LUISS

Department
of Business and Management

Bachelor’s degree in Management and Computer Science

The use and application of AI and ML in the automotive industry

Prof. Giuseppe Italiano

SUPERVISOR

246751

CANDIDATE

Academic Year 2021/2022

INDEX:

LIST OF ABBREVIATIONS

AI Artificial intelligence

ML Machine learning

ADAS Advanced Driver Assistance System

ECUs Electronic central units

SVM Support Vector Machines

DMS Driver monitoring system

LBP Local binary pattern

AD Autonomous driving

AV Autonomous vehicle

CNN Convolutional neural network

SLAM Simultaneous localization and mapping

LIST OF TABLES/IMAGES

1.0 INTRODUCTION

At the basis of this study there is the willingness of make clearer what ADAS and Autonomous Driving are, and moreover to try to make understand how they are extremely based and dependent from Artificial Intelligence and Machine Learning.

I've chosen this topic because in average a person spends one hour and half a day driving, so 17 days a year. Technology is by definition the application of scientific knowledge for practical purposes, so its goal is in simple words to simplify humans' life. Driving can be considered an extremely habitual action made with no effort, but in reality, it's an action that make spend us a lot of energies unconsciously. And here it's the role of technologies: replacing human actions in order to make that effort for them and also to increase security.
These technologies are fundamental developed thanks to the techniques of AI and ML, and this is what I want to show in this work.

The thesis is developed on three main chapters: the first one talk about ADAS: what they are, their main characteristics and describe the functioning of two of the most important and diffused ones.
The second main chapter talk about autonomous driving in particular about the three main block that permit it's functioning, that obviously include the use of the already defined ADAS.
Last chapter talk about the future of this topic looking at possible scenarios of the automotive sector, but also the need of new reglementary for the new scenarios bring by these technologies.

In addition, a short interview to about 50 people was conducted, in order to have real data and opinion about the diffusion and use of these systems.

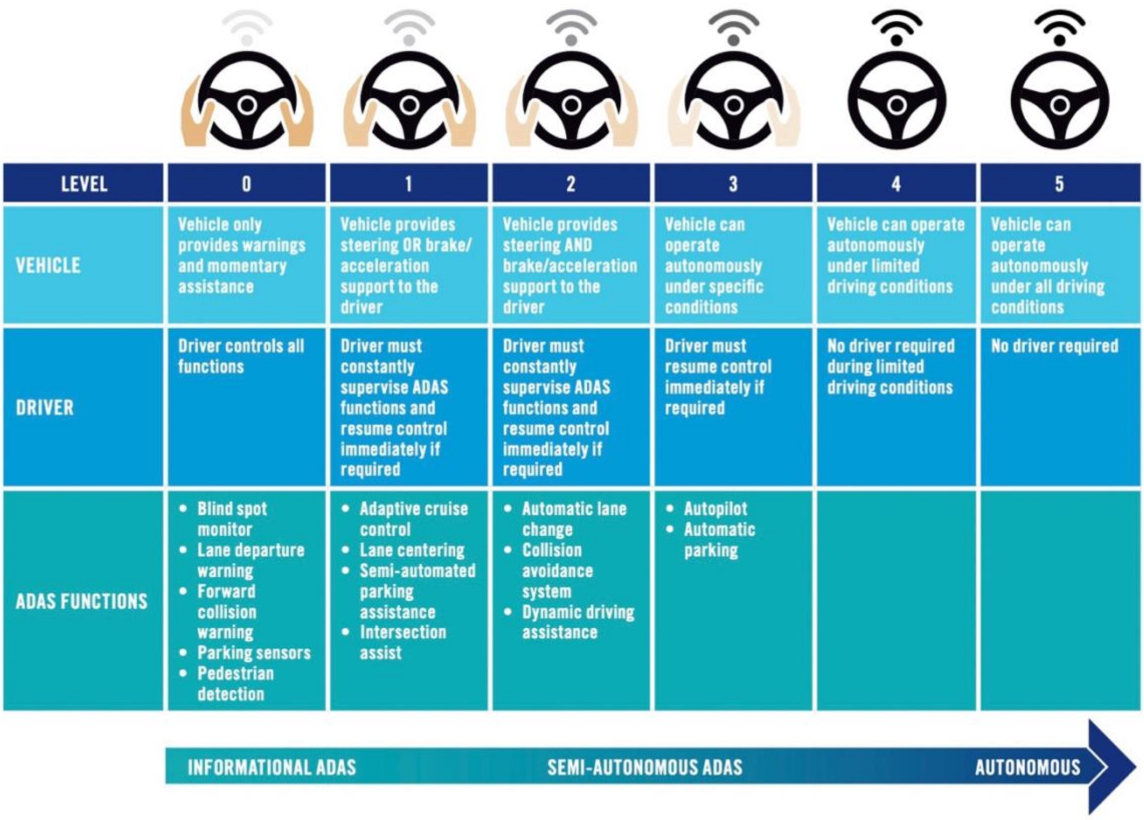
2.0 ADAS SYSTEMS

2.1 DEFINITION AND DIFFUSION

Advanced Driver Assistance System (ADAS) can be defined as a collection of electronic systems that aid drivers by improving safety, comfort and efficiency while driving and parking.
ADAS is the fastest growing technology segment in the automotive market, worth an estimated $24 billion in 2018 and predicted to reach $92 billion by 2025.

ADAS functioning is possible due to a series of technologies that allow on-board computers to perceive the external environment, through an array of on- board sensors and data processing. Cameras, ultrasonic sensors, radar and lately also LiDAR sensor are distributed all around the car with the goal of collecting vast amounts of data. The goal is not only react to external events, but also anticipate them sooner than a human alone.

The great diffusion of different types of ADAS has brought to create a classification system based on the level of autonomy they provide, starting from the basic informational ones, where the driver remains in full control, moving to the fully autonomous vehicles.



LEVEL	0	1	2	3	4	5
VEHICLE	Vehicle only provides warnings and momentary assistance	Vehicle provides steering OR brake/ acceleration support to the driver	Vehicle provides steering AND brake/acceleration support to the driver	Vehicle can operate autonomously under specific conditions	Vehicle can operate autonomously under limited driving conditions	Vehicle can operate autonomously under all driving conditions
DRIVER	Driver controls all functions	Driver must constantly supervise ADAS functions and resume control immediately if required	Driver must constantly supervise ADAS functions and resume control immediately if required	Driver must resume control immediately if required	No driver required during limited driving conditions	No driver required
ADAS FUNCTIONS	• Blind spot monitor • Lane departure warning • Forward collision warning • Parking sensors • Pedestrian detection	• Adaptive cruise control • Lane centering • Semi-automated parking assistance • Intersection assist	• Automatic lane change • Collision avoidance system • Dynamic driving assistance	• Autopilot • Automatic parking		

Similar to safety-related systems in other domains, the whole development process of ADAS needs to comply with industry-wide standards as well as other governing regulations.
ADAS have to face different challenges. The main challenge is developing safe systems that work reliable everywhere. With safety we also intended that a car's system must be protected from unwanted influences and cyber-attacks.
Other challenges in ADAS include the necessity of reliable functionality wherever vehicles may be used. Traffic laws and sign are not globally standardized, in addition, there can be temporary change in speed limits and other posted restrictions. ADAS must be able also to work dependably in different and sometime very harsh physical environment.

Despite advanced ADAS being relatively new and only available from premium market segment, we are seeing rising driver acceptance of self-driving technology. For example, a recent study by MIT's Human Centered AI Institute on Tesla journey data found drivers relied on Tesla's Autopilot for over 30% of their total miles driven. Currently, tech-savvy consumers show the greatest interest in self driving features, with 69% interested in using autonomous vehicles when they become available, compared to just 49% of all surveyed drivers.
Consumer interest varies by ADAS feature and their associated autonomy levels. For example, two of the most popular ADAS functions today are blind spot warning (Level 0) and parking assist (Level 1), which achieve 84% and 81% customer satisfaction respectively.
Informational ADAS features are now widely accepted whilst newer, more advanced ADAS features are gaining acceptance, with Level 1 features such as ACC and automatic emergency braking achieving 75% and 69% customer satisfaction respectively.
As ADAS technology improves and such features become more cost-effective, consumer interest in the mid-range market segment will increase – 79% of drivers state they would choose an ADAS-enabled vehicle if it was available at no additional cost when compared with a conventional vehicle.

2.2 ADAS ARCHITECTURE

Modern vehicles can contain up to 90 electronic central units (ECUs) which process the large amount of data coming from the different subsystems. For example, an ECU responsible for ultrasonic parking assist will process the ultrasound data to determine, at a basic level, the immediate surroundings of a vehicle during parking.
These useful, but relatively simple, Informational ADAS functions are facilitated by local processing at the sensor node. These subsystems operate independently from each other and the central computer.

Increasing the level of autonomy of ADAS, so their complexity, it increases also the needed communication between the various subsystems and the ECUs; in fact, they have to communicate with each other to permit vehicle automation. The highest level of ADAS must be able not only to detect, but also model and perceive the surrounding environment. Consequently, they require greater collaboration between the central decision taking ECUs and individual subsystems. For example, an ACC system may use cameras and radar to detect headway distance, and local processing to transform this raw data into object level data before communicating with the central ECU that governs steering and braking.
Two main kinds of architecture are today the mostly used: centralized and distributed.

Centralized architectures are strongly diffused in vehicles that offer a high level of autonomy as Tesla and in general the premium sector. These vehicles integrate high performance computing platform into the central ECU to permit the processing of thousands of data in few second, needed for the functioning of the complex system of automation. The main disadvantage of this architecture is the cost, starting from the cost of the material itself, that doesn't permit the diffusion of this systems to mid-range vehicles, to the needed of adoption of a performance communication infrastructure.

Distributed architecture is diffuse due to the mid-range vehicles, in order to permit simpler form of automation at a cheaper price. This alternative approach involves distributing some or all of the intelligence to the sensor nodes to facilitate local processing of raw sensor data. This results in a hybrid or fully distributed system architecture. For example, a camera could perform its own object detection locally, only sending the relevant information to the central computer for further processing and decision making. So the main disadvantage respect to the centralized architecture is the level of performance they can face, but at the same time they permit the diffusion of these technologies, that brings to have more investments on it and so in the future the diffusion also of the most complex application not only for the premium users.

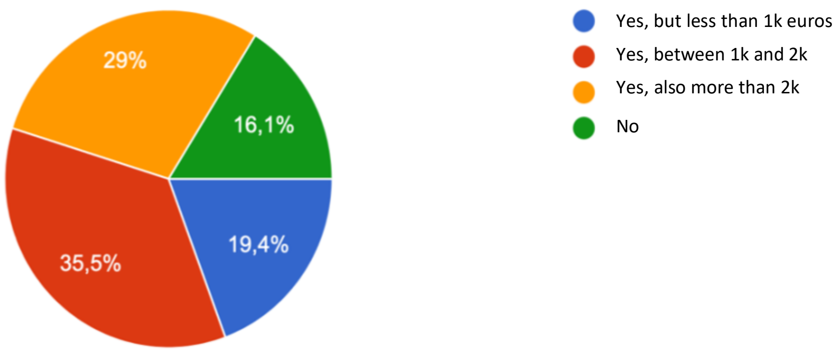
Making a local interview, on a data set of about 50 people, about ADAS we can observe different things.
First thing we can observe is that cars produced from 2018 have all at least one simple ADAS, both if they are cheap cars, so cost less than 20.000 euros, or popular and premium cars. Premium car will have obviously a more complete package of ADAS than cheap ones that usually have only simple cruise control or the blind spot monitor.
For cars from 2015 to 2018 the presence or not of ADAS depend from car to car, it mainly depend on the cost of the car and on the car outfitting and optional.

On cars produce before 2015 it’s quite difficult to find the presence of ADAS, only on the most luxury cars of the premium brands it’s possible to find the first one that were experienced.
So, on my dataset, I can observe that the 58% states that his/her car has at list one ADAS, the 39% don’t have and the 3% don’t know.

Between the percentage of people whose car has ADAS I can observe that the most diffuse of them it’s normal or adaptive cruise control, present for the 76% of cases, followed by automatic emergency brake with the 52% and from the line keeping assist system with the 38%. Traffic sign recognition is present only in the 24% of cases, while the automatic parking is the less diffuse with the 9%.

The presence of ADAS in the car doesn’t mean that used by the owners. In fact the 39%, despite have ADAS, don’t use any of them with regularity. Cruise control is not only the most diffused, but also the one used with more frequency.

Another important observation that can be made, it’s the willingness to pay of the interweaved. It can be observed that the majority, the 84%, is willing to pay more money to have ADAS in their car, in particular an average between 1.000 and 2.000 euros is the most popular amount.



The last observation that can be made is about future developments of this technologies, in particular on advanced driving system and on the possible advent of the first kind of autonomous driving. The 42% of interweaved wouldn’t trust on advanced systems that brake or steer independently.

2.3 TRAFFIC SIGN RECOGNITION

Traffic sign recognition is one of the most diffuse kinds of ADAS. It's a level 0 system because the driver remains in total control of all the function of the car, and he only receives warnings. In fact, the main porpoise of this system is filling the gap of inattention of driver due to many different situations.

Modern cars typically use a combination of camera-based traffic sign recognition and map input to provide highly accurate speed limit information. In the case of variable speed limit situations, the camera-detect speed limit takes priority of the information supplied by the map material.
Thanks to the developing of ADAS today speed limits are not the only kind of traffic sign that cars are able to recognize. Artificial intelligence in fact permits cars to recognize no entry/wrong way, no passing, yield and stop sign and last development are also able to understand the colour of traffic lights.
Cameras are the most use and diffuse source of input information for the functioning of traffic sign recognition, but recent studies propose also new sensor that are hybrid between cameras and lidar sensor to improve overall system accuracy.

The diffusion of T.S.R increases in the late 2000s, but it is developing start in the mid 1990s. the first functioning method that was take as standard was based on two steps: first the detection of circular and triangular sign by performing a geometrical analysis of edges, then the sign recognition step performs by comparing the normalized signs, after pre-processing them to allow comparison, with imagine template from a database.
Nowadays there are two main approaches: the hand crafter feature approach and the deep learning approach.

The hand crafter feature approach usually involves a first transformation of the detect signs into a manually chosen feature vector, in order to recognize the sign's attributes; secondly the training of a classification algorithm, such as Random Forest or Support Vector Machine, using the output of the feature vector.

.

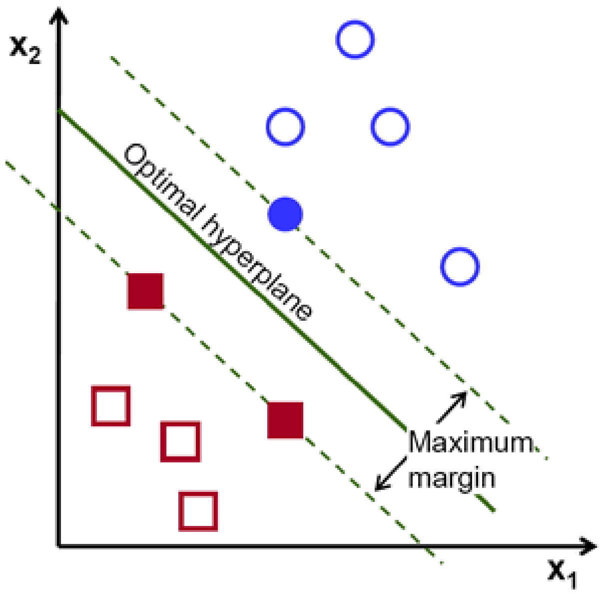
To understand better this first method let's firstly defined "hand crafted" as the features that refer to properties coming from the use of different algorithms that extract the information present in the image itself. For example, it can understand the shape of a sign, so the presence of edges and corners. An imagine is simply a 2D matrix, so an edge detector algorithm works by finding areas where the matrix suddenly changes.
Let's then understand better the second part of the process. A classification algorithm is one of the supervised learning techniques. Given a dataset or some observations, the algorithm learn from the given data and classify them in classes called labels or categories. The algorithm that implements the classification work

on the dataset is called classifier and it can be of two kinds: binary, if the classification has only two possible outcome, or multi-class if there more possible outcome. The classifier algorithm for traffic sign is a multi-class one.
Support Vector Machines are a kind of classification algorithms. The objective of them is to find a hyperplane in an-dimensional space that distinctly classifies the data point. The hyperplane is the plane that has the maximum margin, that is the maximum distance between data point of all the recognize labels. Support vector are data points that are closer to the hyperplane and influence the position of it and are used to maximise the margin of classifier. SVM is a good functioning method for both linearly separable and not linearly separable data, using the Kernel Trick.

SVM are highly performing for traffic sign classification. A Hong Kong' research shows it accuracy; in fact, on a data set of 350 traffic sign shapes and 250 speed limit sign SVM have achieved 100% accuracy on sign shapes classification and 99% accuracy on speed limit signs classification. The performance of SVM model highly depends on the choice of model parameters.

Random forest is another kind of classification algorithm. It's based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. It works by building a large number of Classification And Regression trees and to make prediction for a new observation each tree votes an outcome, and the most voted outcome is take as the final one.

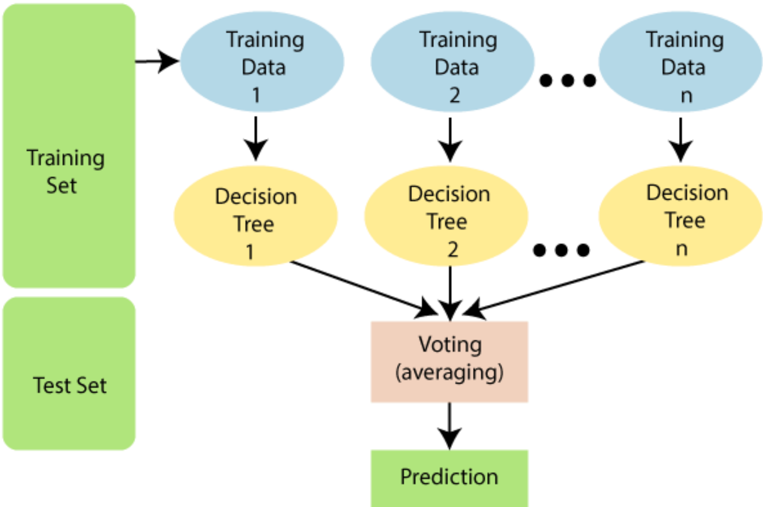
Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output.



The functioning of Random Forest is based on four steps: first a random k data points are picked from the training set, then the decision tree associated to these trees are builted. After deciding the number of trees that have to be built the first two steps are repeated. For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

Random Forests have received increasing interest because they can be more accurate and robust to noise than single classifiers. Another advantage of Random Forests is their ease of use in the sense that they have only two parameters, the number of variables in the random subset at each node and the number of trees in the forest.

The second main approach is the deep learning one, that doesn't require any hand crafter feature. The system is trained to build its own internal representation, in order to be able to output and accurate sign classification based on the training data. This is typically done by training a deep neural network with an extremely high amount of data.



Deep Learning is a type of Machine Learning, inspired by the structure of a human brain. Its algorithms attempt to draw similar conclusions as humans would by continually analyzing data with a given logical structure. To achieve this, deep learning uses a multi-layered structure of algorithms called neural networks. The typical neural network architecture consists of several layers. We call the first layer as the input layer, the central layers as the hidden layer and the last layer as the output one.
We can divide the learning process in two important phases: the "going" of information called forwardpropagation and the "return" called backpropagation. The first phase forwardpropagation occurs when the network is exposed to the training data and these cross the entire neural network. This means passing the input data through the network in such a way that all the neurons apply their transformation to the information they receive from the neurons of the previous layer, through the activation functions, and sending it to the neurons of the next layer. Next, a loss function is used to estimate the loss, or error, and to compare and measure how good/bad the prediction.
Once the loss has been calculated, this information is propagated backwards. Starting from the output layer, that loss information propagates to all the neurons in the hidden layer that contribute directly to the output. However, the neurons of the hidden layer only receive a fraction of the total signal of the loss, based on the relative contribution that each neuron has contributed to the original output. This process is repeated layer by layer. Now that we have spread this information back, we can adjust the weights of connections between neurons. What we are doing is

making the loss as close as possible to zero the next time we go back to using the network for a prediction.
The deep learning approach has the advantage of avoiding the difficult task of feature engineering and has high rate of accuracy, in fact today it’s the most used approach between the two. However, it’s more computationally expensive and require a lot more training data. Fortunately, there are several public traffic sign datasets avaible that were collected from various countries such as Germany, Sweden and the United States.

The performance ADAS traffic sign recognition has improved significantly over the last decade, there still some challenges to face for both manufactures and tech suppliers. For example, there can be many local variations on a given sign type or class. More sign variations mean more data is needed to train the chosen recognition system and more effort required to collect the data, as public dataset might be limited. Also, even the number of recognizable sign type is increasing, there is generally still quite a limited set that can be recognized until now.

2.4 DRIVER MONITORING SYSTEM

One of the main goals of AI in the automotive sector is helping prevent accidents caused by driver inattention or bad decisions, in fact human factors are still the main reason for 90% of traffic accidents. Distraction is defined as any activity that takes drivers’ attention away from the driving task.
One way in which AI helps reduce distractions is through intelligence user interaction concepts in the vehicle’s infotainment system, that allow vehicle operation without the driver to take their attention off the road.
Another way is to monitor the driver’s state and to warn them upon sign of sleepiness or distraction. Sleepiness is described as a state of being awake, but with increased tendency to fall asleep.

A DMS based on visual appearance needs to cover potential applications such as Fatigue Analysis, Distraction Detection or Behavior Analysis. These high-level features can only be provided by building a chain of lower-level modules. Low-level functions read images and detect the face or the eyes, while mid-level functions analyze specific features of the detected elements, such as gaze and eyelid aperture. Based on this information, the system classifies the driver visual attention in three categories: on-road, off-road or mirrors/instruments cluster. If driver attention is not on the road for too long or too much frequently, the system will trigger a warning.

Gaze estimation is one of the fundamental steps. Its algorithms can be categorized into two new approaches: geometry-based and appearance-based.
Geometry-based, also called model-based approach, estimates the point of gaze using geometric calculation based on a constructed 3D model of the face or the eyes. Depending on the algorithm used, the model might be constructed the estimated head pose usually in conjunction with estimated position of relevant facial landmarks such as pupil center, eyeball center or mouth.
The appearance-based method uses machine learning to determine the point of gaze based on direct eye or face imagine as input. Visual descriptor, local binary patterns or multi-scale histogram of oriented gradients are extracted from each imagine and processed through a machine learning regressor algorithms, for numerical output, or classifier algorithms, for category output.
Visual descriptor or image descriptors are descriptions of the visual features of the content in images. They are the first step to find out the connection between pixels in a digital image. They describe elementary characteristics such as the shape, the color and the texture.
Local binary pattern is a kind of visual descriptor used for the property of high discrimination power, that is the degree of accuracy with which a predictor is able to differentiates outcomes into categories. LBP labels each pixel in an image by comparing the gray level with the neighboring pixels and then assigning a binary number.
Histogram of oriented gradients is another type of visual descriptor. The HOG focuses on the structure or the shape on element detected in the imagine. This technique counts occurrences of gradient orientation in the localized portion of image, in fact the essential thought behind it, it is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. The image is divided into small, connected regions called cells, and for the pixels within each cell, a histogram of gradient directions is compiled.
One big advantage of the appearance-base method is that it is less demanding that its geometrical counterpart, since it doesn't require high resolution input images to perform the gaze estimation.

Gaze estimation employs convolutional neural network and other deep learning methods both to replace the manual feature extraction step, or to act as end-to-end system. They take direct eye or face images as input, and then outputs directly the estimated gaze direction without any human pre-defined intermediate step such as the feature extraction.

Before the eyes can be monitored the system has to perform several pre-processing steps such as the face detection and the eyes detections. It's important to remember that the expressions face detection and face recognition mean two different things. Face detection indicates the recognition of a face in a imagine, while face recognition is about identifying the person by matching the detection with a face database.

One of the most popular algorithms for face detection is the Viola-Jones algorithm. This algorithm can detect faces very fast, despite it is painfully slow to train, but perform reliably on frontal face images only. Viola Jones algorithm is named after two computer vision researchers who proposed the method in 2001, Paul Viola and Michael Jones. Given an image, the algorithm works on grayscale image, the algorithm looks at many smaller subregions and tries to find a face by looking for specific features in each subregion. It needs to check many different positions and scales because an image can contain many faces of various sizes.
The Viola Jones algorithm has four main steps:

I. Selecting Haar-like features: they are digital image features used in object recognition. All human faces share some universal properties of the human face like the eye's region is darker than its neighbor pixels, and the nose region is brighter than the eye region. A simple way to find out which region is lighter or darker is to sum up the pixel values of both regions and compare them. The sum of pixel values in the darker region will be smaller than the sum of pixels in the lighter region.

II. Creating an integral image: also known as a summed-area table, it is the name of both a data structure and an algorithm used to obtain this data structure. It is used as a quick and efficient way to calculate the sum of pixel values in an image or rectangular part of an image.

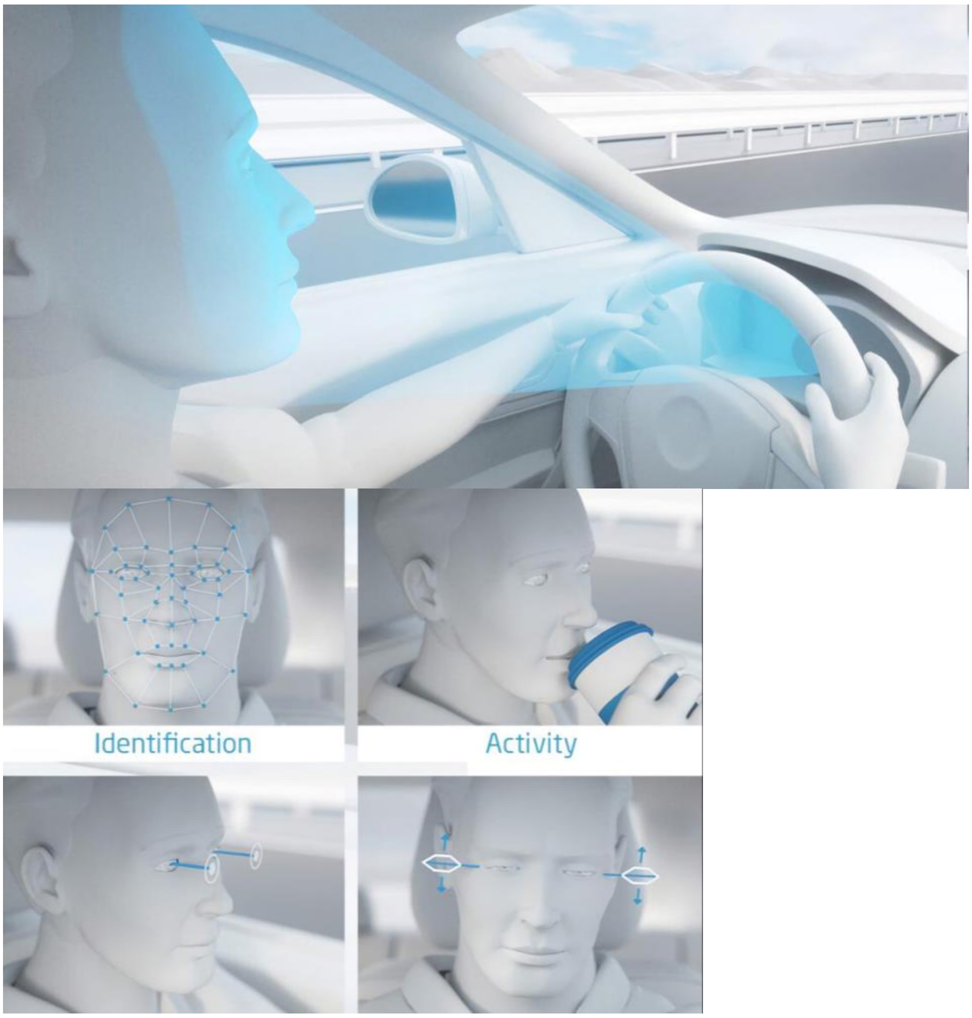
III. Running AdaBoost training: it's another kind of algorithm. In an image the features detected can be thousands. So, when the AdaBoost is trained to identify important features, information are fed in the form of training data and subsequently it is trained to learn from the information to predict. Ultimately, the algorithm is setting a minimum threshold to determine whether something can be classified as a useful feature or not.

IV. Creating classifier cascades: The job of the cascade is to quickly discard non-faces and avoid wasting precious time and computations. The cascade system divides the process of identifying a face into multiple stages. In the first stage, the subregion passes through the best features, in the next stages the are the remaining features. When an image subregion enters the cascade, it is evaluated by the first stage. If that stage evaluates the subregion as positive, meaning that it thinks it's a face, the output of the stage is maybe.
When a subregion gets a maybe, it is sent to the next stage of the cascade and the process continues as such till we reach the last stage.
If all classifiers approve the image, it is finally classified as a human face.

After the face region is detected, the next stop is to localize the eye region in the face. Eye detection algorithm can be categorized into four main approaches: template matching, feature based, appearance based and hybrid.

The template matching approach compares any template with all areas in the face image and return the area that best match the template.
The feature-based approach looks for eye features, such as color and shape throughout the image.
Appearance based method uses machine learning to classify if a certain region in the image is an eye or not.
The hybrid technique is a combination of the other ones.

Although the AI realms of face detection, gaze estimation and emotion recognition have been studied for years, their real-world application as DMS in cars are still in their infancy. In order to be useful DMS has to perform reliable under many challenging real-driving situations such as a dark cabin environment or drivers wearing glasses per example. Privacy might be an additional issue, considering that the most inattention driver detection techniques use camera as sensor and some non-safety-critical processing might need to be performed outside the vehicle.



2.5 FUTURE DEVELOPMENT: HYUNDAI'S ADAPTIVE CRUISE CONTROL

Hyundai's technology integrates artificial intelligence into the company's Advanced Driver Assistance System. The system, called Machine Learning based Smart Cruise Control (SCC-ML), will go into series production already in Hyundai's next generation of vehicles.

SCC-ML combines AI and SCC into a system that analyses and learns the driver's driving behaviour. Through this machine learning, the new technology independently regulates the speed according to the same pattern as the driver.

The system works as follows: Sensors such as the front camera and radar constantly record driving information and send it to a central computer. This computer extracts the relevant details from the collected information in order to identify the driver's typical driving patterns. This process of machine learning is controlled by AI. The system is programmed in such a way that it recognizes unsafe driving patterns and does not adopt them.

With currently available conventional cruise control systems, the driver manually adjusts the distance to the vehicle in front and the speed to be driven. Without machine learning, it is not yet possible to meticulously adapt the settings required for autonomous driving to the individual preferences of the driver.

The algorithm used captures three aspects of the manual driving pattern: distance to vehicles in front, how strongly the driver accelerates and how quickly he reacts to changes in the driving situation. External driving conditions and current speed requirements are also taken into account.

The system detects whether, for example, the driver is driving slowly in the city or in the overtaking lane of the motorway. When analyzing the data, the system can differentiate between more than 10,000 different patterns and thus flexibly adapt to the driving style of each driver.

The system is so sensitive that it detects changes in the driver's driving style and regularly adapts to them. In addition, SCC-ML is programmed in such a way that it does not learn any unsafe driving patterns and thus increases both reliability and safety.

3.0 AUTONOMUS DRIVING

3.1 DEFINITION AND DIFFUSION

The basic goal of transportation is to move people and goods from A to B, and one key aspect of that is to ensure that the people and the good reach their destination not just quickly but safely. Despite the introduction of several new technologies that helps the drivers, statistics show that this is not enough. In fact, they show that about 94% of serious car crash are caused by human error, highlighting the importance of greater vehicles autonomy, removing human error from the equation.
In addition to safety, higher level of automation allows higher productivity and reduce level of stress, moreover it can help also to reduce the number of vehicles on the streets.

Autonomous driving AD is a good case study of applied artificial intelligence. Many sensor modalities, including cameras or radars, are employed to achieve AD. Data from each of these sensors is processed and understood by both machine learning and artificial intelligence.

The term self-driving is often used interchangeably with autonomous. However, it’s a slightly different thing. A self-driving car can drive itself in some or even all situations, but a human passenger must always be present and ready to take control. Self-driving cars would fall under Level 3 (conditional driving automation) or Level 4 (high driving automation), that are the ones whose basic functioning will be analysed in this document.

KPMG did an extensive study about self-driving cars diffusion and adoption, in which they interviewed 25 industry leaders and experts about their expectations. Although they agreed that they expect an S-curve type of adoption for self-driving cars, as all the new technologies introduced in the market, three different scenarios of adoption were generally given:

1. The first scenario is the one of aggressive adoption. This is a scenario envisioned by among others the CEO of Tesla, Elon Musk. For an aggressive adoption scenario, it’s important that the systems are sufficiently optimized before launching and that the suppliers can make people see the benefits of self-driving cars. The popularity of the system will cause even greater technological developments and will increase the vehicle to environment abilities. This will then make more and more people adopt the system, reaching the critical mass in less than five years.
2. A less aggressive scenario of adoption is the base case scenario. In this case, the systems are sufficiently optimized before launch, and consumers embrace the benefits of self-driving cars. However, it is not immediately

embraced by the proper authorities, causing a delay of adoption. In this scenario, it is however not necessarily assumed that self-driving cars will reach 100 percent adoption, but it will reach the 75% by 2040.

3. Finally, there is the conservative scenario. In this case, because the systems fail to deliver on the initial promises, the greater mass of the consumers will not adopt the system. Eventually, there will be a small rise in adoption as the development increases, but it will never reach the general populace.

Today data are showing that the autonomous vehicles industry rises by 16% globally every year. While there was an estimated total of around 31.4 million autonomous cars globally in 2019, this number should increase to some 54.2 million in 2024.

There's not a single approach to AD. Actually, the most diffused it's the modular approach based on three key building blocks: sensing, planning and acting. In this kind of approach, a pipeline of separate components is employed, linking sensory inputs to actuator outputs.

3.2 PERCEPTION

Perceiving the environment is crucial to assign a traffic situation. For this porpoise autonomous vehicles Avs employs various sensor modalities such as cameras, radars, lidar sensor and ultrasonic sensor. Avs not only employs sensor, but also the right applications of AI to understand data coming from these multiple sensors.
AD system needs to detect, classify and characterized 3D information concerning road users, road boundaries, drivable space, traffic light, traffic sign and other any element that is relevant to driving safely. In high-speed scenarios is important to identify any hazards and threats as soon as possible in the shortest possible time because of high speed implies greater distance travelled before the driver can intervene. So AD perception algorithms need to run fast because they are part of a system where every millisecond count.
Considering the large amount of information that comes and the different sensor that have to be menage at the same time, it's easy to understand that a high-performance ECU is needed. AI algorithms might need to find the proper trade-off between accuracy and available computing resource. Greater performance means more safety.

Researcher and developers have proposed different perception methods which have different particularities for each sensor modalities. In the most recent solutions several sensor modalities are often combined together. Currently the most diffused methods are the ones that use deep learning such as CNNs. These

algorithms that are classified as supervised algorithms, so they're trained with millions of examples.
The functioning of CNN is very interesting because its architecture is analogous to that of the connectivity pattern of Neurons in the Human Brain. In simple word what CNN does is to extract the feature of image and convert it into lower dimension without losing its characteristics. It makes this through different steps:

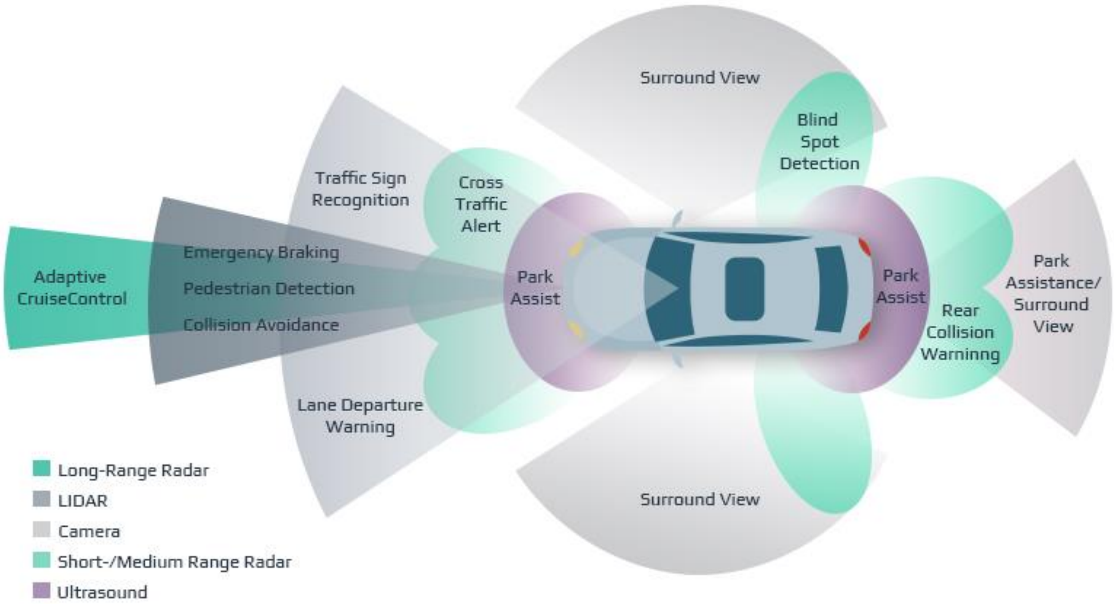
1. The first one is the convolutional operation. Every image can be considered as a matrix of pixel values. Convolution preserves the spatial relationship between pixels by learning image features using small squares of input data. To do this it uses a filtering matrix.
2. A filter extracts features from an image. The products between the filter values and the image pixel values forms the convolution layer. The filter is a matrix of weights. The values in the filter matrix are updated each time the network performs backpropagation. In CNN, there are many different filters which extract various features in an image. As it progresses through the network, the features extracted from an image get more and more specific. A filter in the first layer is used to detect a feature that can be easily recognized, a filter in the second layer would be used to detect something more specific.
3. An additional operation called ReLU has been used after every Convolution operation. The purpose of ReLU is to introduce non-linearity in our CNN, since most of the real-world data are non-linear. ReLU is an element wise operation that replaces all negative pixel values with zero in the feature map.
4. The following step is the pooling step. The pooling operation reduces the dimensionality of each feature map but retains the most important information. The aim of the pooling operation is to reduce the size by getting rid of information not important for feature.
5. The next operation is the flattering one. This step literally goes to flatten the pooled feature map into a column. The reason it does this is because of the necessity of insert this data into an artificial neural network.
6. The last step is the full connection operation. This is the step where artificial neural networks and convolutional neural networks collide. The input layer contains the vector of data that was created in the flattening step. The role of the artificial neural network is to take this data and combine the features into a wider variety of attributes that make the convolutional network more capable of classifying images. By the end of this channel, the neural network issues its predictions, considering also an error factor.

In general, classical approaches use different methodologies for the different kind of sensor.
For cameras the approaches can be divided in object based and pixel based.
Object based methods are used to detect and classify object that could be encapsuled in a 2D bounding box as vehicles, cyclist or pedestrian. In contrast pixel

based ones use imagines segmentation techniques for assigning a class to each pixel. They are used for elements that can't be encapsuled as a road surface.
Object base methods usually use single-shot detector algorithms and region-based algorithms.
Single-shot detector has two components: a backbone model and SSD head. Backbone model usually is a pre-trained image classification network as a feature extractor. The SSD head is just one or more convolutional layers added to this backbone.
Region based algorithms, or two-stage methods, first detect region of interest and then apply classification on them.
Region based produce better object recognition and localization, while single-shot detector is faster and so more used in real-time application.

3D information collection and reconstruction can be performed both by a stereo camera system or by the lidar sensor, that it's more accurate.
Lidar's basic principle it's to measure the time it takes for a light beam to travel from its source, bounce back from a surface and return to the source. The sensor can transmit hundred of thousand of such lights beams per second into predefined directions, their reflection produces a cloud of 3D points that indicate spatial positions. Since it's assumed that each 3D point belongs to an object in the scene, after identification each point need also to be clustered, so assigned to the correct object.

Detection and classification are two important building blocks, but perception extends beyond that: perceiving the environment also includes the use of relevant algorithms for object tracking and multi-sensor data fusion to allow proper understanding of surrounding conditions.



3.3 PLANNING

Once the vehicle has identified and understood all the surrounding elements it's time to plan how to act. Planning is a complex task that involves different layers: global planning means find the best route from A to B following the traffic rules, behavior planning means estimating the most likely behavior of all the external actors in the scene and local planning means generate a continuous trajectory in the sense of planning immediate actions to bring the vehicle to next desired position.
So once that Avs have perceived its surroundings, the vehicle needs to contemplate multiple hypotheses and choose the one most suitable to bring the car to its final destination, as safety and quickly as possible. To achieve this goal the vehicle needs to localize itself in a given map in order to calculate the best route, and at the same time it needs to predict likely behavior from other actors and compute the most immediate trajectory to follow.
We assume that the final destination it's given by the combination of on-board global navigation satellite systems and sophisticated localization algorithms, that are used to localize and associate static objects with previously mapped objects. Localization is fundamental step, in fact sometimes it's even considered as an independent step before planning.
There are many different techniques to help an Avs to locate themselves:

1. Kalman filter
2. Particle Filter
3. SLAM

1- A Kalman filter can be used for data fusion to estimate the state of a dynamic system in the present (filtering), the past (smoothing) or the future (prediction). Sensors embedded in autonomous vehicles emit measures that are sometimes incomplete and noisy. The inaccuracy of the sensors, also called noise is a very important problem and can be handled by the Kalman filters.
The state vector is composed of a position p and a velocity v. At each estimate, we associate a measure of uncertainty P. Using the uncertainty is great, because we can consider the fact that Lidars are more accurate than RADARs in our measurements.
A Kalman filter is an implementation of a Bayesian filter, an alternation between prediction and update over and over again.
To make an estimate a Kalman Filter only needs the current observations and the previous prediction. The measurement history is not necessary.
The mathematic behind the Kalman filters is made of additions and multiplications between the prediction matrices and the update matrices.

2- A Particle Filter is another implementation of the Bayes Filter.

In a Particle Filter, we create particles throughout the area defined by the GPS and we assign a weight to each particle. The weight of a particle represents the probability that our vehicle is at the location of the particle.

3- SLAM, simultaneous localization and mapping, is a method used for autonomous vehicles that lets you build a map and localize your vehicle in that map at the same time. SLAM algorithms allow the vehicle to map out unknown environments.
There are two types of technology components used to achieve SLAM. The first type is sensor signal processing, including the front-end processing, which is largely dependent on the sensors used. The second type is pose-graph optimization, including the back-end processing, which is sensor-agnostic.
The front-end processing component can be of two kind: visual SLAM or lidar SLAM.

Visual SLAM uses images acquired from cameras and other image sensors. It can be implemented at low cost with relatively inexpensive cameras. In addition, since cameras provide a large volume of information, they can be used to detect landmarks. Visual SLAM algorithms can be classified into two categories: sparse methods match feature points of images use algorithms such as PTAM, Parallel Tracking and Mapping algorithm; Dense methods use the overall brightness of images and use algorithms such as DTAM, Dense Tracking and Mapping algorithm.

Lidar SLAM is a method based obviously on the use of lidar and laser sensors. The laser sensor point cloud provides high-precision distance measurements and works very effectively for map construction with SLAM. Generally, movement is estimated sequentially by matching the point clouds. For lidar point cloud matching, registration algorithms such as iterative closest point (ICP) and normal distributions transform (NDT) algorithms are used.

Once's that the localization step is completed the best route is calculated, given different parameters such as traffic rules, a prior traffic knowledge, current traffic conditions, distance and many others.
The relevant information about streets are part of the existing maps, that are usually represented in the form of graph while the information about them are available through backend server communication as Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure.

So, maps are considered as graph formed by nodes and edges, therefore in order to plan a route a graph has to be traversed, that means evaluate its information, to find the best path for the desired goal, that can be for example the fastest or the shortest in term of distance. To do so several graph-traverse algorithms have been used. The most diffused kind is the so-called Best-First Search algorithm, which consist of moving through the graph by following the most promising nodes

given certain predefined rules. Good algorithm of this kind are the Dijkstra and A* ones. They both look for the path that minimizes a certain weight or cost between two nodes, that represent the current and final position. Of these two algorithm A* is more accurate and efficient.
In order to make these methods functioning on large-scale maps incremental graph are usually employed. Methods like A* relay on some predefined statements, whereas some more recent approaches use AI to optimize the selection of such rules or even search new ways to approach route planning with the use of Recurrent Neural Network. They are obviously a class of artificial neural networks where connections between nodes can create a cycle, allowing output from some nodes to affect subsequent input to the same nodes.

About behavior planning, there is extensive research analyzing other road user possible behaviors. Human driving behavior can bel learned by imitating real human driving experience data. But there are some corner cases, such as incidents or dangerous situation where data can't be collected or there are few available. Simulated data can be used to train such situations using Deep Reinforcement Learning.
DRL is the combination of two AI techniques: reinforcement learning and deep learning. Its goal is to multiple layers of Artificial Neural Networks that are present in an architecture to replicate the working of a human brain. There are two main types of Reinforcement Learning algorithms:

- Model-based algorithms
- Model-free algorithms

Model-based algorithm use the transition and reward function to estimate the optimal policy. Model-free algorithms find the optimal policy with very limited knowledge of the dynamics of the environment. They don't have any transition/reward function to judge the best policy. They estimate the optimal policy directly from experience.

To plan the best trajectory implementation are based in construing a curve representation for a set of possible way point. This can be done using several different curve construction models, as polynomials, splines or clothoids, that are useful for generating smooth trajectories. Other implementations use random sampling or optimization of the possible navigation area as for example the Rapidly exploring Random Tree that is a random-sampling trajectory planning algorithm. Optimization-based methods can produce optimal trajectories, but their main disadvantage is that they need high computational processes.
The current trend in automotive are focused on learning-based approach for example for planning a optimal lane-change trajectory. Other new methods focus on the spatial temporal features of neural networks.

3.4 MOTION CONTROL

The next step is to convert the plan into executable actions. The right signals must be sent to the various actuators, such as the breaking or the steering system, engine or transmission. Each of this actuator may need a different kind of input, in fact the main goal of the motion control is to map the intended action into signals that the actuators can understand.
Moreover, the motion control relies on real time knowledge of vehicle state information for which it’s necessary to estimate the right vehicle dynamics model. The used techniques can be divided into two groups: model based, which are variation of the Kalman filter and Particle filter, they optimize the measurement of the vehicle when there is measurement error; data-driven-base which estimate vehicle state and parameters through learning from historical and real-time data thanks to the use of neural networks.

Once the proper vehicle dynamic parameters are collected, motion control can send the commands in order to keep the vehicle as close as possible to the optimal trajectory. This can made for example through Proportional-Integral-Derivate (PID) controllers. They use a control loop feedback mechanism to control process variables and are the most accurate and stable controller. The working principle behind a PID controller is that the proportional, integral and derivative terms must be individually adjusted or "tuned." Based on the difference between these values a correction factor is calculated and applied to the input.

There is a growing trend toward the employing of learning-controllers, which make use of training data. Learning controllers are capable of anticipating reputing vehicles motion effect and disturbances, incrementing the overall safety and comfort of the system. They use CNNs as the Iterative Learning Control for path tracking, or Model Predictive Control which computes control action by solving an optimization problem.

4.0 THE FUTURE OF AI IN CARS

The future of autonomous driving will strictly follow the development of AI. In fact, already nowadays the dependence for the functioning of the overall system from AI is very strong.
The landscape of partnership and collaboration between the automotive Original Equipment Manufacturers, robotic start-up and high-tech software and hardware provider is continuously evolving, so it's difficult to predict if car's manufacture will in the future develop internally the AI solutions or they will keep collaborating with external partners. Moreover, the complexity and cost of the autonomous driving endeavour might produce significant changes in the landscape of manufactures through merges, acquisition and bankruptcies.

In general, as all the new technologies, autonomous vehicle technology will follow an S-curve development pattern. Autonomous vehicles are currently in development and testing stages. They will need significantly more technical improvements, real-world testing, regulatory approval and commercial development in order to become reliable, widely available and affordable, and therefore common in the vehicle fleet.
Many current vehicles have Level 2 and 3 technologies such as cruise control, hazard warning and automated parallel parking. Several companies are testing Level 4 vehicles in limited conditions. Although some vehicles can now achieve up to 95% operability, meaning that vehicles can reach 95% of desired destination, achieving higher rates is increasingly difficult.

Optimists predict that most vehicles will operate autonomously by 2030, but most of them have financial interests in autonomous vehicle industries and base their predictions on experience with electronic technologies rather than motor vehicle innovations. Most objective experts predict that Level 5 automation will require many more years for development and testing.

The future diffusion of Avs will bring to some implication as a need of a new roadway design with the improvement of lane marking, road sign or the need of wireless repeaters in tunnels to provide internet access. For similar reasons also parking areas may need to be redesigned. Moreover, there are potential conflicts between user and community goals in autonomous vehicle design and programming. For example, if programmed to maximize passenger comfort they may reduce traffic speeds, and if programmed to protect occupants they may increase crash risk to other road users

Development in privacy and cybersecurity are among the hot-button issue to solve because of vehicles are getting more and more connected to different external systems and actors, which increase the risk of cyberattack through the various communication protocols. All the communication are possible access point for cyberattack due to the vulnerability caused by the connection to external counterparts, as smartphone. But the vehicle can also suffer cyberattack that are channelled through internal systems. In fact, any system that is hardwired to the driving ECU could be a potential conduit for attack. Considering also the environment outside the vehicle itself, all the surrounding digital ecosystem, as the smartphone application for the distance control of the car, is a point of vulnerability.
All these vulnerabilities across the entire ecosystem require a cybersecurity strategy for the entire lifecycle of a vehicle. For the porpoise regulations and standard are started to being defined thanks to the UN Economic Commission for Europe.

5.0 CONCLUSIONS

The present work has shown how AI and ML are fundamental for the functioning of ADAS and consequently of Autonomous Driving.

Starting from the general functioning of them, the precise actions that permit their functioning has been analysed explaining the main concept and processes of AI and ML behind them.

Thanks also to a local interview it can be observed also that the diffusion of these technologies is not so capillary yet and the trust about them is not so high yet. In particular the common and simplest ones, as cruise control, start to be quite diffused, but the most technologies-advanced ones that will bring to a real autonomous driving are predicted to be diffused in 10 years.
Moreover, new kind of regulations are need to cope with future scenarios that the advent of this technologies will bring.

BIBLIOGRAPHY

The road to AI-enhanced ADAS | Cambridge Consultants AI in the driving seat - Cambridge Consultants:
https://www.cambridgeconsultants.com/insights/infographic/road-ai-enhanced-adas

AI IN THE DRIVING SEAT - Cambridge Consultants:
https://www.cambridgeconsultants.com/sites/default/files/uploaded-pdfs/AI%20in%20the%20driving%20seat%20-%20Cambridge%20Consultants.pdf

Random Forest – Medium:
https://medium.com/analytics-vidhya/random-forest-4a2981aab4f7

Traffic sign detection and recognition based on random forests:
https://hi.booksc.me/book/52436211/ff5b7a

deep-learning-basics-1dDeep Learning Basics. Basic Concepts for Deep Reinforcement:
https://towardsdatascience.com/deep-learning-basics-1d26923cc24a/

Object Detection Using Haar Cascade Classifier: https://github.com/mansi75/Object-Detection-Using-Haar-Cascade-Classifier

viola-jones-algorithm. How to Detect Face Recognition using Viola Jones Algorithm:
https://www.mygreatlearning.com/blog/viola-jones-algorithm/

Image classification using Convolution Neural Networks (CNN):
https://www.skillsire.com/read-blog/136_image-classification-using-convolution-neural-?mode=night

SLAM (Simultaneous Localization and Mapping) – MathWorks:
https://www.mathworks.com/discovery/slam.html/

How Does a PID Controller Work? - Omega Engineering:
https://www.omega.com/en-us/resources/how-does-a-pid-controller-work