highway Safety (IIHS) and the National Highway Traffic Safety Administration (administered by the NHTSA) [2]. The use of advanced driver assistance systems (ADAS) in vehicles has become a necessity. With features such as parking assist, adaptive lighting, blind spot monitoring and collision avoidance technology, this system supports drivers and improves safety. While there are more hybrid electric vehicles their security concerns increase. It is important to guarantee that the high voltage electrical system is crashworthy, the battery is integrated and safe [3]. Car safety precautions include child safety. This includes instructing parents to check child rear seats before leaving the vehicle, using child safety seats, and integrated child safety doors Cybersecurity is a growing threat due to increased connectivity and carry reliance on software in vehicles therefore. Protecting vehicles from cyberattacks and ensuring the integrity of critical software systems is more important than ever [4].

Application of machine learning techniques in automotive safety brings a new era of aggressive crash avoidance and occupant protection Advanced Driver Assistance Systems (ADAS) are commonly used machine learning systems role results. These systems use algorithms to analyze sensor data from cameras, radar, lidar and ultrasonic sensors to detect potential hazards in crashes, turns and even drowsy driving [5]. This technology provides real-time warnings or interventions to help drivers avoid accidents. Machine learning is essential for autonomous vehicles to recognize and understand their surroundings, recognize traffic signs, people, and other moving objects and make decision calls to help them travel safely Predictive maintenance. which in turn uses machine learning to monitor the vehicles' own health and predict any problems before failure results [6]. This technology provides real-time warnings or interventions to help drivers avoid accidents.

Machine learning is essential for autonomous vehicles to recognize and understand their surroundings, recognize traffic signs, people, and other moving objects and make decision calls to help them travel safely Predictive maintenance. which in turn uses machine learning to monitor the health of the vehicles themselves [7]. By helping to identify and prevent any potential cyberattacks on the vehicle's systems, machine learning helps improve automotive cybersecurity. Vehicle software can be protected from unauthorized access and data breaches using machine learning algorithms that can detect anomalies and strange behavior [8]. Machine learning in automotive safety using real-time data analytics, predictive capabilities and continuous algorithmic improvements is an ongoing research area that can improve road traffic safety necessary to reduce traffic-related incidents , injuries and deaths to improve driving effectiveness and overall safety [9].

Deep learning has been shown to be a revolutionary force in automotive safety prediction, radically changing how we predict and avoid accidents on the road. With large amounts of data coming from many sources including as sensors, cameras and vehicle control systems, deep learning methods especially neural networks are used These methods can identify complex patterns, identify threats a it is possible and shows important information in real time. Deep learning algorithms play a key role with advanced driver assistance systems (ADAS) by testing the environment, detecting objects and predicting collision hazards. This enables the system to issue warnings in a timely manner or even preventive measures to prevent accidents. Deep learning in predictive maintenance is important because it can analyze sensor data to identify technical problems in the vehicle before they become dangerous.

Feature extraction is necessary to identify pedestrians from other features in and around an image or video frame Before manual feature engineering; However, with the development of deep learning, this approach is now widely incorporated into web design [10]. The conventional Histogram of Oriented Gradients (HOG) is one of the more popular extraction techniques. Unlike computer histograms that show gradient orientations in these cells by image segmentation to extract important edge information and shape information, local binary models (LBP) provide information about garment textures by encoding local texture patterns by comparing the pixel intensities with their neighbors is very useful in distinguishing between a pedestrians [10]. Also, a deep learning technique, convolutional neural networks (CNNs) are popular in feature extraction. These networks have greatly enhanced pedestrian recognition by learning and extracting features such as shapes, textures, and context [11].

Ensuring the safety of motorists and pedestrians sharing the road is of utmost importance. Accurate pedestrian detection and identification, enabled by state-of-the-art deep learning algorithms, is essential to the success of our efforts. By incorporating cognitive algorithms into a convolutional bi-directional gated recurrent unit (Bi-GRU) model for pedestrian recognition, this research aims to push the limits of autonomous vehicle safety [12]. This process improves model understanding in dynamic pedestrian scenarios using Bi-GRU to store time context in addition to the ability of convolutional neural networks (CNNs) to capture detailed scene details Feature those in ambiance, those in ambiance and crowds [13]. This ubiquitous approach is intended to greatly improve the accuracy and reliability of pedestrian detection methods, making autonomous vehicles safer and more responsive is more adept at dealing with real-world obstacles. In this work, we explore the key features and techniques we use in our model and create findings that reveal the potential to change the automotive safety feature of and about the whole new.

Academics and engineers have made great strides to improve autonomous vehicle safety as new techniques for deep learning model design and data preprocessing This review focuses on the use of Min-Max normalization, which provides the input data is better and increases the accuracy of the model in pedestrian detection Established it is optimized and standardized, resulting in more consistent and reliable results. In addition, the addition of the Bi-GRU (Bidirectional Gated Recurrent Unit) layer in the model framework introduces an additional period of pedestrian detection This layer provides the model with an important contextual understanding about dynamic pedestrian conditions because it can capture time dependence in input data Bidirectional information processing is possible, and it greatly improves the safety of the autonomous vehicle by enabling the image to carry protective information. In addition, the addition of focus improves the accuracy of the model in detecting pedestrians, especially in complex and multidimensional environments where pedestrian access or movement may be obstructed is unobservable, this dynamic distribution of weights across data points assures that the model focuses on the most relevant cases. Attention is a variable that enhances pedestrian visibility and contributes significantly to the overall safety of involved vehicles by reaching critical objects faster. If driven together, these developments make a significant contribution to the field and point to future advancements in autonomous driving safety and reliability.

Key contributions of this framework are,

- The proposed system increases the ability of the model to accurately detect pedestrians and enhance the input data quality by using Min-Max normalization. This pre-processing step provides reliable and consistent results.
- The two GRU layer uses the temporal relationships of input data to provide pedestrians with a description of the relevant context. The use of two-way information processing enhances the understanding of the dynamic pedestrian model and contributes to the safety of autonomous vehicles.
- The model focuses on the most pertinent information thanks to the attention layer, which dynamically distributes weights to different data regions. This is essential for precise pedestrian detection in complex scenarios. This technique enhances autonomous vehicle safety by more effectively acquiring and processing critical attributes.
- The convolutional layer improves the model's capacity to identify complex shapes, patterns, and textures in images, which is crucial for precise pedestrian identification in a variety of settings. By incorporating it, autonomous vehicle safety is further reinforced by

ensuring a more thorough grasp of dynamic pedestrian circumstances.

The research's remaining portions are listed below. Section II provides a review of the literature. The issue statement is covered in Section III. The suggested technique for detecting neurological diseases is then covered in Section IV. Section V discusses the results and discussions. Section VI discusses the conclusion.

II. Related Work

Chen et al. [14] proposes a thorough method for deep learning-based pedestrian recognition and tracking in challenging circumstances, with an emphasis on problems like small-target people and partially obstructed pedestrians in crowded areas. A notable improvement to the YOLO detection algorithm involves the addition of channel attention as well as spatial attention modules. The capacity of the model to represent crucial feature information is eventually improved by these modules, which aid in amplifying important feature data in both channels and spatial dimensions. These modules' incorporation into Darknet-53's backbone network exemplifies an organized method of feature extraction. It is a well-known choice to use the DeepSort tracking technique in conjunction using the Kalman filter algorithm to estimate the mobility status of pedestrians. The tracking procedure is further strengthened by the Mahalanobis distance and evident feature for similarity estimates and the Hungarian method for ideal target matching. This method strengthens the tracking system by integrating deep learning for identification with conventional tracking methods. It is necessary to do preliminary testing of the enhanced YOLO pedestrian detection system and DeepSort tracking technique in the same setting. The findings show a considerable decrease in missed detections and false positives, a more robust handling of tracking failures caused by occlusion, and an essential increase in detection precision for small-target pedestrians. These encouraging results point to the suggested approach's efficacy in enhancing the functionality of intelligent vehicle-pedestrian identification and tracking systems, particularly in challenging and congested environments.

Lu et al. [15] proposes an innovative method for predicting vehicle trajectory for connected and autonomous cars that operate in mixed traffic situations. For these vehicles to be safe and sustainable, accurate forecasting of trajectory is essential. This task is complicated by several variables, including individual vehicle motions, the state of the road, and interactions with other nearby cars. This problem is successfully addressed by the suggested method, Heterogeneous Context-Aware Graph Convolutional Networking with Encoder-Decoder architecture, which simultaneously extracts hidden contexts from each historical trajectory, the dynamic driving scene, and the interactions between vehicles. The design of the method, which uses 2-dimensional Convolutional Networks with temporal attention to represent the changing scene context from scene photos and Temporal Convolutional Networks to capture personal environments from previous vehicle trajectories, is praiseworthy. Spatial-temporal dynamic Graph Convolutional Networks, which incorporate both individual and scene settings as node representations, are a powerful option for modelling inter-vehicle interaction patterns. It is well thought out to combine these three circumstances and use them as input to the decoder to produce future trajectories. The credibility of the study is increased by the verification of the proposed framework on real-world datasets comprising various driving circumstances. The findings validate the assertion that the model beats state-of-the-art techniques in terms of prediction accuracy and stability regardless of vehicle conditions. The vehicle trajectory prediction field for connected and autonomous cars gains significantly from the work presented in this paper, which offers a thorough and practical solution to complicated real-world traffic situations [16].

Pustokhina et al. [17] addresses an important problem essential to improving the safety of vulnerable road users: the actual detection of anomalies in pedestrian areas. The expanded network of surveillance cameras and the amount of film recorded make manual verification and labeling of anomalies difficult and time-consuming As a result, surveillance systems automation, especially those based on deep learning theories, has become increasingly popular in computer vision. To overcome these challenges, our research focuses on the development of a deep learning anomaly detection technique (DLADT-PW) for pedestrian roads. The aim of the DLADT-PW method is to detect and classify different types of deformation in pedestrian traffic, such as the presence of cars, skateboarders, or jeeps Mask area convolutional neural network (Mask-RCNN) incorporating a dense connected mesh work in models after the pre-processing stage in order to reduce noise and detect. The image can be enhanced before phase. The study provides evidence of the effectiveness of the DLADT-PW method by running detailed simulations and analyzing the results from different angles. The findings confirm the excellent performance of the DLADT-PW model and demonstrate its potential for accurate identification. By developing a computerized anomaly detection algorithm that can quickly and segmentally detect anomalies in pedestrianized roadways, this research significantly enhances the field of computer vision, including pedestrian safety. In this context, deep learning models are used, especially masked RCNN and DenseNet, and emphasis is placed on implementing state-of-the-art techniques. The capabilities of the model are further confirmed through comprehensive analysis and empirical analysis, indicating the potential for useful application in real-world monitoring systems for providing vulnerable road users security has increased.

Islam et al. [18], it introduces a vision-based approach to customized safety messages (PSMs), which addresses an important part of passenger safety in Internet-connected vehicles in vehicular Internet-connected (V2P) communication is needed in situations out of this, but pedestrians face a major obstacle as it relates to the lack of low-level connectivity with nearby connected vehicles -which involve telephony or especially- long distance communication (DSRC) devices. The main contribution of the study is a vision-based system of real-time PSM using video feeds from roadside traffic cameras according to SAE J2945 According to the paper’s analysis, safety for pedestrians has been provided the planned netted vehicle is an example of the feasibility of this strategy. The presented results highlight the effectiveness of the vision-based system in real-time collision warning to avoid future vehicular and pedestrian accidents, especially in time-to-collision (TTC) values in the emphasis. The study confirms that, in a connected vehicle context, our vision-based pedestrian safety notification system satisfies the latency criteria for the PSCW safety application. The important topic of pedestrian security within the environment of linked automobiles is addressed in this research, which is noteworthy. Its vision-based methodology provides a novel approach to the problem of pedestrian communication in the absence of specialized equipment. The creation of a useful safety application and the approach's proven effectiveness in real-time collision warning serves to further highlight its potential. Its legitimacy and application in the automobile sector are improved by adhering to SAE standards. To offer a more thorough understanding of the suggested approach, the study might utilize more information on the technical facets of the vision-based system, the specific data streams employed, and potential difficulties in real-world application.

The examined literature focuses on important advancements in deep learning and computer vision for problems relating to pedestrian safety and traffic safety. In the study, a better YOLO-based system for pedestrian tracking and recognition is introduced. By using attention modules to solve problems such as tiny targets and occluded pedestrians, the system improves detection accuracy and provides stable tracking in congested situations. In different research, a novel method for simulating complicated vehicle interactions and improving trajectory prediction accuracy is presented for predicting vehicle trajectories in mixed traffic settings using graph convolutional networks. A thorough analysis of current LSTM-based short-term traffic prediction algorithms is presented, together with a classification and assessment of their advantages and disadvantages, to help researchers choose the best method for certain traffic situations. Another study explores automated anomaly detection at pedestrian crossings and presents a Deep Learning anomaly detection technique that can identify a wide range of anomalies and improve pedestrian safety. Using roadside traffic cameras and adhering to industry standards, a vision-based system for generating Personal Safety Messages in connected vehicle contexts is discussed, enabling real-time collision warnings and ultimately enhancing pedestrian safety and communication in the absence of specialized equipment. Through creative approaches and useful solutions, these studies jointly increase traffic forecasts, intelligent transportation systems, and pedestrian safety.

III. Problem Statement

Pedestrian safety in the context of autonomous vehicles is still a major concern, according to the literature review, which draws from the discussion above. This is especially true in situations where there are small targets among the pedestrians, partial obstructions in crowded areas, and no personal communication devices. While recent research shows that deep learning, better detection algorithms, and vision-based techniques have potential, a holistic and cohesive solution is required to maximize pedestrian recognition and tracking for autonomous vehicle safety. Thus, Convolutional Bi-GRU, a concept-based model that combines traditional detection methods with deep learning for accurate and reliable pedestrian identification [19]. The main objective is to increase the detection accuracy of target pedestrians, and to reduce missed detections, false positives, and failures due to occlusion and tracking failures so especially the research objective of improving the safety of autonomous vehicles in harsh and crowded situations is maximized. The advantages of recurrent and convolutional neural networks are well combined in the Conv-Bi-GRU model. CNNs perform exceptionally well when extracting features from images, and this integration improves the model’s ability to extract and analyze spatial and temporal complexity from video data, and provides improve our understanding of the temporal relationships observed in Bi-GRU walking trajectories.

IV. Proposed Framework of Attention Based Convolutional Bi-GRU Model

The approach adopted in this study is based on a structured design with several main components. In order to train and test the model, in the first phase of data collection, a large dataset of pedestrians should be obtained. The next step in processing the data is minimum-maximum normalization, which involves scaling the pixel values to the standard range. Concept-based convolutional two-GRU model, a deep learning system designed for pedestrian detection is the core of the methodology. This model has several layers, such as a conceptual approach for highlighting relevant semantic features in the data and bidirectional gated repetitive units (Bi-GRU) for capturing time descriptions. The data are extracted with convolutional layers of space objects to better understand pedestrians’ events. A fully connected layer and SoftMax activation provide the final result, enabling the model to better classify pedestrians. Considering all things, this technology is designed to improve the safety of autonomous vehicles by preventing crashes and accidents, in addition to being exceptionally effective at detecting pedestrians and giving way it is possible to approach the front with an autonomous vehicle. Fig. 1 shows the block diagram of the proposed model.

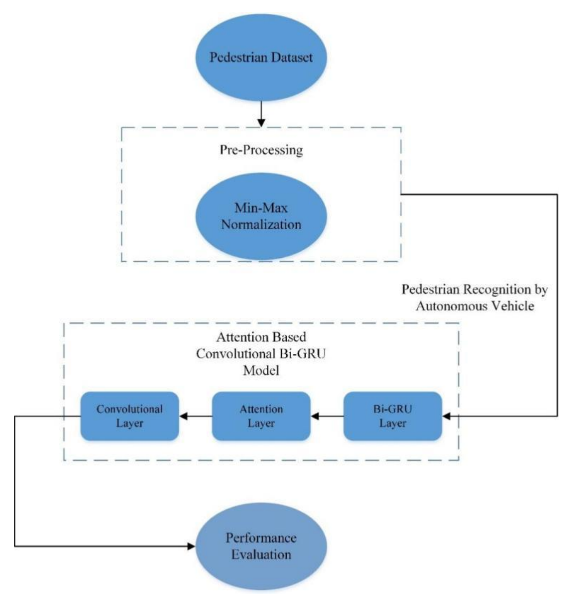

Fig. 1. Block diagram of proposed attention based convolutional Bi-GRU model.

A. Data Collection

The pedestrian dataset used in this study, sourced from Kaggle, consists of extensively documented videos capturing people using crosswalks in various contexts. The dataset is divided into three distinct segments, each providing valuable insights into pedestrian behavior. The first segment, titled "Crosswalk," is a 12-second clip showing individuals safely crossing the street using designated means, offering a clear understanding of compliance with pedestrian safety regulations. The second segment, a 25-second video titled "Night," highlights pedestrian behavior in low-light conditions, providing insights into how lighting influences crosswalk use. The dataset includes accurate bounding box annotations for pedestrians, meticulously encoded in a .csv file, making it highly suitable for analyzing and training machine learning models. For training and testing, the dataset was split into 70% for training and 30% for testing, ensuring a balanced evaluation of the model's performance [20].

A. Pre-Processing using Min-Max Normalization

Min-max scaling is a data processing technique, sometimes called normalizing, in which the pixel values of images in a data set are converted to a specific range, usually 0 to greater in computer vision, including safety-related applications for pedestrian detection and new autonomous vehicles. If you want to make sure that each pixel value falls within a defined, constrained range, min-max scaling comes in handy. This creates a model of deep learning and a smooth meeting [21]:

$$x_{sv} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

whereas, the scaled value is x_{sv}. The initial pixel value is represented by x. The smallest pixel value in the collection is denoted by x_{min} . The dataset's highest pixel value is represented by the value x_{max}.

This preprocessing step enhances model performance by reducing the impact of varying pixel intensities and ensuring that the model learns from data with consistent value ranges. In the context of pedestrian detection, this consistency improves the model's ability to accurately recognize and differentiate pedestrians from background elements, leading to better detection accuracy and robustness in safety-critical applications.

B. Pedestrian Recognition Using Attention-based Convolutional Bi-GRU Model to Improve Autonomous Vehicle Safety

Using the Attention-Based Convolutional Bi-GRU Model and deep learning for pedestrian recognition, this model optimizes autonomous vehicle safety by utilizing pre-trained feature vectors in the input layer to represent the important attributes related to pedestrian recognition. This study may effectively examine and comprehend the many attributes of pedestrians due to these vectors, which are pre-configured to encode pertinent information about them. Preprocessed data is fed into the input layer in the form of images. These images have usually been through operations like scaling, cropping, and potentially color normalization. The input layer is where the data enters the neural network; it doesn't do any computation.

$$X_i = (x_1, x_2, \dots \dots \dots x_{n-1}, x_n) \tag{2}$$

In Eq. (2), the feature vector of the i^{th} pedestrian data instance is represented by X_i, the attribute vector of the i^{th} characteristic inside the pedestrian data is represented by x_i, and the length or dimensionality of the input data is indicated by X. An image sequence's temporal dependencies are intended to be captured by the Bidirectional Gated Recurrent Unit (Bi-GRU) layer. It has the ability to analyze data both forward and backward, which allows it to gradually capture context. The temporal properties of the input data are represented as a series of feature vectors as this layer's output. As illustrated in Fig. 2, this model utilizes visual features that reflect pedestrian scenarios as input data. The Bi-GRU layer is split into two halves that each evaluate the image features in both forward and reverse directions at the same time. The GRU algorithm then processes these successive picture features and produces an output vector with specified dimensions. Four crucial computing components are involved in the GRU action. The reset gate is part of the initial element.

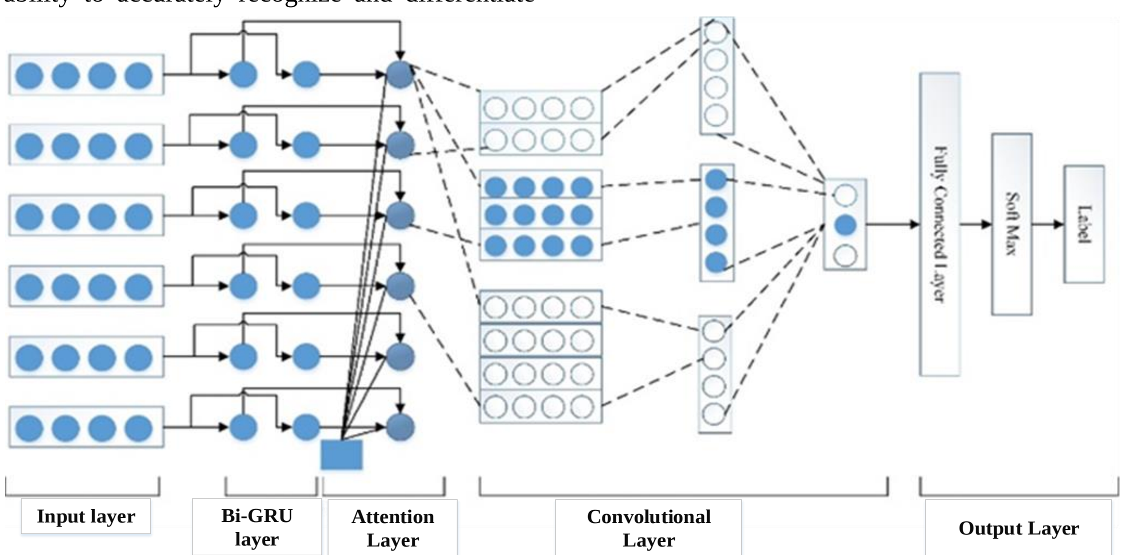

Fig. 2. Architecture of proposed attention-based convolutional Bi-GRU model.

$$Re_t = \tau(w_r x_t + u_r h_{t-1} + b_r) \tag{4}$$

$$\mathbb{Z}_t = \tau(w_z x_t + u_z h_{t-1} + b_z) \tag{5}$$

In order to determine the output of the current instant, GRU first computes the candidate memory content from Eq. (6), where w and u represent weight information and b represents bias.

$$h^*{}_t = \tanh(wx_t + uRe_t h_{t-1} + b) \tag{6}$$

The results from the preceding mentioned operations are the basis on which the GRU ultimately computes the output using Eq. (7). The Bi-GRU layer can record contextual information about every component in the input data for pedestrian identification.

$$h_t = (1 - \mathbb{Z}_t)h^*{}_t + \mathbb{Z}_t h_{t-1} \tag{7}$$

To concentrate on pertinent areas or frames within the input data, the attention mechanism is essential. In order to emphasize areas of interest, it dynamically assigns weights to various input sequence segments. This is especially crucial for spotting pedestrians in complex circumstances. The input is represented in an attention-weighted manner by this layer. Each Bi-GRU layer output is given weight by the attention mechanism, as shown in Fig. 2. The weight's magnitude indicates how important the semantic material is. To put it simply, locations with greater weights attract more "Attention," a sign that they have a significant influence on the final pedestrian recognition categorization result. For computing attention, the formula is denoted by Eq. (8) and Eq. (9). The attention weight a_{ik} and the related weight parameters (w_{a2} and w_{a1}) are represented in Eq. (10), together with the sequence data length t_h.

$$G_t = \sum_{t=1}^{t_h} a_{ik}\, h_t \tag{9}$$

$$a_{ik} = softmax(w_{a2} \tanh(w_{a1} h_t)) \tag{10}$$

Applying the Convolutional Layer to the spatial representation of the input data is common practice in Convolutional Neural Networks (CNNs). It is essential for recognizing pedestrians' visual traits, such as edges, forms, and textures, since it extracts pertinent aspects and patterns from the attention-weighted input. The computation of the convolution layer is carried out as follows:

The convolution layer receives its input from the attention layer's generated intermediate semantic information, where $G_{i:k}$ denotes the grouping of features pertaining to the i^{th} and k^{th} items in the context of pedestrian recognition.

$$G_{i:k} = G_i \oplus G_{i+1} \oplus \ldots \oplus G_k \tag{11}$$

Three different convolutional kernels are applied to obtain deep features. Throughout this procedure, 'f' stands for a hyperbolic tangent function, w for the weight parameters, ' m ' for the convolution kernel width, and ' b ' for the bias term.

$$t_i = f(wG_{i:k+m-1} + b) \tag{12}$$

After that, the model uses maximum pooling to merge the convolution results and identify important features. The final output of the entire convolution layer is then formed by combining the pooled results. 'n' stands for the number of convolution results, and 'j' for the number of convolution kernels. The computations are performed as follows.

$$t_k = [t_1, t_2, \ldots \ldots, t_{n-1}, t_n\,] \tag{13}$$

$$\widehat{t_k} = \max(t_k) \tag{14}$$

$$\hat{t} = [\widehat{t_1}, \widehat{t_2}, \widehat{t_3}] \tag{15}$$

In this study, the input of the fully connected layer is the output of the convolutional layer of the output layer. Then, use the SoftMax function to find as much as possible in each segment. The pedestrian detection part of the input has the highest potential. The calculations are written as the following formula, where WD represents the weighting parameters of the mesh and BD represents the bias term.

$$d_k = w_d \hat{t} + b_d \tag{16}$$

$$o_i = \frac{\exp(d_k)}{\sum_k \exp(d_k)} \tag{17}$$

To improve the safety of autonomous vehicles, a specific pedestrian detection method using a concept-based convolutional bi-GRU model approach is presented using deep learning methods and are mixed, such as convolutional layers, attention mechanisms, and bidirectional gated recurrent units (Bi- GRU). Important attributes are extracted by the convolutional layer using weights assigned by the model to key semantic descriptions by attention mechanism Pedestrians should be classified reliably by the model due to SoftMax activation and the fully connected layer it provides for the final output. This approach shows promise for increasing the safety of autonomous vehicles by helping to avoid collisions, and for better pedestrian detection.

V. Results and Discussion

Autonomous vehicle safety was experimentally optimized using a deep learning-based pedestrian recognition system, analyzed with a concept-based convolutional bi-GRU model. The approach in this study is organized with several key features. In the first phase of data collection, a large dataset of pedestrians was obtained. This dataset, used for benchmarking, includes extensively documented videos with accurate bounding box annotations, capturing pedestrian behavior in various scenarios such as crosswalks and low-light conditions. The next step in processing the data involved min-max normalization, which scales pixel values to a standard range. The key component of the method is the concept-based convolutional bi-GRU model, a deep learning system built for pedestrian recognition. This model consists of several layers, enabled by bidirectional gated recurrent units (Bi-GRU) to capture temporal patterns and important semantics in the data. Crucial output for the attention mechanism, which highlights essential sections and enables accurate pedestrian classification, is provided by a fully connected layer and SoftMax activation. The goal of this technology is to enhance the safety of autonomous vehicles by preventing collisions and accidents through highly effective pedestrian detection, offering a promising direction for the future of driverless transportation.

The Table I provides a comprehensive summary of a neural network model's architecture and parameters. It details each layer's type, output shape, and the number of parameters involved. The model begins with a 1D convolutional layer (`Conv1D`) followed by a max pooling layer to down-sample the feature maps. It then uses a flattening layer to reshape the data for dense layers, which include two fully connected (`Dense`) layers with 128 units each and one final dense layer with a single output unit. The total number of parameters, including those in the convolutional and dense layers, is 131,905, which are all trainable, indicating the model's complexity and capacity for learning from data.

TABLE I. MODEL ARCHITECTURE AND PARAMETERS SUMMARY

Layer (type)	Output Shape	Param #
conv1d_1 (Conv1D)	(None, 28, 64)	448
max_pooling1d_1 (MaxPooling1D)	(None, 14, 64)	0
flatten_1 (Flatten)	(None, 896)	0
dense_2 (Dense)	(None, 128)	1,14,816
dense_3 (Dense)	(None, 128)	16,512
dense_4 (Dense)	(None, 1)	129
Total params	**1,31,905**	**(514 KB)**
Trainable params	**1,31,905**	**(514 KB)**
Non-trainable params	**0**	**(0.00 Byte)**

A. *Performance Evaluation*

For comparison the SVM, Random Forest Classifier and Faster R-CNN methods performance is compared with the proposed Attn-Based Convolutional Bi-GRU model. Precision, recall, F1-score, and accuracy were utilized as evaluation criteria for comparison. The model was evaluated using these parameters.

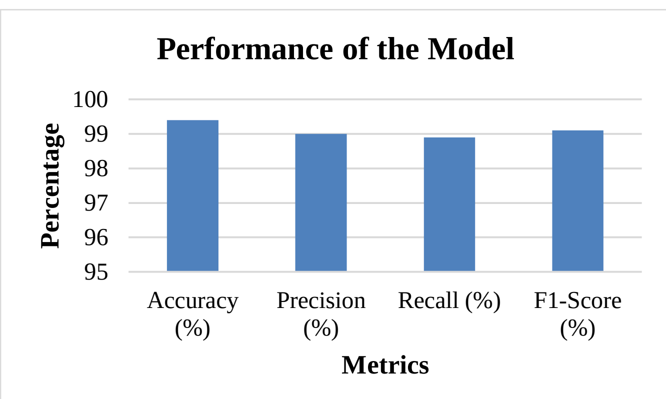

Fig. 3. Performance of the proposed attn-based convolutional Bi-GRU model.

A novel pedestrian recognition method is assessed using an Attn-Based Convolutional neural networks Bi-GRU model; the outcomes are displayed in Fig. 3 of the suggested model's performance metrics. The proposed model works remarkably well, demonstrating its high overall consistency in pedestrian recognition with an accuracy of 99.4%. It also shows exceptional accuracy at 99%, suggesting a minimal amount of false identifications, and remarkable recall at 98.9%, exhibiting its ability to reliably identify the majority of genuine pedestrians. The model's remarkable pedestrian recognition ability is validated by the F1-Score, which stands for an ideal balance between recall and precision and is notably high at 99.1% (see Table II).

TABLE II. PERFORMANCE METRICS OF ATTN.-BASED CONVOLUTIONAL BI-GRU MODEL WITH EXISTING METHODS

Method	Accuracy	Precision	Recall	F1-Score
SVM [22]	82.3(%)	88.16(%)	72.92(%)	79.82(%)
Faster R-CNN [23]	84(%)	96.6(%)	92.6(%)	94.6(%)
Random Forest [22]	90.2(%)	90(%)	80.8(%)	85.18(%)
Proposed Attn-Based Convolutional Bi-GRU Model	99.4(%)	99(%)	98.9(%)	99.1(%)

Accuracy, precision, recall, and F1-Score are only a few of the performance variables used in the Table I to compare different models from machine learning. Despite a modest recall of 72.92%, the Support Vector Machine, also known as the SVM, accurately classifies positive cases, achieving an accuracy of 82.3% and a precision of 88.16%. With an accuracy of 84%, a far better precision of 96.6%, as well as a much better recall of 92.6%, the faster R-CNN algorithm performs better than SVM, resulting in a strong F1-Score of 94.6%. With an F1-Score of 85.18%, the Random Forest algorithm performs exceptionally well, with accuracy of 90.2%, good precision of 90%, and recall of 80.8%. But the most effective model is known as the Proposed Attn-Based Convolutional Bi-GRU Model, which achieves an amazing 99.4% accuracy, 99% precision, 98.9% recall, and a remarkable 99.1% F1-Score, indicating that it can accurately classify either positive or negative instances. This shows that the suggested model performs better than any other approaches, which makes it a viable option for the specified classifying task.

The proposed attention-based convolutional bi-GRU model was compared with SVM, Faster R-CNN, and Random Forest due to their established use and performance in pedestrian detection tasks. SVM was chosen for its robustness in classification but is limited in capturing complex patterns and temporal dependencies. Faster R-CNN, a deep learning model known for high accuracy in object detection, served as a strong benchmark, though it struggles with temporal context. Random Forest, an ensemble learning method, was included as it performs well in various recognition tasks but may not handle the spatial and temporal complexities of urban environments effectively. These comparisons highlight the superior performance of the proposed model, particularly in accuracy, precision, recall, and F1-score, demonstrating its potential to significantly enhance pedestrian detection in autonomous vehicle systems.

The proposed Attn.-Based Convolutional Bi-GRU Model, Random Forest, Faster R-CNN, and Fig. 4 shows the Support Vector Machine (SVM) over all the remainder of the four machine learning models together with their accuracy, precision, recall, and overall F1-Score performance metrics. Notably, the SVM produces an F1-Score of 79.82% accompanied by a somewhat low recall of 72.92%, a strong precision of 88.16%, and a reasonable accuracy of 82.3%. Faster R-CNN, on the other hand, obtains an incredible F1-Score of 94.6% with an increase in accuracy of 84%, a significantly higher precision of 96.6%, and an astounding recall of 92.6%. With an accurate rate of 90.2%, the approach known as Random Forest achieves a great F1-Score of 85.18% by balancing recall at 80.8% and precision at 90%. On the other hand, the Proposed Attn.-Based Convolutional Bi-GRU Model is clearly the strongest performance, with an astonishing F1-Score of 99.1%, near-perfect precision of 99%, recall of 98.9%, and astounding accuracy of 99.4%. This graph demonstrates how the proposed model outperforms the other methods across the board, suggesting that it has a lot of potential for solving the present categorization challenge.

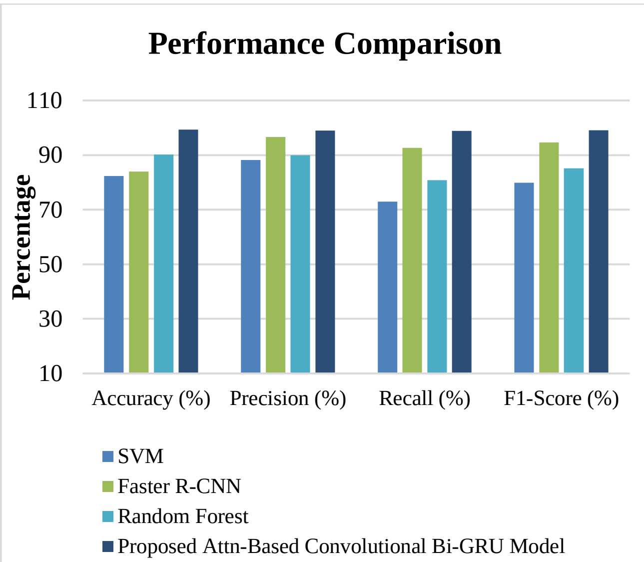

Fig. 4. Graphical depiction of the performance metrics of proposed Attn.-based convolutional Bi-GRU model with existing approaches.

Fig. 5 shows the training accuracy graph, within the framework of the proposed Attn-Based Convolutional Bi-GRU model, shows how well the model classifies pedestrians using the training dataset throughout its training iterations, or epochs. This graph shows how the model is learning and if it is becoming more accurate or if the training data may be overfitting. However, the effectiveness of the model on a separate, untested dataset that was not used for training is depicted on the testing accuracy graph. This graph sheds light on how well the model can use its knowledge of pedestrian detection to situations outside of the training set. A high testing accuracy shows the model's competence in consistently identifying pedestrians in a variety of real-world scenarios, and those is essential for improving protection for autonomous vehicle applications. A substantial training accuracy suggests that the framework successfully acquired from the training data.

Fig. 6 shows a graphical depiction of the training and testing loss for the proposed Attn-Based Convolutional Bi-GRU model, which illustrates how the model's loss function varies across the training and assessment stages.

Fig. 7 shows that the ROC curve depicts the performance of the model in binary classification tasks, especially for the proposed Attn-based Convolutional Bi-GRU model. When the threshold for classifying pedestrians is different, the sensitivity of the model changes, as detected by the ROC Curve. The model's ability to accurately identify pedestrians while avoiding false positives is reflected by a high number of true positives. In this case, the ROC curve generally exhibits a positive trend as the threshold increases from 0 to 0.7, indicating that the model is good at discriminating pedestrians from pedestrians, which is important for the safety of the application of the autonomous vehicle.

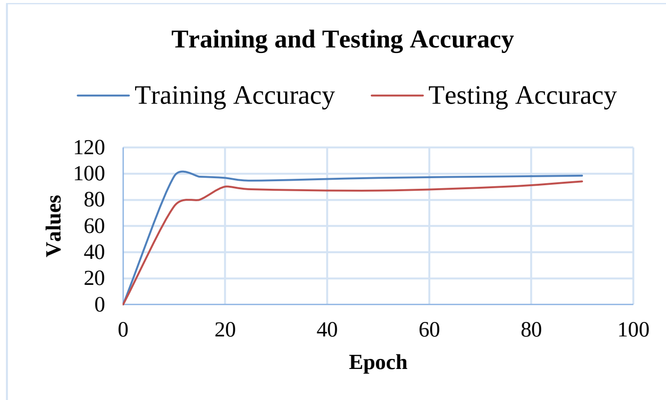

Fig. 5. Graphical depiction for training and testing accuracy of proposed Attn.-based convolutional Bi-GRU model.

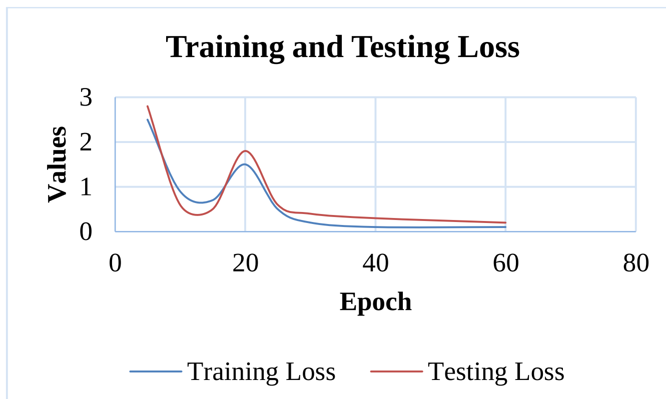

Fig. 6. Graphical depiction for training and testing loss of proposed Attn-based convolutional Bi-GRU model.

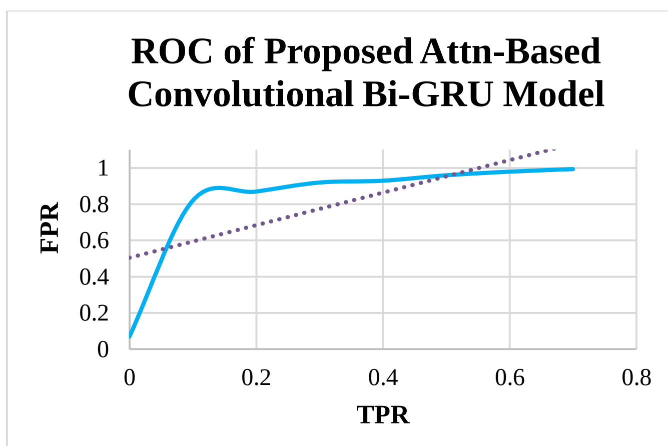

Fig. 7. ROC curve of proposed Attn-based convolutional Bi-GRU model.

B. Discussion

Improved road safety and autonomous vehicles were the two main outcomes of the study. When it comes to detecting pedestrians, the algorithm is exceptionally accurate at 99.4%, reducing the probability of accidents and increasing road safety Challenges in real-world urban environments, such as poorly lit areas, are prevented , and complex pedestrian behaviors, which are often too numerous for available algorithms to handle, and can pose risks to user safety [19]. Convolutional two-GRU architecture's attention methods allow the model to focus on relevant data, improving its performance in complex urban environments False alarm reduction with 99% accuracy ensures autonomous vehicles work more efficiently, safely and efficiently on public roads Advanced deep learning algorithms This project is a great illustration of how it can accelerate adoption and ultimately change the landscape of safe and efficient travel.

In pedestrian detection, attention mechanisms and Bidirectional Gated Recurrent Units (Bi-GRU) offer significant advantages by improving the model’s ability to focus on relevant features and capture temporal dependencies. Attention mechanisms allow the model to emphasize critical regions of the input data, such as specific areas where pedestrians are likely to appear, enhancing the model’s ability to differentiate between pedestrians and other objects or background noise. This selective focus leads to more accurate and reliable detection. Meanwhile, Bi-GRU units, which process data in both forward and backward directions, capture temporal patterns and contextual information from sequential data, such as video frames. This bidirectional approach helps the model understand the dynamic and evolving nature of pedestrian movements over time, improving its performance in complex and variable environments. Together, these techniques enhance the model's ability to accurately identify pedestrians and adapt to changing conditions, crucial for autonomous vehicle safety.

The proposed attention-based convolutional bi-GRU model outperforms existing models primarily due to its integrated approach, which leverages the strengths of both attention mechanisms and Bidirectional Gated Recurrent Units (Bi-GRU). The attention mechanism enhances the model’s ability to focus on crucial parts of the input data, effectively distinguishing pedestrians from background elements and improving detection accuracy. Bi-GRU units, by capturing temporal dependencies from both directions, provide a comprehensive understanding of dynamic pedestrian movements, which is vital in real-world scenarios where pedestrians’ positions and actions constantly evolve. However, potential weaknesses include the increased computational complexity and resource requirements associated with training and deploying such a sophisticated model. Additionally, while the model excels in controlled environments, its performance may degrade in highly unpredictable or extremely cluttered scenarios. Despite these limitations, the combined use of attention mechanisms and Bi-GRU units represents a significant advancement in pedestrian detection, offering a robust solution that enhances autonomous vehicle safety.

The research presented in this study, while demonstrating significant advancements in pedestrian detection accuracy, is subject to several limitations that warrant consideration. One primary limitation is the increased computational complexity associated with the attention-based convolutional bi-GRU model, which may restrict its real-time applicability in resource-constrained environments. Additionally, the model's robustness has primarily been tested in controlled scenarios, raising concerns about its performance in highly unpredictable or densely cluttered urban settings. Furthermore, the reliance on specific datasets for training and testing poses questions regarding the model's generalizability across different geographical regions and pedestrian behaviors. For future work, efforts should focus on optimizing the model’s computational efficiency, potentially through the development of lightweight architectures or the use of hardware accelerators like GPUs or TPUs. Expanding the evaluation to include diverse environmental conditions and pedestrian behaviors will be crucial to ensure the model's robustness and applicability. Additionally, integrating the model with other sensor modalities, such as infrared or radar, and exploring multi-sensor fusion frameworks could further enhance its effectiveness and contribute to the overall safety of autonomous vehicles.

VI. Conclusion and Future Work

The use of Python as the implementation device is a crucial factor of the model's adaptability and applicability. This paper provides a comprehensive and beneficial technique to enhancing self-sustaining vehicle protection. The first and most important step in developing a strong pedestrian dataset is facts collecting. The subsequent level of pre-processing standardizes the information the use of Min-Max normalization to get it ready for the advanced model. The Attention-Based Convolutional Bi-GRU Model, a deep studying architecture with awesome pedestrian identity abilities, is the brains in the back of this tactic. The version consists of bidirectional gated recurrent units (Bi-GRU) for temporal context seize, attention mechanisms for emphasizing semantic records, and convolutional layers for extracting spatial traits. A 99.4% accuracy price is executed, that's an outstanding improvement over preceding models via a median of approximately 17.1%. The results are outstanding. The demanding situations of identifying pedestrians in elaborate and dynamic metropolitan contexts are greatly addressed by using this accomplishment, which will ultimately lead to safer self-sustaining vehicle operation. This method offers a promising first step in the direction of the introduction of safer and extra dependable self-reliant motors, which may lessen dangers and improve the security of pedestrians and other street customers as autonomous transportation structures maintain to broaden. To increase the Attention-Based Convolutional Bi-GRU Model's performance and adaptability to changing real-world settings, extra improvements and adjustments may be investigated in future study.

References

[1] J. Zhang et al., “An infrared pedestrian detection method based on segmentation and domain adaptation learning,” Comput. Electr. Eng., vol. 99, p. 107781, 2022, doi: https://doi.org/10.1016/j.compeleceng.2022.107781.

[2] P. Jabłoński, J. Iwaniec, and W. Zabierowski, “Comparison of Pedestrian Detectors for LiDAR Sensor Trained on Custom Synthetic, Real and Mixed Datasets,” Sensors, vol. 22, no. 18, Art. no. 18, Jan. 2022, doi: 10.3390/s22187014.

[3] S. Iftikhar, Z. Zhang, M. Asim, A. Muthanna, A. Koucheryavy, and A. A. Abd El-Latif, "Deep Learning-Based Pedestrian Detection in Autonomous Vehicles: Substantial Issues and Challenges," Electronics, vol. 11, no. 21, p. 3551, 2022.

[4] J. Kolluri and R. Das, "Intelligent multimodal pedestrian detection using hybrid metaheuristic optimization with deep learning model," Image Vis. Comput., vol. 131, p. 104628, 2023, doi: https://doi.org/10.1016/j.imavis.2023.104628.

[5] W. Wang, X. Li, X. Lyu, T. Zeng, J. Chen, and S. Chen, "Multi-Attribute NMS: An Enhanced Non-Maximum Suppression Algorithm for Pedestrian Detection in Crowded Scenes," Appl. Sci., vol. 13, no. 14, Art. no. 14, Jan. 2023, doi: 10.3390/app13148073.

[6] "NMS-Loss: Learning with Non-Maximum Suppression for Crowded Pedestrian Detection | Proceedings of the 2021 International Conference on Multimedia Retrieval," ACM Conferences. Accessed: Oct. 25, 2023. [Online]. Available: https://dl.acm.org/doi/10.1145/3460426.3463588

[7] H. Kulhandjian, J. Barron, M. Tamiyasu, M. Thompson, and M. Kulhandjian, "Pedestrian Detection and Avoidance at Night Using Multiple Sensors and Machine Learning," in 2023 International Conference on Computing, Networking and Communications (ICNC), 2023, pp. 165–169. doi: 10.1109/ICNC57223.2023.10074081.

[8] D. Yang, H. Zhang, E. Yurtsever, K. A. Redmill, and Ü. Özgüner, "Predicting Pedestrian Crossing Intention With Feature Fusion and Spatio-Temporal Attention," IEEE Trans. Intell. Veh., vol. 7, no. 2, pp. 221–230, 2022, doi: 10.1109/TIV.2022.3162719.

[9] U. Gawande, K. Hajari, and Y. Golhar, "Real-Time Deep Learning Approach for Pedestrian Detection and Suspicious Activity Recognition," Procedia Comput. Sci., vol. 218, pp. 2438–2447, 2023, doi: https://doi.org/10.1016/j.procs.2023.01.219.

[10] H. Xu, M. Guo, N. Nedjah, J. Zhang, and P. Li, "Vehicle and Pedestrian Detection Algorithm Based on Lightweight YOLOv3-Promote and Semi-Precision Acceleration," IEEE Trans. Intell. Transp. Syst., vol. 23, no. 10, pp. 19760–19771, 2022, doi: 10.1109/TITS.2021.3137253.

[11] A. M. Obeso, J. Benois-Pineau, M. S. G. Vázquez, and A. Á. R. Acosta, "Visual vs internal attention mechanisms in deep neural networks for image classification and object detection," Pattern Recognit., vol. 123, p. 108411, 2022, doi: https://doi.org/10.1016/j.patcog.2021.108411.

[12] H. Lv, H. Yan, K. Liu, Z. Zhou, and J. Jing, "YOLOv5-AC: Attention Mechanism-Based Lightweight YOLOv5 for Track Pedestrian Detection," Sensors, vol. 22, no. 15, Art. no. 15, Jan. 2022, doi: 10.3390/s22155903.

[13] X. Zhou and L. Zhang, "SA-FPN: An effective feature pyramid network for crowded human detection," Appl. Intell., vol. 52, no. 11, pp. 12556–12568, Sep. 2022, doi: 10.1007/s10489-021-03121-8.

[14] X. Chen, Y. Jia, X. Tong, and Z. Li, "Research on Pedestrian Detection and DeepSort Tracking in Front of Intelligent Vehicle Based on Deep Learning," Sustainability, vol. 14, no. 15, Art. no. 15, Jan. 2022, doi: 10.3390/su14159281.

[15] Y. Lu, W. Wang, X. Hu, P. Xu, S. Zhou, and M. Cai, "Vehicle trajectory prediction in connected environments via heterogeneous context-aware graph convolutional networks," IEEE Trans. Intell. Transp. Syst., vol. 24, pp. 8452–8464, Aug. 2023, doi: 10.1109/TITS.2022.3173944.

[16] A. Khan, M. M. Fouda, D.-T. Do, A. Almaleh, and A. U. Rahman, "Short-Term Traffic Prediction Using Deep Learning Long Short-Term Memory: Taxonomy, Applications, Challenges, and Future Trends," IEEE Access, vol. 11, pp. 94371–94391, 2023, doi: 10.1109/ACCESS.2023.3309601.

[17] I. V. Pustokhina, D. A. Pustokhin, T. Vaiyapuri, D. Gupta, S. Kumar, and K. Shankar, "An automated deep learning based anomaly detection in pedestrian walkways for vulnerable road users safety," Saf. Sci., vol. 142, p. 105356, Oct. 2021, doi: 10.1016/j.ssci.2021.105356.

[18] M. Islam, M. Rahman, M. Chowdhury, G. Comert, E. D. Sood, and A. Apon, "Vision-Based Personal Safety Messages (PSMs) Generation for Connected Vehicles," IEEE Trans. Veh. Technol., vol. 69, no. 9, pp. 9402–9416, 2020, doi: 10.1109/TVT.2020.2982189.

[19] T. Aminah, "FRAMEWORK FOR PEDESTRIAN WALKING BEHAVIOUR RECOGNITION TO MINIMIZE ROAD ACCIDENT".

[20] "Pedestrian Dataset." Accessed: Oct. 25, 2023. [Online]. Available: https://www.kaggle.com/datasets/smeschke/pedestrian-dataset

[21] J. Thaker, N. Jadav, S. Tanwar, P. Bhattacharya, and H. Shahinzadeh, Ensemble Learning-based Intrusion Detection System for Autonomous Vehicle. 2022. doi: 10.1109/SCIoT56583.2022.9953697.

[22] X. Peng and J. Shan, "Detection and Tracking of Pedestrians Using Doppler LiDAR," Remote Sens., vol. 13, no. 15, p. 2952, Jul. 2021, doi: 10.3390/rs13152952.

[23] C. Amisse, M. E. Jijón-Palma, and J. A. S. Centeno, "FINE-TUNING DEEP LEARNING MODELS FOR PEDESTRIAN DETECTION," Bol. Ciênc. Geodésicas, vol. 27, no. 2, p. e2021013, 2021, doi: 10.1590/s1982-21702021000200013.