

## Article

# Assistive Self-Driving Car Networks to Provide Safe Road Ecosystems for Disabled Road Users

Juan Guerrero-Ibañez <sup>1</sup>, Juan Contreras-Castillo <sup>1</sup>, Ismael Amezcua-Valdovinos <sup>1</sup>  
and Angelica Reyes-Muñoz <sup>2,\*</sup>

<sup>1</sup> Faculty of Telematics, University of Colima, Colima 28040, Mexico; antonio\_guerrero@ucol.mx (J.G.-I.); juancont@ucol.mx (J.C.-C.); ismaelamezcua@ucol.mx (I.A.-V.)

<sup>2</sup> Computer Architecture Department, Polytechnic University of Catalonia, 08034 Barcelona, Spain

\* Correspondence: angelica.reyes@upc.edu; Tel.: +52-3121144354

**Abstract:** Disabled pedestrians are among the most vulnerable groups in road traffic. Using technology to assist this vulnerable group could be instrumental in reducing the mobility challenges they face daily. On the one hand, the automotive industry is focusing its efforts on car automation. On the other hand, in recent years, assistive technology has been promoted as a tool for consolidating the functional independence of people with disabilities. However, the success of these technologies depends on how well they help self-driving cars interact with disabled pedestrians. This paper proposes an architecture to facilitate interaction between disabled pedestrians and self-driving cars based on deep learning and 802.11p wireless technology. Through the application of assistive technology, we can locate the pedestrian with a disability within the road traffic ecosystem, and we define a set of functionalities for the identification of hand gestures of people with disabilities. These functions enable pedestrians with disabilities to express their intentions, improving their confidence and safety level in tasks within the road ecosystem, such as crossing the street.

**Keywords:** deep learning; neural networks; pedestrians; pedestrians with disabilities; recurrent neural networks; road users; self-driving cars



**Citation:** Guerrero-Ibañez, J.; Contreras-Castillo, J.; Amezcua-Valdovinos, I.; Reyes-Muñoz, A. Assistive Self-Driving Car Networks to Provide Safe Road Ecosystems for Disabled Road Users. *Machines* **2023**, *11*, 967. <https://doi.org/10.3390/machines11100967>

Academic Editor: Yahui Liu

Received: 4 September 2023

Revised: 6 October 2023

Accepted: 12 October 2023

Published: 17 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

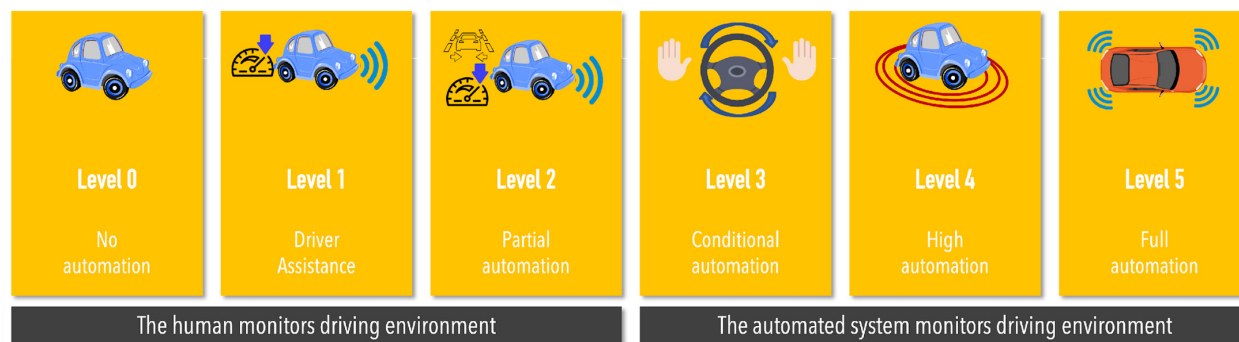
According to the World Health Organization (WHO), road traffic injuries are a severe public health problem worldwide. WHO data show that approximately 1.35 million people die in road traffic crashes, and 20–50 million suffer some form of non-fatal injury [1]. More than 90% of road deaths occur in third-world or developing countries, and more than half involve vulnerable road users.

The term vulnerable road users (VRUs) refers to unprotected users within the road environment, including cyclists, pedestrians, motorcyclists, and people using transport devices to move their bodies [2]. The WHO explicitly states that older adults, children, and people with disabilities are the road users with the highest risk of being involved in a road accident [3].

Data released by the WHO in 2022 established that 1.3 billion people worldwide have some significant disability. Furthermore, 45 million people worldwide are blind or visually impaired, and 45 million are deaf [4]. In addition, around 87% of people with disabilities live in developing countries.

Factors such as crossing decisions and walking speed are pedestrians' most studied risk factors [3]. The studies show that it is the actions and decisions of the disabled pedestrian rather than environmental or social characteristics that pose the most significant risk [4]. Studying these factors has contributed to understanding how people move and has helped to define the way forward for designing more inclusive and functionally valuable environments.

The automotive industry focuses on creating vehicles with systems that provide greater intelligence to have a fully autonomous car as the final product. Organizations such as the National Highway Traffic Safety Administration (NHTSA) and the Society of Automotive Engineers (SAE) define five levels of vehicle automation (Figure 1), ranging from zero, where the driver has complete control of driving tasks, to five, where the vehicle uses automation systems to perform driving tasks without human intervention [5].



**Figure 1.** Automation levels for self-driving cars.

The goals of self-driving cars are to reduce traffic accidents, congestion, and pollution levels substantially and to reduce transportation costs. Although there are still technical challenges to be solved, self-driving cars' real success or failure is related to social acceptance and integration into the road traffic ecosystem [6,7]. According to Kaur and Rampersad, reliability and trust are two of the main concerns of people when using self-driving cars [8]. Reliability refers to the ability of the car to overcome unexpected situations, while trust relates to safeguarding a car's safety. Replacing human drivers with autonomous control systems could also create a severe social interaction problem: constant interaction between pedestrians and self-driving cars is required to ensure smooth and safe traffic flow, especially at pedestrian crossings. We can identify several interaction tasks between the self-driving car and disabled pedestrians, such as pedestrian identification, pedestrian movement prediction, pedestrian behavior analysis, car-pedestrian communication, and car-pedestrian feedback. There are also tasks related to car-to-car communication to inform about pedestrian and car agreement dissemination using broadcasting mechanisms and edge computing.

A self-driving car must be able to handle obstacles to navigate safely through the vial traffic zone. Therefore, VRU detection is a major task to be performed by self-driving cars to significantly reduce the probability of accidents within the vehicular traffic ecosystem [9]. The authors in [10] classify VRUs into six categories: (i) distracted road users, (ii) road users inside the vehicle, (iii) special road users, (iv) users of transport devices, (v) animals, and (vi) road users with disabilities (Table 1).

The latter category represents a particular type of pedestrian moving within the road environment who has a disability (such as blind people, deaf people, and people in wheelchairs, among others). This category is referred to as Disabled VRU (D-VRU). We believe self-driving cars will open many possibilities for people with disabilities (independence, with the improvement in quality of life that this represents) by explicitly interacting with them to enhance their mobility, independence, and confidence when crossing roads.

A key element for pedestrians with disabilities to feel safe within the vehicular traffic ecosystem will depend on the ability to communicate with vehicles and how well both parties understand such interaction. Hence, a comprehensive communication framework is needed to define the interaction between pedestrians and self-driving vehicles and between self-driving vehicles. Such a framework should include modules for detecting D-VRU, communication mechanisms for the interaction between D-VRUs and self-driving vehicles, and networking communication between vehicles using wireless technology.

**Table 1.** Categories of vulnerable road users.

Category	Description
Distracted road users	A type of pedestrian walking on the road but is distracted by an additional activity (such as using a cell phone, conversing with another person, or thinking about something else).
Road users inside the vehicle	Category refers to the occupants of an automated or a conventional vehicle.
Special road users	Pedestrians with very low walking speeds, such as older people and children, are included in this category.
Users of transport devices	Users who use transport devices such as skates, scooters, roller skis or skates, and kick sleds or kick sleds equipped with wheels.
Animals	Refer to animals within the driving zone, including dogs, horses, and cats.
Road users with disabilities	People who move within the driving zone and have a disability, such as blindness, deafness, or wheelchair users.

Current efforts have focused on applying technology to assist these people, giving rise to Assistive Technology (AT). According to the Assistive Technology Industry Association (ATiA), AT refers to any equipment, item, software program, or product that enhances the capabilities and functionalities of people with a disability [11]. The WHO states that AT focuses on enabling people to live independently, healthily, and productively and to integrate into society more naturally. AT can be applied to different areas such as health, education, or sports.

This paper proposes an architecture that integrates assistive technologies into the road traffic ecosystem. The goal is to increase the safety, security, and confidence of D-VRUs when crossing pedestrian crosswalks. The vision towards integrating AT with autonomous cars leads us to define the Assistive Self-Driving Car (ASC) concept.

This paper explores integrating machine learning and wireless technologies to detect disabled pedestrians. To detect and identify D-VRUs, neural network (NN) models and wearable devices based on wireless technologies are used to establish direct communication between D-VRUs and self-driving vehicles.

We further extend the interaction between D-VRUs and self-driving cars by designing and testing an NN-based model for identifying different hand gestures to express the intentions of actions from the D-VRU to the self-driving car. This vital step has not been thoroughly studied. The primary purpose is to achieve a higher level of agreement between D-VRUs and self-driving cars, giving confidence to pedestrians in their actions and therefore improving self-driving cars' acceptance level in our modern society.

The main contributions of this paper are as follows:

- The definition of the ASC framework is composed of modules for D-VRU recognition, hand gesture interaction and its corresponding feedback, and a network architecture for self-driving car interaction.
- The definition, challenges, and requirements for an ASC to improve its interaction with D-VRUs.
- A mechanism based on NN and a wearable device that accurately identifies pedestrians with disabilities and their specific type of impairment to enable interaction with the pedestrian through an appropriate interface for feedback.
- An algorithm based on recurrent NN that identifies hand gestures as a means for the interaction between the D-VRU and the self-driving assistive car.

This paper is structured as follows: Section 2 overviews the assistive self-driving car concept. Section 3 includes a state-of-the-art literature review of pedestrian–self-driving car interactions. Section 4 describes the interaction process between assistive self-driving cars

and D-VRUs and some solutions to perform the interaction process. Finally, we present our conclusions and future work.

## 2. State of the Art

Pedestrian–vehicle interaction is a topic that has been under study for the past years in the context of intelligent vehicle environments. As a result, many surveys have been published on pedestrian detection within vehicle environments, using different sensors, cameras, and artificial intelligence algorithms [12–16].

This section discusses how pedestrian–car interactions occur in the current road-driving ecosystem. We then analyze how this interaction will change as self-driving cars integrate into the road-driving ecosystem. Finally, we describe some requirements that must be met for the interaction between D-VRUs and ASCs.

### 2.1. Current Pedestrian–Car Communication

In a typical road traffic environment, the communication between pedestrians and car drivers is based on informal communication including making eye contact, facial expressions, hand gestures, and even specific car horn sounds. Such everyday communications allow pedestrians and drivers to infer, based on the experiences of both parties, the set of required actions to follow to avoid accidents [17]. This non-verbal communication indicates the vehicle's movements (stop and give way to the pedestrian, continue driving, etc.) and the pedestrian's actions (stop, cross the street, etc.) to avoid a possible eventuality. Generally, in countries where road-crossing regulations are permissive, the pedestrian attempts to make eye contact with the driver to identify that the driver has already detected them. With this informal communication, the pedestrian can cross the street more safely and securely, considering that the probability of suffering an eventuality is almost zero.

However, the same informal communication mode does not work with people with disabilities. A blind person, for example, would not be able to make eye contact with the vehicle's driver. If the driver honks the horn, the blind pedestrian will not know whether the sound is meant for them to cross the road or wait for the car to pass. It is necessary to develop friendly mechanisms for disabled pedestrians to help them safely move through the road ecosystem.

The road-driving ecosystem is evolving rapidly; consequently, automated vehicles are starting to take to the roads. In this new road-driving environment, traditional (non-automated) cars, semi-automated cars, and VRUs will be circulating and will impact the current mode of interaction. As the level of automation advances, the interaction process will change. While some researchers believe that informal communication will disappear as self-driving cars are integrated into the road ecosystem, others further acknowledge that the disappearance of informal communication will increase distrust of self-driving cars [18]. Informal communication should not be outdated but adapted to the new conditions of vehicle intelligence. More robust, safer, and more accessible forms of communication need to be developed to enable two-way communication between self-driving cars and D-VRUs. At the same time, these new forms of communication will need to be inclusive so that every pedestrian can move safely in the road environment.

### 2.2. Pedestrian–Self-Driving Car Communication

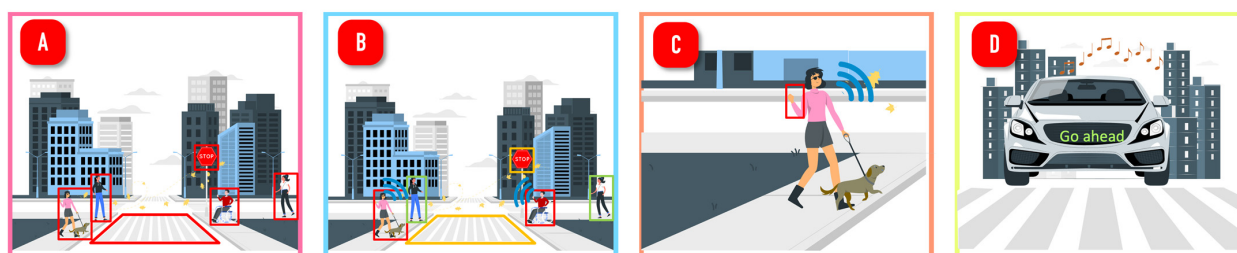
The ability to interact with pedestrians is one factor determining the success of self-driving cars. However, research focused on creating mechanisms to facilitate the interaction between self-driving cars and pedestrians is still in its early stages. Poor communication between self-driving cars and pedestrians can have fatal consequences. For example, suppose the vehicle knows what the pedestrians intend to do. In that case, the car can react and avoid colliding with the pedestrian and potentially causing severe damage to the pedestrian, including loss of life. The absence of a driver in an automated vehicle makes pedestrians wary of not knowing what the self-driving car will do [19,20].

Most interaction efforts have focused on sending messages from self-driving cars to pedestrians. These messages inform the pedestrian of the action to be taken. The interaction technologies developed have been classified into three categories: (i) visual, (ii) acoustic, and (iii) anthropomorphic interfaces [21]. Some works have proposed the development of visual interfaces, such as LED strips [22–24], screens displaying text or icons [25–27], and holograms or projections. All these interfaces are an attempt to communicate the action to the pedestrian. More research extends the functionality of visual interfaces to include acoustic devices that help convey the message to those with visual impairments [28,29]. Anthropomorphic interfaces attempt to inform pedestrians of actions to be taken by using human characteristics (simulating eyes to represent eye contact with the pedestrian) [30–32].

However, more progress has yet to be made in proposing a way for pedestrians to communicate with self-driving cars. One of the critical issues will be the car's understanding of pedestrians' intentions. To successfully communicate from pedestrian to self-driving car, the self-driving car must be trained to correctly understand the pedestrian's messages. Several papers consider the posture [33–35], the position of the head [36,37], and the trajectory and kinematics of the pedestrian [38,39]. The main problem with these methods is that they rely on implicit communication. The self-driving car makes inferences about the pedestrian's intentions based on the pedestrian's actions in this type of communication. This work focuses on explicit communication among pedestrians and self-driving cars using simple hardware devices that use IEEE 802.11p radio. The self-driving car then broadcasts pedestrian alerts and information to nearby self-driving cars in a privacy-aware manner while supporting hand gestures to express the pedestrian's intention to use the intersection.

### 2.3. Requirements for Pedestrian–Self-Driving Car Communication

For the two-way communication mechanism of communication between self-driving cars and pedestrians to succeed, several essential requirements must be satisfied (such as being a simple, adaptable, safe, and secure mechanism, among others). The authors in [40] mention that for the self-driving car to communicate with the pedestrian, there are two essential steps that the car needs to perform. The first step focuses on detecting and identifying pedestrians circulating within the travel environment of the self-driving car. The second stage describes that autonomous cars will have to notify pedestrians what action the car will take. Based on the stages defined by the authors in [40], additional ones will be necessary for the specific case of the interaction process between the ASC and D-VRU, which we present in Figure 2.



**Figure 2.** Interaction process between an ASC and D-VRUs: (A) object detection, (B) zebra crossing, VRU, and D-VRU classification, (C) hand gesture and intention detection, and (D) two-way communication.

### 2.4. Pedestrian Detection

As part of the autonomous navigation functions of ASCs, detecting all objects or obstacles within the car's travel environment is essential. Object detection will allow the self-driving car to calculate its trajectory or establish the action to be taken to avoid a situation that may endanger or put the people around it at risk. The main task of the ASC is to detect objects, calculate their distance, identify their position, and keep a comprehensive record of all things (static and moving). However, object detection and classification are



challenging, as it demands a high level of processing for a large amount of data. Therefore, machine learning and deep learning technologies are used for such tasks.

In the context of ASCs, there are two main mechanisms to detect and classify objects: feature extraction and object classification. Feature extraction is performing dimensionality reduction of data so that raw input data is divided and reduced to create smaller groups that are easier to process. This process optimizes detecting features such as shapes, angles, or movement in digital images or videos. Some of the current methods used in feature extraction include the Histogram of Oriented Gradients [41–43], Local Binary Pattern [44–46], Deformable Part Model [47–49], and Aggregate Channel Feature (ACF) [50–52].

Different types of deep learning (DL)-based algorithms detect and classify VRUs. DL models, specifically Convolutional Neural Networks (CNNs), can obtain classification rates with up to 100% accuracy [53]. Region-based algorithms define candidate areas or regions where the object is expected to be and then apply a CNN model to obtain a detection box. Within such network models, Regional Convolutional Neural Networks (R-CNNs) [54,55], Fast and Faster R-CNNs [56,57], and Mask Region-based CNNs [58,59] are being used.

Additionally, regression-based algorithms do not employ the concept of regions. Still, the input image is only processed once, and the category and the target border can be regressed on multiple image positions. Algorithms within this category are You Only Look Once (YOLO) [60], Single-Shot MultiBox Detector (SSD) [61], and RetinaNet [62]. YOLO is one of the most widely used models for real-time object detection, allowing faster detection than other models with video up to 30 frames per second at the cost of sacrificing detection accuracy.

### 3. Materials and Methods

This section details our assistive interaction mechanism between self-driving cars and D-VRUs.

#### 3.1. Assistive Self-Driving Cars

Reducing the number of road accidents is one of the main goals of assistive self-driving cars. Achieving this objective requires that autonomous vehicles implement a set of capabilities that enable them to detect and identify the intentions of pedestrians.

Self-driving cars should have a mechanism for expressing their intentions to create a safer pedestrian environment. Modern cars are equipped with assistance systems known as ADASs (Advanced Driver Assistance Systems), which have the ability to recognize pedestrians in general, without making specific discriminations as to whether they are disabled or not. Therefore, current systems cannot provide an assistive environment that facilitates interaction between the car and a disabled pedestrian. Considering the preceding, we propose integrating assistive technologies into autonomous cars.

The self-driving car defines a technical and functional architecture to structure integration between hardware and software within a car and the processing block of all the tasks such a vehicle must perform to work efficiently. The autonomous car control tasks are classified into (i) perception, (ii) planning and decision, and (iii) motion and control.

The perception block creates a representative model of the travel environment based on all the data collected from the different sensors included in the self-driving car (LiDAR, cameras, ultrasonic, and radar), static data (digital maps, rulers, and routes), and environmental conditions (weather conditions and real-time location).

The planning and decision block oversees generating a real-time safe and efficient navigation action plan. To create the plan, the self-driving car combines the representative model developed in the perception block with environmental data such as destination points, digital maps, and traffic rules.

The vehicle motion and control block must execute the trajectory generated by the previous blocks using motion commands that control the self-driving car actuators.

There are three requirements that the self-driving assisted car must meet to be able to communicate correctly with disabled pedestrians: (i) identify the different types of dis-

abilities, (ii) locate the global position of the disabled person, and (iii) define an interaction mechanism for this situation.

The first requirement is extending the pedestrian detection capabilities by identifying pedestrians to better respond to each pedestrian's situation. Identifying the specific type of disability will allow the autonomous car to select the most appropriate medium or interface to communicate with the pedestrian. The second requirement is focused on locating the global position of the pedestrian with a disability, which is essential for the self-driving car to respond to the pedestrian's needs in the best possible way. Lastly, the third requirement defines a series of interaction mechanisms among self-driving cars, pedestrians with a disability, and other self-driving cars. The idea is to define a framework in which D-VRUs and self-driving cars interact to determine a set of actions or steps each participant is to follow.

### 3.2. Scenario

In this work, a specific scenario was defined to illustrate the operation of the proposed assistive system. The example scenario is shown in Figure 3.



**Figure 3.** Representation of the usage scenario for the proposed assistive system.

The scenario analyzes the process of a D-VRU trying to cross the road. In this scenario, the D-VRU, communicating through their handheld device, sends a series of messages to enable the self-driving car to detect their presence within its driving environment. The self-driving car identifies the type of disability of the pedestrian. Next, it identifies the intention of the pedestrian based on the analysis of the hand gestures made by the pedestrian with a disability. Finally, it provides different adaptive responses depending on the disability. Then, based on the signals emitted by the pedestrian's wearable device, the self-driving car can make a targeted response to improve the assistance provided to the pedestrian.

### 3.3. Proposed Architecture

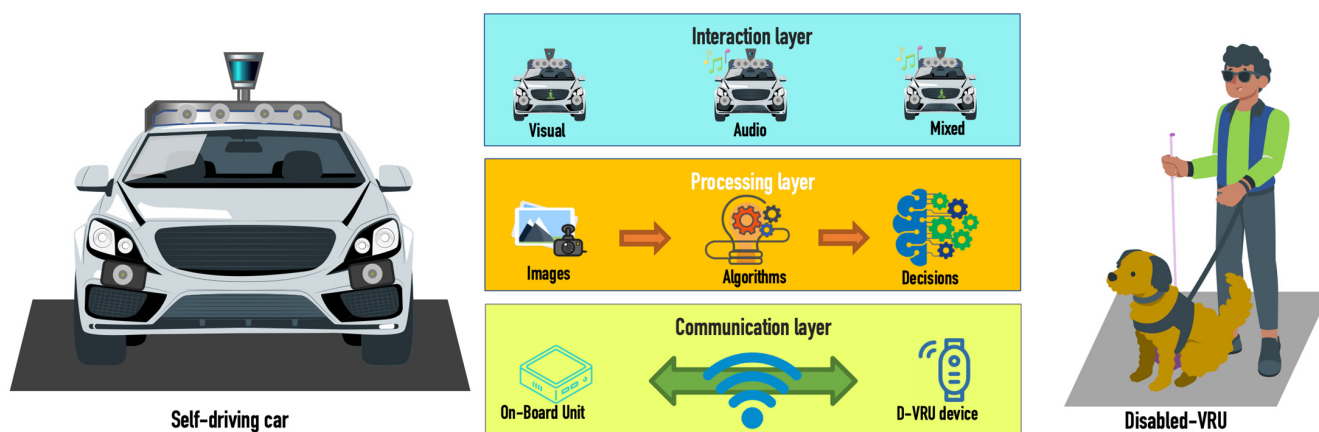
Our proposal incorporates assistive technology into self-driving vehicles and D-VRUs, creating the assistive self-driving car concept.

The D-VRU's information is very sensitive and must be protected from theft or undesired sharing. For this reason, we propose an architecture based on federated learning [63] where each user can decide whether to share information with the central node, which,

in this case, is the autonomous vehicle. The data is anonymously transmitted; at the destination, even the transmitter's source is anonymized. In this decentralized approach, the model is trained locally using the raw data on edge devices (D-VRU smartphones or handheld devices). These devices have significantly higher latency, lower-throughput connections, and are only intermittently available for training.

The key idea is that the data and model are both maintained locally at the mobile device or the handheld of the D-VRU, and only the learning configuration parameters are shared between the local nodes and the vehicles to generate a global model shared by all nodes.

The architecture consists of three essential layers, as shown in Figure 4.



**Figure 4.** Representation of the layers for the proposed architecture.

### 3.3.1. Communication Layer

The communication layer regulates the exchange of messages between the self-driving car and the D-VRU. When there is a D-VRU on the road, their mobile devices periodically share location data via broadcast with the rest of the users, the vehicles, and the road infrastructures. With each location, the autonomous vehicle receives the D-VRU control point when the global model is activated.

If the D-VRU cannot or prefers not to use a mobile device, then the pedestrian with a disability can wear a handheld device that incorporates 802.11p technology. The self-driving car will be equipped with 802.11p communications technology. This proposal defines procedures and messages for identifying VRUs using an 802.11p wireless device (smartphone or handheld device). It also establishes a set of messages that will be exchanged between all the elements that make up the proposed architecture, allowing the location and identification of the type of disability of the D-VRU (see Section 3.3).

### 3.3.2. Processing Layer

The processing layer provides intelligence to the architecture proposal, treating each D-VRU's model as a whole entity. This layer is responsible for identifying and interpreting the D-VRU's intentions to respond with the interface best suited to their disabilities. Using federated learning, we generate effective personalized model parameters for each D-VRU. Afterwards, the processing layer evaluates the information from different D-VRUs in similar contexts to achieve customized model aggregation.

The processing layer takes, as input, images of the driving environment captured by a camera located in the front of the vehicle, processes the images in real-time to detect the D-VRU's hand gestures, interprets the D-VRU's intention, and reports the obtained result to the interaction layer, which includes (1) the type of disability of the D-VRU, (2) the D-VRU's intention, and (3) the D-VRU's location.



### 3.3.3. Interaction Layer

All road users negotiate with other road users to achieve their social-traffic goals. In the traffic ecosystem we analyze, where autonomous vehicles will become part of our daily lives, the D-VRUs require special attention through new communications protocols. The interaction layer works with any pair of agents occupying a region of interest in the traffic system simultaneously. It disappears once any agents move outside of the area of interest. The interaction layer has a mechanism for filtering the appropriate message between both agents based on the results obtained by the previous layers. Having the core of each model at the D-VRU device makes for fewer iterations of high-quality updates to produce a good model, which means training the models can use much less communication.

This layer can choose visual communication if it detects someone who is deaf or hard of hearing, audio interaction if the interaction layer detects someone who is visually impaired, or mixed interaction if it detects several people with different types of disabilities. Storing D-VRU data in the cloud is unnecessary, and we also preserve privacy using aggregation protocols of federated learning [64].

An important aspect to consider for ensuring the safety of the system is redundancy. According to our proposed architecture, the different layers in which our system is based interact with each other to provide feedback. In this sense, the decisions made at one layer are based on the contextual information provided by other layers. For example, the processing layer employs different algorithms for the detection and identification of D-VRUs and hand gestures for interacting with the vehicle once the pedestrian is detected. This layer, however, indirectly depends on the communication and interaction layers because the vehicle receives notifications through broadcast messages sent using IEEE 802.11p technology. Furthermore, it requires the interaction with one or more interfaces equipped on the vehicle used to interact with the pedestrian. If for some reason the D-VRU wireless device stops sending broadcasts, the redundancy in the feedback of the systems allows the system to continue working by using only the detection and identification layers. Likewise, if for some reason the video-capturing devices stop working properly in the vehicle, it is possible to continue the interaction with the pedestrian through the IEEE 802.11p technology to come to an agreement.

It is also possible to include processing service redundancy. Along with the communication, detection, and interaction processes in the vehicle, the information gathered by all sensors can be sent and processed in edge servers. By employing edge computing, all decisions processed in the vehicle can be corroborated by the output of the same algorithms running on a cloud server with the same contextual information. This helps in enhancing the overall safety of the system.

### 3.3.4. Validation System

The purpose of this article is the definition of an architecture where pedestrians with disabilities can be recognized, can interact with the intelligent vehicle, and can make decisions jointly to improve the safety of both. To validate the proposed architecture, different unit tests were performed, which are described below.

A first validation test was performed at the communication layer. This test consisted of demonstrating the viability of the D-VRU device where a disabled pedestrian crosses a two-way avenue. The test was developed using the Veins simulator [65]. The D-VRU device periodically sends broadcast messages to nearby vehicles to identify itself as a disabled passerby using IEEE 802.11p technology. Specifically, inside Veins a pedestrian was generated that moves at a speed of 2 km per hour. Also, eight vehicles are generated sequentially at different time intervals with the purpose of defining a platoon lead vehicle which first receives the message from the D-VRU, calculates the distance from the pedestrian, and begins to slow down to avoid a collision. In addition, the vehicle also begins to send periodic messages to the other vehicles, which upon receiving the messages consider the actions that the platoon lead vehicle has carried out to perform their own actions, such as reducing speed depending on the distance from the vehicle, the pedestrian, and the

platoon lead vehicle. The results were published in [66] where the test scenario is described in greater detail.

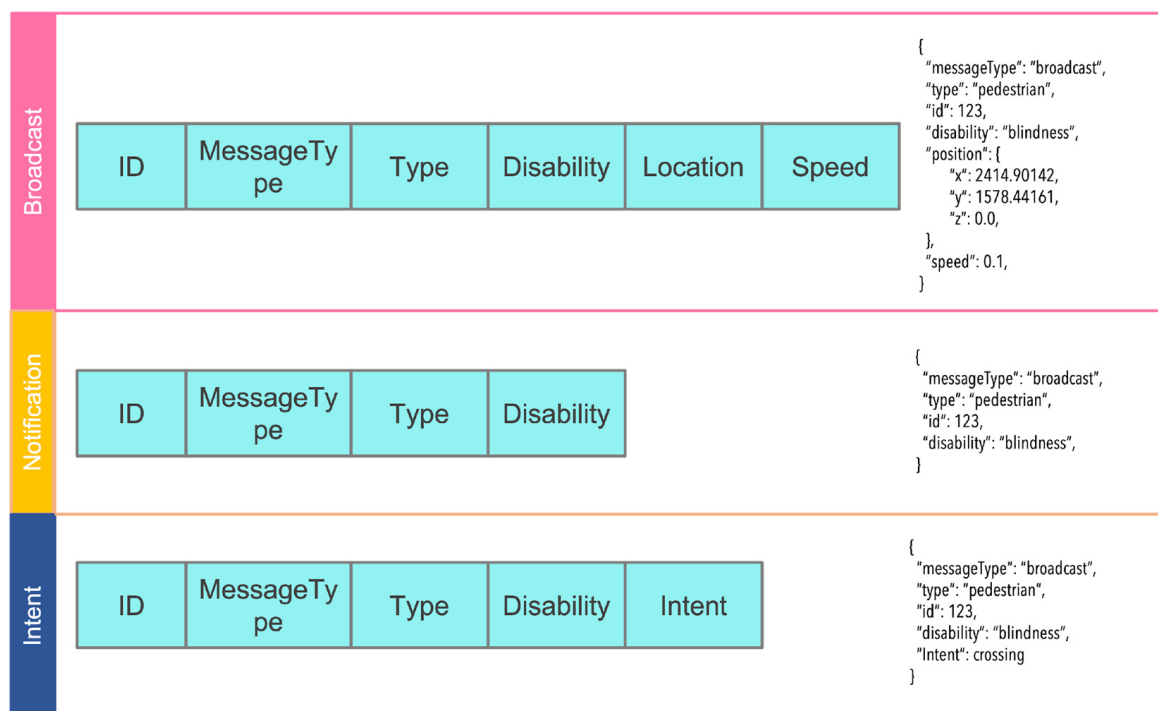
A second and third validation test were performed at the processing layer. The second test consisted of the use of machine learning techniques for the visual detection of disabled pedestrians through video cameras and the detection and identification as vulnerable depending on their aids such as guide dogs or canes. Section 3.6 shows this process in detail. The third test was carried out on the interaction between the D-VRU and the vehicle through hand gestures using deep learning for proper identification where the accuracy of identifying the pedestrian's intention is demonstrated. Section 3.7 shows this process in detail.

### 3.4. D-VRU Device

A pedestrian with a disability will use his smartphone or a device equipped with 802.11p technology. The device will informatively process physiological and body movement data without any personal information and share the information with the self-driving car, allowing the car to fully identify the type of disability, the location, the speed, and the direction of travel, among other things.

### Message Structure

Three types of messages are defined for 802.11p communication: broadcast, notification, and intent, as shown in Figure 5.



**Figure 5.** Structure of the different message types proposed for 802.11p-based communication.

The “broadcast” message is sent by the D-VRU periodically, every second, so that ASCs driving near the D-VRU’s environment can identify the disabled pedestrian at an early stage. The structure of this message type consists of several fields, described below. First, an “ID” field represents the sender’s ID. The “MessageType” field indicates the type of message that is being sent. In this case, its value is set to “broadcast”. The “Type” field indicates the type of node that is sending the message. This type of message’s value is set to “D-VRU”. The “Disability” field shows the pedestrian’s disability type, including blindness, deafness, etc. The “Location” field contains the D-VRU’s location coordinates. Finally, the “Speed” field reports the pedestrian’s speed. Speed and location can be calculated using

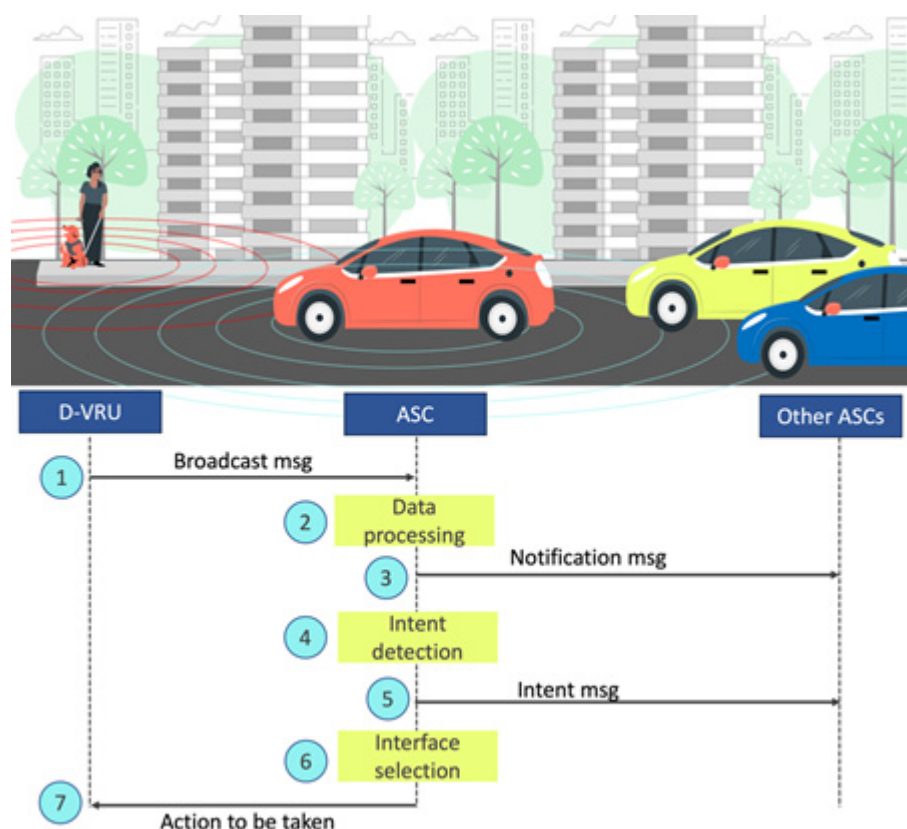
applications used to measure driving performance, such as a GPS speedometer/mileage app. To measure GPS accuracy, the USA government provides the GPS signal in space with a global average user range rate error (URRE) of  $\leq 0.006$  m/sec over any 3 s interval, with 95% probability [67].

The ASC sends the notification message to other cars near its environment to inform them of a D-VRU in the environment. This message has an “ID” field representing the sender’s ID. The “MessageType” field specifies the message type. In this case, its value is “notification”. The “Type” field indicates the kind of sender of the message. In this case, its value is “vehicle”. Finally, the type of disability of the D-VRU detected in the environment is indicated in the “Disability” field.

Once the car has identified the intention of the D-VRU, it sends an intent message to indicate to the other cars what action the D-VRU will take. The other cars can then take appropriate precautions to keep the mobility environment safe. The structure of the message is like the “notification” message. However, a field is added to indicate the intention of the D-VRU.

### 3.5. D-VRU/ASC Interaction Process

Figure 6 shows the interaction process between the self-driving car and the disabled pedestrian. The interaction procedure generally works like this:



**Figure 6.** Representation of the interaction process between the D-VRU and the ASC.

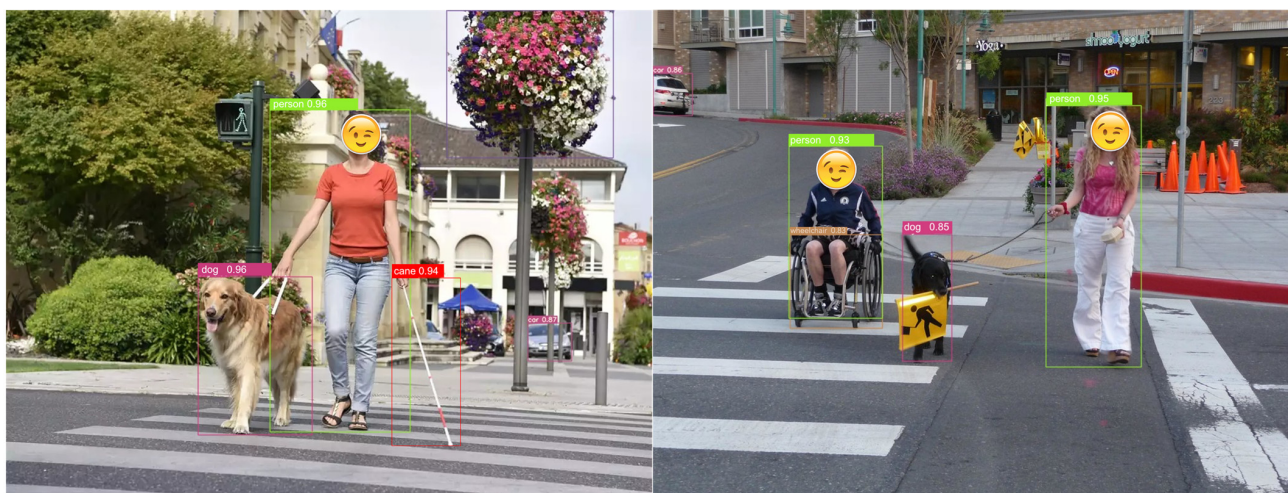
1. The disabled pedestrian sends a “broadcast” message to assistive self-driving cars near his environment every second.
2. Nearby ASCs receive and process the broadcast messages.
3. All ASCs receiving the “broadcast” message send “notification” messages so that vehicles behind them out of range of the D-VRU’s broadcast message can be alerted to the presence of a person with a disability. The “notification” message informs vehicles out of range of the presence of a person with a disability at the intersection. To extend

- the range of the notification, the receiving vehicles forward the message. After three hops, the message is discarded.
4. When the car detects the D-VRU, the hand gesture recognition process starts to identify the pedestrian's intention.
  5. When the ASC detects the intent of the D-VRU, it sends an "intent" message to nearby vehicles. This message indicates what action the pedestrian will take.
  6. The ASC chooses the interaction interface that best suits the physical conditions of the D-VRU.
  7. Finally, the ASC notifies the D-VRU, via the appropriate interface, of the action to be taken to establish a secure environment.

An important point to note is that since each D-VRU sends its own broadcast message, it is possible to have multiple D-VRUs with different limitations at the crosswalk. The ASCs can process all messages and communicate with each D-VRU through the interface best suited to the limitations of each disabled person.

### 3.6. D-VRU Detection Algorithm

A YOLO-based network expanded the self-driving car detection models while integrating the assistive concept into the detection process. In this case, as an example of how to integrate the assistive concept into the self-driving car environment, the model detects blind people or those in wheelchairs. For blind people, the detection process seeks a cane or a guide dog as a specific characteristic for this type of pedestrian. In addition, the model training process used three free and available datasets. The first is composed of 155 images [68], the second is composed of 514 images [69], and the third dataset contains 9308 images [70]. In addition, 23 images were obtained from the Internet to create an experimental dataset of 10,000 images. All images from the testing dataset were later exported into the YOLO7 format to train the model to detect pedestrians using canes and pedestrians using wheelchairs. After training the model, detection tests were performed with images from the Internet. Figure 7 shows the results of the detection process obtained by the algorithm used.



**Figure 7.** Example of D-VRU detection using the YOLO network-based model. Original images were obtained from Flickr.

### 3.7. Hand Gesture Detection Algorithm

For assistive technologies to be successful and become a reality, it is necessary to correctly identify the action (pedestrian intention from now on) that the D-VRU intends to perform as fast and efficiently as possible. We believe defining an inclusive mechanism that allows complete interaction between D-VRUs and ASCs is necessary while preserving both parties' privacy. In addition, the self-driving car must understand pedestrian commu-

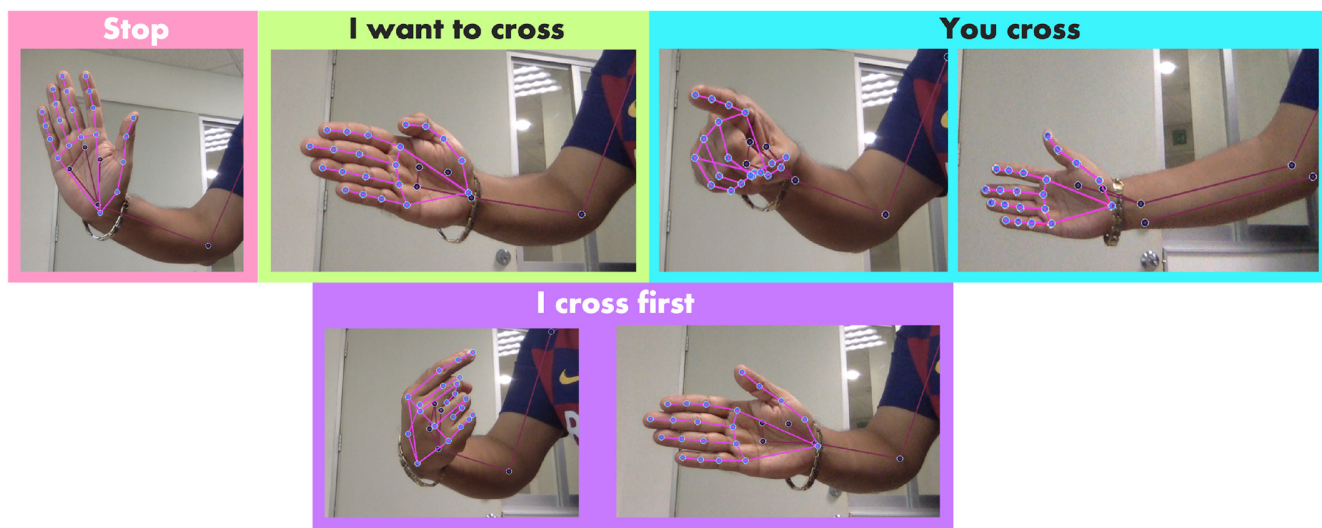


nication within the vehicular environment. In vehicle intelligence, neural networks have become essential for pedestrian intent detection.

Pedestrian intention detection efforts have focused on the study of pose estimation and direction of motion [71–73]. Pedestrians should be able to communicate with the self-driving car through hand gestures. Hand gestures help pedestrians and drivers communicate non-verbally. For example, in some countries, hand gestures are considered traffic rules. Drivers use them to indicate their intentions when a tail or brake light is not working [74,75]. The ASC should be able to detect the hand movements of a pedestrian near a crosswalk and interpret the pedestrian signals to determine the pedestrian's intent. Analyzing the sequence of hand movements is required to identify and classify the signal.

The use of hand gestures to communicate with autonomous cars has been worked on in [76], where the authors proposed the GLADAS system, a deep learning-based system that they evaluated by virtual simulation. This system obtained 94.56% and 85.91% accuracy in the F1 score metric.

However, this proposal was not applied to a real gesture identification scenario. Therefore, our proposal develops a model for a real-life environment. To test communication using a sign model, we defined four specific signs to identify the pedestrian's intention of “stop”, “I want to cross”, “you cross”, and “I will cross first” (Figure 8).



**Figure 8.** Representation of the signals used for communication between ASC and D-VRU.

Deep learning is one of the techniques used for pattern identification. Specifically, recurrent neural networks are a class of networks used to process and obtain information from sequential data [77–79]. Neural networks have an activation function that only acts in one direction, from the input layer to the output layer, which prevents it from remembering previous values. A recurrent neural network (RNN) is similar but includes connections that point “backward” allowing feedback between neurons within the layers. Unlike other neural networks, which process one piece of data at a time, RNNs can process sequences of data (videos, conversations, and texts) but are not used to classify a particular piece of data [80–82]. Instead, they can generate new sequences and incorporate feedback, which creates temporality, allowing the network to have memory.

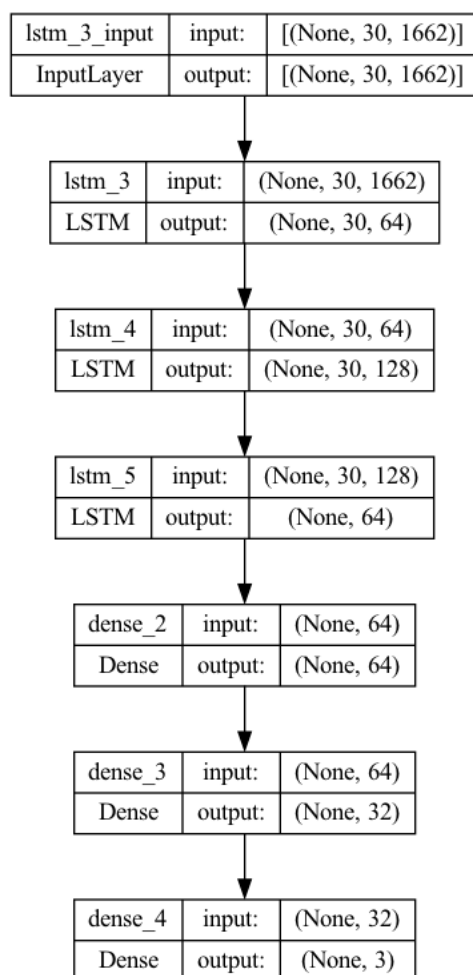
Recurrent neural networks use the concept of recurrence to generate an output (referred to as activation). The generated and recurring inputs use a “temporary memory” to obtain the desired output. One of the most widely used recurrent networks is LSTM (long short-term memory), which “remembers” a relevant piece of data in the sequence and preserves it for several moments, allowing them to have both short-term and long-term memory [83].



This work proposes a model based on the LSTM network to detect the defined hand gestures that may enable ASC and D-VRU communication. Although hand gestures may vary depending on the country and culture, the idea is only to exemplify how gesture recognition is applied to self-driving cars and creating a model based on LSTM that allows the self-driving car to identify such gestures.

For the hand gesture detection model, the MediaPipe framework (MP) was used [83], an open-source machine learning framework for creating models for face feature detection, hand tracing, and object detection, among other things. MP Holistic makes use of the MediaPipe models (pose, face mesh, and hands) to create 543 reference points related to pose (33 points), face (with 468 points), and 21 reference points for each hand. We used the tool for hand tracking, creating 21 3D landmarks with multi-hand support based on high-performance palm detection and hand landmark modeling.

Routines extracted key points using MP. Subsequently, these routines collected the key points of 400 sequences of each defined gesture, preprocessed the data, and created data labels and features. The model for hand gesture identification design uses the LSTM-based neural network. Figure 9 shows the structure of the developed network. The total number of parameters used was 596,675.



**Figure 9.** Representation of the layer structure of the proposed model for hand gesture recognition.

#### 4. Results and Discussions

This section analyzes and discusses the results obtained in the performance evaluation of the proposed model.

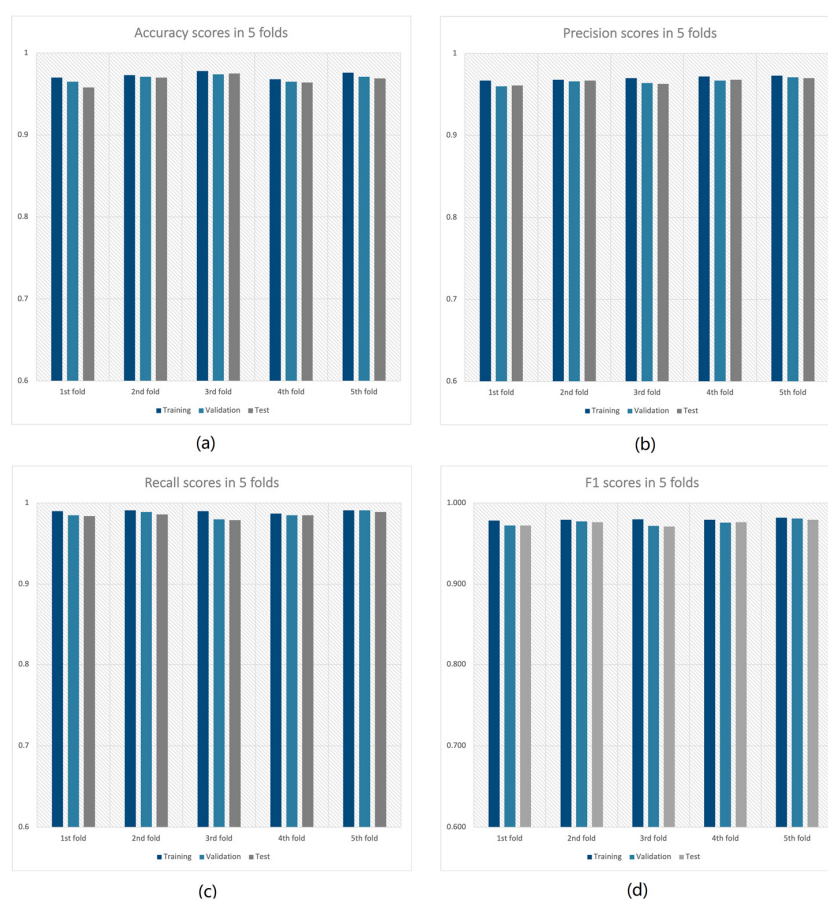
#### 4.1. Evaluation Metrics

One of the essential points for a classification process is the determination of the correct estimator. These models must be evaluated to determine their effectiveness. Performance evaluation metrics are based on the total number of true positives, true negatives, false positives, and false negatives.

There are four performance metrics to evaluate the classification model's effectiveness: (I) accuracy, which reflects the model's ability to predict all classes; (II) precision, which shows the model's ability to detect positive classes from all predicted positive classes; (III) recall accuracy shows the model's ability to detect positive classes of all current positive classes, and (IV) F1 score represents the harmonic mean of precision and recall. Generally, the input is the image containing the object(s) to be classified, and as output, a class label is placed inside the image. The most widely used algorithms are logistic regression, Naïve Bayes, stochastic gradient descent, k-nearest neighbors, decision trees, random forests, and support vector machines [84].

#### 4.2. Data Validation

Typically, the dataset is randomly split to generate one subset for training and another for validation. In most models, 70 to 80% of the data is reserved for training and the rest for validation. However, when the data is limited, this technique could be more effective because some of the information in the data may be omitted during the training phase, causing a bias in the results. K-folds are used to ensure that all dataset features are considered during training. This technique is highly recommended, especially when there is limited data. In this work, the k-fold technique was used. The k-value parameter was set to 5 to split the dataset. Figure 10 shows the results of this evaluation.



**Figure 10.** Results obtained from the data validation process: (a) accuracy, (b) precision, (c) recall, and (d) F1 score.

The results are an indication of the stability of the model. Regarding accuracy, an average value of 97.3% was obtained for the training phase, 96.9% for the validation phase, and 96.7% for the test phase. For the precision metric, the mean values obtained were 97% for training, 96.5% for validation, and 96.5% for testing. The recall measure's mean values were 0.989 for the training phase, 0.986 for the validation phase, and 0.984 for the test phase. Finally, the F1 score parameter had values of 0.9797, 0.9756, and 0.9751 for the training, validation, and test phases, respectively. Therefore, very similar behavior is observed with all five folds in the different phases of the model, which shows that the model's data are accurate and not under- or over-fitted.

#### 4.3. ASC–D-VRU Interaction

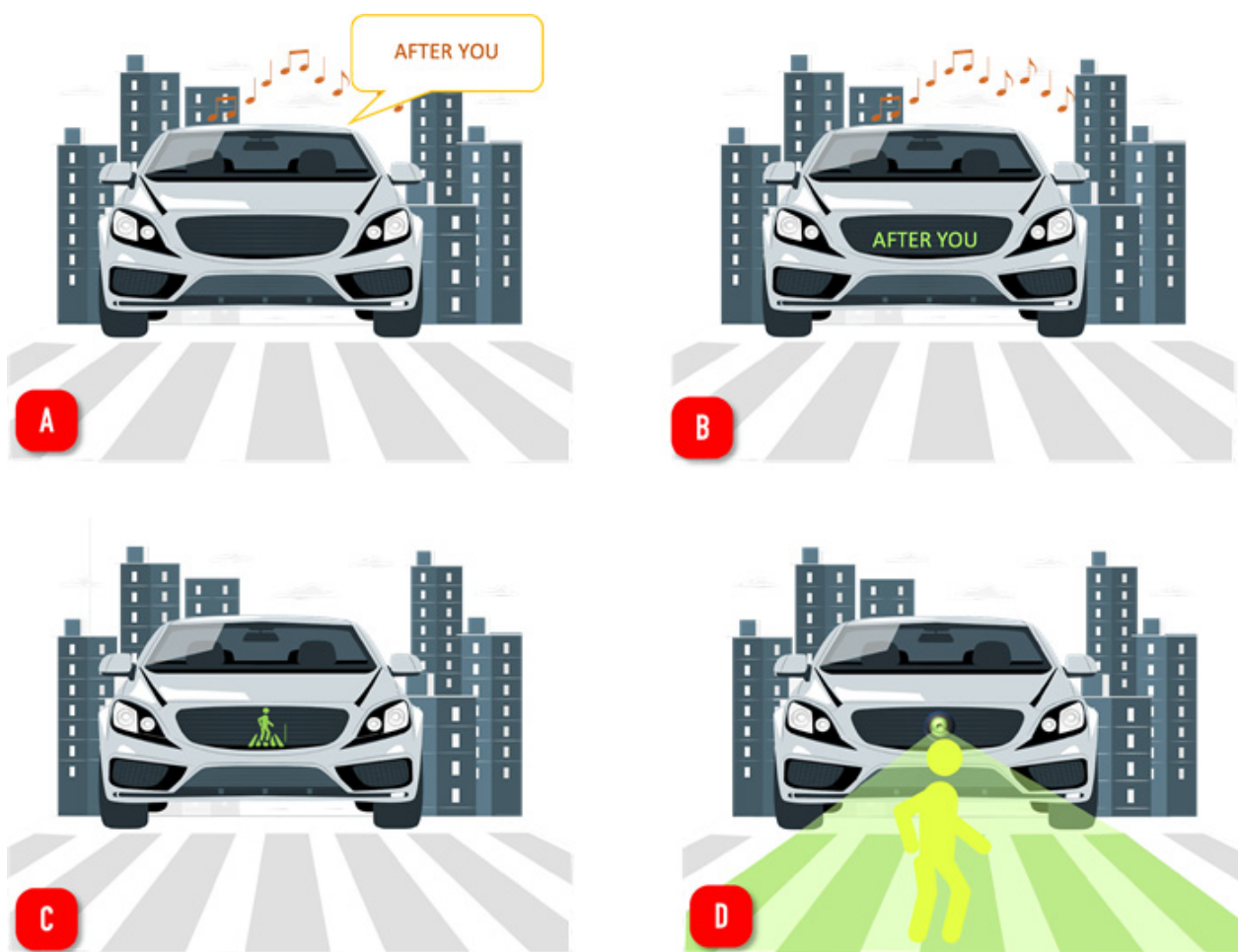
Finally, the interaction process focuses on how the car will notify the pedestrian about the intentions or actions to follow.

Interaction technologies fall into three categories. (i) Visual interfaces that allow information display through text, icons, holograms, or projections. (ii) Acoustic signals to transmit the message to visually impaired people. (iii) Anthropomorphic interfaces that use human characteristics to communicate with the pedestrian to provide a sense of safety. For example, some researchers propose simulating eye contact with pedestrians using an interface that affects eye movement.

In recent years, work has been ongoing on developing human–computer interfaces for self-driving cars. These interfaces are called external Human–Machine Interfaces (eHMIs) and are installed outside the car.

Once the eHMI locates the disabled pedestrian, it proceeds to identify whether the pedestrian communicates with the car through a series of hand gestures. Based on this assessment, the pedestrian's intention is recognized, and a series of measures improve the functional capabilities of the disabled pedestrian when crossing. One of the first steps is the selection of the most appropriate communication interface to notify the pedestrian with a disability of the following action. The interface selection is directly related to the type of D-VRU identified, for example, by using the acoustic interface when the detected pedestrian is a blind person. In this case, the acoustic interface could allow the assistive self-driving car to emit a voice message such as “after you”, “can cross”, or “stop” (Figure 11A). If the type of pedestrian detected is a person in a wheelchair, then the self-driving car would use the visual interface to display messages that could be text messages such as “may cross”, “stop”, and “after you”. Messages are complemented with icons or holographic projections to improve understanding of the text due to a lack of language knowledge or inability to read (Figure 11B–D).

The proposal presented in this paper has the potential to make a significant difference to the mobility and safety of people with disabilities in the road environment. However, there may be a number of situations that may affect the performance of the proposed solution. For example, one of the problems of the proposal may be the inclusion problem. Let us analyze the scenario in which many pedestrians are together with a disabled individual. The proposal could detect the disabled person through the wireless communication process. However, there may be a situation where the pedestrians obstruct the disabled person, which hinders the interaction process between the ASC and the D-VRU. Therefore, this proposal should be expanded to include safe and predictable ways for disabled people to interact with other road users, such as drivers and cyclists, to avoid confusion and dangerous situations. To inform other users of the intentions of the pedestrian with a disability, clear and universally understood communication interfaces, such as light signals or sounds, can be developed, which could reduce occlusion problems.



**Figure 11.** Representation of technological interfaces: (A) acoustic interface to indicate the action to follow by the D-VRU, (B) acoustic and visual interface, (C) display on the front of the vehicle showing information on what the D-VRU should do, and (D) projection of message on the road with visual elements to indicate to the D-VRU the option of “safe crossing”.

## 5. Conclusions

Self-driving cars will interact more closely with other human agents on social roads. For this reason, having quantitative models to predict these interactive behaviors has become increasingly important. This article analyzes the interaction process between disabled people and self-driving cars. The idea is to integrate the concept of assistive technology into the self-driving car environment, creating an assistive self-driving car that extends the pedestrian detection capabilities to a process of identification of disabled people using deep learning technology. A bi-directional interaction mechanism between pedestrians with disabilities and the self-driving car was proposed. Through an algorithm based on recurrent neural networks for hand gesture detection and external human–computer interfaces, a bidirectional interaction to increase the safety and reliability of the disabled pedestrian to perform activities within the road environment, such as crossing a street, was achieved.

Unlike other works proposed in the literature, this proposal does not require installing or implementing additional infrastructure. Instead, it uses the processing capacity already implemented in self-driving cars, saving investment costs and extending its coverage to any area of the road traffic ecosystem. Furthermore, although the proposal focuses on pedestrians with disabilities, all other pedestrians can use the hand gesture mechanism to communicate with self-driving cars. As a commonly used communication mode in today’s driving environment, it does not imply a cognitive burden for pedestrians to learn something new.

**Author Contributions:** Conceptualization, J.G.-I., J.C.-C., I.A.-V., and A.R.-M.; methodology, J.G.-I., J.C.-C., I.A.-V., and A.R.-M.; software, J.G.-I., J.C.-C., I.A.-V., and A.R.-M.; validation, J.G.-I., J.C.-C., I.A.-V., and A.R.-M.; formal analysis, J.G.-I., J.C.-C., I.A.-V., and A.R.-M.; investigation, J.G.-I., J.C.-C., I.A.-V., and A.R.-M.; resources, J.G.-I., J.C.-C., I.A.-V., and A.R.-M.; data curation, J.G.-I., J.C.-C., I.A.-V., and A.R.-M.; writing—original draft preparation, J.G.-I., J.C.-C., I.A.-V., and A.R.-M.; writing—review and editing, J.G.-I., J.C.-C., I.A.-V., and A.R.-M.; visualization, J.G.-I., J.C.-C., I.A.-V., and A.R.-M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially funded by the project “Drone fleet monitoring and optimization of commercial operations flight plans” PID2020-116377RB-C21. Research Challenges: R&D&I Projects, Spanish Research Agency.

**Data Availability Statement:** Data available on request due to privacy restrictions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. World Health Organization. Road Traffic Injuries. Available online: <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries> (accessed on 9 February 2023).
2. European Commission. ITS & Vulnerable Road Users. 2015. Available online: [https://transport.ec.europa.eu/transport-themes/intelligent-transport-systems/road/action-plan-and-directive/its-vulnerable-road-users\\_en](https://transport.ec.europa.eu/transport-themes/intelligent-transport-systems/road/action-plan-and-directive/its-vulnerable-road-users_en) (accessed on 11 October 2023).
3. Schwartz, N.; Buliung, R.; Daniel, A.; Rothman, L. Disability and pedestrian road traffic injury: A scoping review. *Health Place* **2022**, *77*, 102896. [CrossRef] [PubMed]
4. Kraemer, J.D.; Benton, C.S. Disparities in road crash mortality among pedestrians using wheelchairs in the USA: Results of a capture–recapture analysis. *BMJ Open* **2015**, *5*, e008396. [CrossRef] [PubMed]
5. Society of Automotive Engineers. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. 2014. Available online: [https://www.sae.org/standards/content/j3016\\_202104/](https://www.sae.org/standards/content/j3016_202104/) (accessed on 11 October 2023).
6. Lahijanian, M.; Kwiatkowska, M. *Social Trust: A Major Challenge for the Future of Autonomous Systems*; AAAI Association for the Advancement of Artificial Intelligence: Palo Alto, CA, USA, 2016.
7. Rasouli, A.; Tsotsos, J.K. Autonomous Vehicles That Interact With Pedestrians: A Survey of Theory and Practice. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 900–918. [CrossRef]
8. Kaur, K.; Rampersad, G. Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars. *J. Eng. Technol. Manag.* **2018**, *48*, 87–96. [CrossRef]
9. Ragesh, N.K.; Rajesh, R. Pedestrian Detection in Automotive Safety: Understanding State-of-the-Art. *IEEE Access* **2019**, *7*, 47864–47890. [CrossRef]
10. Reyes-Muñoz, A.; Guerrero-Ibáñez, J. Vulnerable Road Users and Connected Autonomous Vehicles Interaction: A Survey. *Sensors* **2022**, *22*, 4614. [CrossRef]
11. ATiA. *What Is AT?* Assistive Technology Industry Association: Chicago, IL, USA, 2015; Available online: <https://www.atia.org/home/at-resources/what-is-at/> (accessed on 11 October 2023).
12. Zhou, Y.; Li, G.; Wang, L.; Li, S.; Zong, W. Smartphone-based Pedestrian Localization Algorithm using Phone Camera and Location Coded Targets. In Proceedings of the 2018 Ubiquitous Positioning, Indoor Navigation and Location-Based Services (UPINLBS), Wuhan, China, 22–23 March 2018; pp. 1–7. [CrossRef]
13. Yang, L.; Zou, J.; Li, Y.; Rizos, C. Seamless pedestrian navigation augmented by walk status detection and context features. In Proceedings of the 2016 Fourth International Conference on Ubiquitous Positioning, Indoor Navigation and Location Based Services (UPINLBS), Shanghai, China, 2–4 November 2016; pp. 20–28. [CrossRef]
14. Shit, R.C.; Sharma, S.; Puthal, D.; James, P.; Pradhan, B.; van Moorsel, A.; Zomaya, A.Y.; Ranjan, R. Ubiquitous Localization (UbiLoc): A Survey and Taxonomy on Device Free Localization for Smart World. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3532–3564. [CrossRef]
15. Chen, L.; Lin, S.; Lu, X.; Cao, D.; Wu, H.; Guo, C.; Liu, C.; Wang, F.-Y. Deep Neural Network Based Vehicle and Pedestrian Detection for Autonomous Driving: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 3234–3246. [CrossRef]
16. Zheng, G.; Chen, Y. A review on vision-based pedestrian detection. In Proceedings of the 2012 IEEE Global High Tech Congress on Electronics, Shenzhen, China, 18–20 November 2012; pp. 49–54. [CrossRef]
17. Guéguen, N.; Meineri, S.; Eyssartier, C. A pedestrian’s stare and drivers’ stopping behavior: A field experiment at the pedestrian crossing. *Saf. Sci.* **2015**, *75*, 87–89. [CrossRef]
18. Rothenbücher, D.; Li, J.; Sirkin, D.; Mok, B.; Ju, W. Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles. In Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), New York, NY, USA, 26–31 August 2016. [CrossRef]
19. Merat, N.; Louw, T.; Madigan, R.; Wilbrink, M.; Schieben, A. What externally presented information do VRUs require when interacting with fully Automated Road Transport Systems in shared space? *Accid. Anal. Prev.* **2018**, *118*, 244–252. [CrossRef]



20. Reig, S.; Norman, S.; Morales, C.G.; Das, S.; Steinfeld, A.; Forlizzi, J. A Field Study of Pedestrians and Autonomous Vehicles. In Proceedings of the AutomotiveUI '18: The 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Toronto, ON, Canada, 23–25 September 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 198–209. [\[CrossRef\]](#)
21. Löcken, A.; Golling, C.; Riener, A. How Should Automated Vehicles Interact with Pedestrians? A Comparative Analysis of Interaction Concepts in Virtual Reality. In Proceedings of the AutomotiveUI '19 the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Utrecht, The Netherlands, 21–25 September 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 262–274. [\[CrossRef\]](#)
22. Vinkhuyzen, E.; Cefkin, M. Developing Socially Acceptable Autonomous Vehicles. *Ethnogr. Prax. Ind. Conf. Proc.* **2016**, 2016, 522–534. [\[CrossRef\]](#)
23. Habibovic, A.; Fabricius, V.; Anderson, J.; Klingegård, M. Communicating Intent of Automated Vehicles to Pedestrians. *Front. Psychol.* **2018**, 9, 1336. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Habibovic, A.; Andersson, J.; Lundgren, V.M.; Klingegård, M.; Englund, C.; Larsson, S. External Vehicle Interfaces for Communication with Other Road Users? In *Road Vehicle Automation 5*; Meyer, G., Beiker, S., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 91–102.
25. Strauss, T. Breaking down the Language Barrier between Autonomous Cars and Pedestrians. 2018. Available online: <https://uxdesign.cc/wave-breaking-down-the-language-barrier-between-autonomous-cars-and-pedestrians-autonomy-tech-a8ba1f6686> (accessed on 3 May 2022).
26. Autocar. The Autonomous Car that Smiles at Pedestrians. 2016. Available online: <https://www.autocar.co.uk/car-news/new-cars/autonomous-car-smiles-pedestrians> (accessed on 3 May 2022).
27. Kitayama, S.; Kondou, T.; Ohyabu, H.; Hirose, M. Display System for Vehicle to Pedestrian Communication. SAE Technical paper 2017-01-0075. 2017. Available online: <https://www.sae.org/publications/technical-papers/content/2017-01-0075/> (accessed on 11 October 2023).
28. Deb, S.; Strawderman, L.J.; Carruth, D.W. Investigating pedestrian suggestions for external features on fully autonomous vehicles: A virtual reality experiment. *Transp. Res. Part F Traffic Psychol. Behav.* **2018**, 59, 135–149. [\[CrossRef\]](#)
29. Costa, G. Designing Framework for Human-Autonomous Vehicle Interaction. Master's Thesis, Keio University Graduate School of Media Design, Yokohama, Japan, 2017.
30. Chang, C.-M.; Toda, K.; Sakamoto, D.; Igarashi, T. Eyes on a Car: An Interface Design for Communication between an Autonomous Car and a Pedestrian. In Proceedings of the AutomotiveUI '17: The 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany, 24–27 September 2017; p. 73. [\[CrossRef\]](#)
31. Ochiai, Y.; Toyoshima, K. Homunculus: The Vehicle as Augmented Clothes. In Proceedings of the AH '11: The 2nd Augmented Human International Conference, Tokyo, Japan, 13 March 2011; Association for Computing Machinery: New York, NY, USA, 2011. [\[CrossRef\]](#)
32. Jaguar Land Rover. The Virtual Eyes Have It. 2018. Available online: <https://www.jaguarlandrover.com/2018/virtual-eyes-have-it> (accessed on 16 October 2023).
33. Le, M.C.; Do, T.-D.; Duong, M.-T.; Ta, T.-N.-M.; Nguyen, V.-B.; Le, M.-H. Skeleton-based Recognition of Pedestrian Crossing Intention using Attention Graph Neural Networks. In Proceedings of the 2022 International Workshop on Intelligent Systems (IWIS), Ulsan, Republic of Korea; 2022; pp. 1–5. [\[CrossRef\]](#)
34. Mínguez, R.Q.; Alonso, I.P.; Fernández-Llorca, D.; Sotelo, M.Á. Pedestrian Path, Pose, and Intention Prediction Through Gaussian Process Dynamical Models and Pedestrian Activity Recognition. *IEEE Trans. Intell. Transp. Syst.* **2019**, 20, 1803–1814. [\[CrossRef\]](#)
35. Fang, Z.; López, A.M. Intention Recognition of Pedestrians and Cyclists by 2D Pose Estimation. *IEEE Trans. Intell. Transp. Syst.* **2020**, 21, 4773–4783. [\[CrossRef\]](#)
36. Perdana, M.I.; Anggraeni, W.; Sidharta, H.A.; Yuniarno, E.M.; Purnomo, M.H. Early Warning Pedestrian Crossing Intention from Its Head Gesture using Head Pose Estimation. In Proceedings of the 2021 International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya, Indonesia, 21–22 July 2021; pp. 402–407. [\[CrossRef\]](#)
37. Rehder, E.; Kloeden, H.; Stiller, C. Head detection and orientation estimation for pedestrian safety. In Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Qingdao, China, 8–11 October 2014; pp. 2292–2297. [\[CrossRef\]](#)
38. Quan, R.; Zhu, L.; Wu, Y.; Yang, Y. Holistic LSTM for Pedestrian Trajectory Prediction. *IEEE Trans. Image Process.* **2021**, 30, 3229–3239. [\[CrossRef\]](#)
39. Huang, Z.; Hasan, A.; Shin, K.; Li, R.; Driggs-Campbell, K. Long-Term Pedestrian Trajectory Prediction Using Mutable Intention Filter and Warp LSTM. *IEEE Robot. Autom. Lett.* **2021**, 6, 542–549. [\[CrossRef\]](#)
40. Mahadevan, K.; Somanath, S.; Sharlin, E. Communicating Awareness and Intent in Autonomous Vehicle-Pedestrian Interaction. In Proceedings of the CHI '18 the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 1–12. [\[CrossRef\]](#)
41. Jaber, A.K.; Abdel-Qader, I. Hybrid Histograms of Oriented Gradients-compressive sensing framework feature extraction for face recognition. In Proceedings of the 2016 IEEE International Conference on Electro Information Technology (EIT), Grand Forks, ND, USA, 19–21 May 2016; pp. 442–447. [\[CrossRef\]](#)

42. Zhang, L.; Zhou, W.; Li, J.; Li, J.; Lou, X. Histogram of Oriented Gradients Feature Extraction Without Normalization. In Proceedings of the 2020 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Ha Long, Vietnam, 8–10 December 2020; pp. 252–255. [\[CrossRef\]](#)
43. Sasongko, A.; Sahbani, B. VLSI Architecture for Fine Grained Pipelined Feature Extraction using Histogram of Oriented Gradient. In Proceedings of the 2019 IEEE 7th Conference on Systems, Process and Control (ICSPPC), Melaka, Malaysia, 13–14 December 2019; pp. 143–148. [\[CrossRef\]](#)
44. Liu, G.; Liu, W.; Chen, X. An Improved Pairwise Rotation Invariant Co-occurrence Local Binary Pattern Method for Texture Feature Extraction. In Proceedings of the 2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 29–31 March 2019; pp. 431–436. [\[CrossRef\]](#)
45. Kaur, N.; Nazir, N. Manik. A Review of Local Binary Pattern Based texture feature extraction. In Proceedings of the 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 3–4 September 2021; pp. 1–4. [\[CrossRef\]](#)
46. Ansari, M.D.; Ghrera, S.P. Feature extraction method for digital images based on intuitionistic fuzzy local binary pattern. In Proceedings of the 2016 International Conference System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 25–27 November 2016; pp. 345–349. [\[CrossRef\]](#)
47. Li, J.; Wong, H.-C.; Lo, S.-L.; Xin, Y. Multiple Object Detection by a Deformable Part-Based Model and an R-CNN. *IEEE Signal Process. Lett.* **2018**, *25*, 288–292. [\[CrossRef\]](#)
48. Jie, G.; Honggang, Z.; Daiwu, C.; Nannan, Z. Object detection algorithm based on deformable part models. In Proceedings of the 2014 4th IEEE International Conference on Network Infrastructure and Digital Content, Beijing, China, 19–21 September 2014; pp. 90–94. [\[CrossRef\]](#)
49. Tang, J.; Lin, Z.; Zhang, Y. Rapid Forward Vehicle Detection Based on Deformable Part Model. In Proceedings of the 2017 2nd International Conference on Multimedia and Image Processing (ICMIP), Wuhan, China, 17–19 March 2017; pp. 27–31. [\[CrossRef\]](#)
50. Huang, K.; Li, J.; Liu, Y.; Chang, L.; Zhou, J. A Survey on Feature Point Extraction Techniques. In Proceedings of the 2021 18th International SoC Design Conference (ISOCC), Jeju Island, Republic of Korea, 6–9 October 2021; pp. 201–202. [\[CrossRef\]](#)
51. Sajat, M.A.S.; Hashim, H.; Tahir, N.M. Detection of Human Bodies in Lying Position based on Aggregate Channel Features. In Proceedings of the 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), Langkawi, Malaysia, 28–29 February 2020; pp. 313–317. [\[CrossRef\]](#)
52. Ragb, H.K.; Ali, R.; Asari, V. Aggregate Channel Features Based on Local Phase, Color, Texture, and Gradient Features for People Localization. In Proceedings of the 2019 IEEE National Aerospace and Electronics Conference (NAECON), Dayton, OH, USA, 15–19 July 2019; pp. 351–355. [\[CrossRef\]](#)
53. Li, Y.; Cui, F.; Xue, X.; Chan, J.C.-W. Coarse-to-fine salient object detection based on deep convolutional neural networks. *Signal Process. Image Commun.* **2018**, *64*, 21–32. [\[CrossRef\]](#)
54. Chen, E.; Tang, X.; Fu, B. A Modified Pedestrian Retrieval Method Based on Faster R-CNN with Integration of Pedestrian Detection and Re-Identification. In Proceedings of the 2018 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, China, 16–17 July 2018; pp. 63–66. [\[CrossRef\]](#)
55. Shi, P.; Wu, J.; Wang, K.; Zhang, Y.; Wang, J.; Yi, J. Research on Low-Resolution Pedestrian Detection Algorithms based on R-CNN with Targeted Pooling and Proposal. In Proceedings of the 2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA), Xi'an, China, 7–10 November 2018; pp. 1–5. [\[CrossRef\]](#)
56. Zhao, Z.; Ma, J.; Ma, C.; Wang, Y. An Improved Faster R-CNN Algorithm for Pedestrian Detection. In Proceedings of the 2021 11th International Conference on Information Technology in Medicine and Education (ITME), Wuyishan, China, 19–21 November 2021; pp. 76–80. [\[CrossRef\]](#)
57. Zhu, K.; Li, L.; Hu, D.; Chen, D.; Liu, L. An improved detection method for multi-scale and dense pedestrians based on Faster R-CNN. In Proceedings of the 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 11–13 December 2019; pp. 1–5. [\[CrossRef\]](#)
58. Malbog, M.A. MASK R-CNN for Pedestrian Crosswalk Detection and Instance Segmentation. In Proceedings of the 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), Kuala Lumpur, Malaysia, 20–21 December 2019; pp. 1–5. [\[CrossRef\]](#)
59. Shen, G.; Jamshidi, F.; Dong, D.; ZhG, R. Metro Pedestrian Detection Based on Mask R-CNN and Spatial-temporal Feature. In Proceedings of the 2020 IEEE 3rd International Conference on Information Communication and Signal Processing (ICICSP), Shanghai, China, 12–15 September 2020; pp. 173–178. [\[CrossRef\]](#)
60. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [\[CrossRef\]](#)
61. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 21–37.
62. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [\[CrossRef\]](#)
63. IEEE. IEEE Draft Guide for Architectural Framework and Application of Federated Machine Learning. In *IEEE P3652.1/D6*; IEEE: Piscataway, NJ, USA, 2020; pp. 1–70.

64. Bonawitz, K.; Kairouz, P.; McMahan, B.; Ramage, D. Federated Learning and Privacy. *Commun. ACM* **2022**, *65*, 90–97. [CrossRef]
65. Sommer, C.; German, R.; Dressler, F. Bidirectionally Coupled Network and Road Traffic Simulation for Improved IVC Analysis. *IEEE Trans. Mob. Comput.* **2011**, *10*, 3–15. [CrossRef]
66. Guerrero-Ibañez, A.; Amezcua-Valdovinos, I.; Contreras-Castillo, J. Integration of Wearables and Wireless Technologies to Improve the Interaction between Disabled Vulnerable Road Users and Self-Driving Cars. *Electronics* **2023**, *12*, 3587. [CrossRef]
67. U.S. Space Force. GPS Accuracy. Available online: <https://www.gps.gov/systems/gps/performance/accuracy/> (accessed on 27 September 2023).
68. Nozaki. Whitecane Dataset, Roboflow Universe. Roboflow, May 2022. Available online: <https://universe.roboflow.com/nozaki/whitecane-mzmlr> (accessed on 11 October 2023).
69. Wheelchair Detection Dataset, Roboflow Universe. Roboflow, November 2021. Available online: <https://universe.roboflow.com/2458761304-qq-com/wheelchair-detection> (accessed on 11 October 2023).
70. Jang, B.H. Visually impaired (whitecane). Available online: <https://www.kaggle.com/datasets/jangbyeonghui/visually-impairedwhitecane> (accessed on 24 April 2023).
71. Yang, J.; Gui, A.; Wang, J.; Ma, J. Pedestrian Behavior Interpretation from Pose Estimation. In Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), Indianapolis, IN, USA, 19–22 September 2021; pp. 3110–3115. [CrossRef]
72. Samant, A.P.; Warhade, K.; Gunale, K. Pedestrian Intent Detection using Skeleton-based Prediction for Road Safety. In Proceedings of the 2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS), Ernakulam, India, 2–4 September 2021; pp. 238–242. [CrossRef]
73. Saleh, K.; Hossny, M.; Nahavandi, S. Intent Prediction of Pedestrians via Motion Trajectories Using Stacked Recurrent Neural Networks. *IEEE Trans. Intell. Veh.* **2018**, *3*, 414–424. [CrossRef]
74. Hand Signals. Available online: [https://static.nhtsa.gov/nhtsa/downloads/NTI/Responsible\\_Walk-Bike\\_Activities/ComboLessons/L3Handouts/8009\\_HandSignals\\_122811\\_v1a.pdf](https://static.nhtsa.gov/nhtsa/downloads/NTI/Responsible_Walk-Bike_Activities/ComboLessons/L3Handouts/8009_HandSignals_122811_v1a.pdf) (accessed on 16 April 2023).
75. DVM. Hand Signals Guide. Available online: <https://www.dmv.org/how-to-guides/hand-signals-guide.php> (accessed on 16 April 2023).
76. Shaotran, E.; Cruz, J.J.; Reddi, V.J. Gesture Learning For Self-Driving Cars. In Proceedings of the 2021 IEEE International Conference on Autonomous Systems (ICAS), Montreal, QC, Canada, 11–13 August 2021; pp. 1–5. [CrossRef]
77. Uçkun, F.A.; Özer, H.; Nurbas, E.; Onat, E. Direction Finding Using Convolutional Neural Networks and Convolutional Recurrent Neural Networks. In Proceedings of the 2020 28th Signal Processing and Communications Applications Conference (SIU), Gaziantep, Turkey, 5–7 October 2020; pp. 1–4. [CrossRef]
78. Xiao, Y.; Keung, J. Improving Bug Localization with Character-Level Convolutional Neural Network and Recurrent Neural Network. In Proceedings of the 2018 25th Asia-Pacific Software Engineering Conference (APSEC), Nara, Japan, 4–7 December 2018; pp. 703–704. [CrossRef]
79. Podlesnykh, I.A.; Bakhtin, V.V. Mathematical Model of a Recurrent Neural Network for Programmable Devices Focused on Fog Computing. In Proceedings of the 2022 Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), Saint Petersburg, Russia, 25–28 January 2022; pp. 395–397. [CrossRef]
80. Song, J.; Zhao, Y. Multimodal Model Prediction of Pedestrian Trajectories Based on Graph Convolutional Neural Networks. In Proceedings of the 2022 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML), Xi'an, China, 28–30 October 2022; pp. 271–275. [CrossRef]
81. Zha, B.; Koroglu, M.T.; Yilmaz, A. Trajectory Mining for Localization Using Recurrent Neural Network. In Proceedings of the 2019 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 5–7 December 2019; pp. 1329–1332. [CrossRef]
82. Ono, T.; Kanamaru, T. Prediction of pedestrian trajectory based on long short-term memory of data. In Proceedings of the 2021 21st International Conference on Control, Automation and Systems (ICCAS), Jeju Island, Republic of Korea, 12–15 October 2021; pp. 1676–1679. [CrossRef]
83. Lugaresi, C.; Tang, J.; Nash, H.; McClanahan, C.; Uboweja, E.; Hays, M.; Zhang, F.; Chang, C.L.; Yong, M.G.; Lee, J.; et al. MediaPipe: A Framework for Building Perception Pipelines. *arXiv* **2019**, arXiv:1906.08172.
84. Khan, I.; Zhang, X.; Rehman, M.; Ali, R. A Literature Survey and Empirical Study of Meta-Learning for Classifier Selection. *IEEE Access* **2020**, *8*, 10262–10281. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.