Deep Learning-based Emotion Recognition for Human-Vehicle Interaction in Autonomous Vehicles

By Dr. Sudarshan Bhattacharyya

Professor of Computer Science, Indian Institute of Technology Kharagpur (IIT Kharagpur)

1. Introduction

[1] The study of emotions plays an important role in scenarios where the interaction with human beings is involved. By studying human emotions, it becomes possible to make predictions about decision-making related to driving. Based on previous studies, which were based on the use of abstract artifacts in interaction, such as visual maps for navigation, it is worth considering the use of artificial intelligence based on facial and vocal expression for the analysis of drivers' performance and affect. Regulating interaction on driver behavior through synthetic companions (or wholly automated agents) could enhance driving pleasure and assist in the prevention of accidents. One of the key tasks linked to reinforcement of machine understanding of emotions is emotion recognition. Learned representations can help build an emotional model of the vehicle environment. In this paper, we present our solution specifically designed to make emotional detection in autonomous vehicle, to stimulate this aspect of human-vehicle interaction. We made this system a focus of interest for studying three main vehicles, passenger seats.[2] In the domain of human-robot interaction (HRI), the principle function of these models is to make persons and robots understand one another well enough to work together as a unit with a purpose. Human driver’s temper is the major issue in automobile security. Emotion technology on the part of the driver’s wheel has pinpointed the problem and made feasible its real-time tracking. When using emotional computing, such a section-based possibility tracking will allow a number of robust car-man systems to be placed within the systems of automobiles. In driver–machine interaction or human–vehicle interaction, identifying drivers and monitoring vehicle user’s emotions is a vital factor. Specifically, more effective understanding with safer traffic is possible by the more effective communication of data between the driver and artificial driver assistance systems; the communication is additionally possible in real time in exacting danger conditions as that data can change safe behavior and favor danger reduction in the sought-after data.

1.1. Background and Motivation

[3] [4]In recent years, the development of artificial intelligence has gained attention. One of the most significant implications of this advancement is the deep learning technology, which has shown its potential in various computer vision tasks, including face recognition, emotion detection from magnet resonance imaging (MRI) images, and action recognition. In fact, deep learning methods are one of the best performing tools for emotion detection from facial expression. Specifically, the use of convolutional neural networks (CNNs) and sequence models in deep learning have significantly improved the performance of facial expression recognition systems. The main interest of industry and academia in developing systems for facial expression recognition is to enable artificial intelligence to understand human emotions, since this would lead to impressive innovations in many fields. For example, IBM's Watson, one of the most advanced cognitive systems for image and speech analysis, is now able to recognize human facial expressions. This represents an important step toward the goal of simulating natural communication and interaction between machines and humans. In the context of vehicular applications, human emotion recognition through CNN-based computer vision systems has important applications in the development of safety systems which could lead to the development of new solutions and products to create a better and safer driving experience for users. The automotive industry is actually expanding its effort towards the development of these systems putting safety and comfort at the center of the attention. In particular, in the design of autonomous and advanced driver-assistance systems (ADAS) vehicles, the knowledge of passengers' level of comfort and stress is an important aspect that can affect the vehicle's driving performance as well as its man–machine interface that should be as much as possible seamless and natural. The results of ANSP (Accident Prediction and Prevention, United States study) survey show that about 21% of accidents are caused by human emotions—apathy, irritation and anxiety are the main causes of accidents due to distraction or slow reaction time. This study demonstrates that the main goal of emotion recognition system in autonomous vehicle is to foresee possible passengers' distraction or forthcoming fatigue during the trip.[5]Regardless of the source, emotion has many different effects on the driver, such as reducing attention, decision making, and decision making ergonomics. For this reason, monitoring the driver's emotional status can be one of the key aspects that can be very useful for advanced applications, such as advanced human–machine interfaces, advanced driver assistance algorithms, infotainment application, e-learning-based

applications (using game theory or cognitive mapping of emotional and cool (unemotional) drivers), and autonomy (e.g., know-how and make the driver aware of the car's decision). In general, the problem of properly managing and interpreting emotions can bring important benefits in various applications. Driven by these facts, it is important to capture a complete record of people's emotions in different situations, especially in a car, where emotions have a strong influence on the vehicle's behavior. This study employs a convolutional neural network for image processing and reshaping, combined with a long short-term memory neural network (LSTM) for sequence signals, in order to enhance emotional patterns accuracy by intelligent processing of selected feature data. In our framework, we use transfer learning to carry out experiments on a public multimodal dataset (MMI LIRIS laboratory, France), which contains vision and physiology data.

1.2. Research Objectives

One of the most relevant innovative thinking assumptions among the various imaginative applications for the typical use time for both drivers and co-drivers is "other things". Many attention-grabbing and different ideas and new application objectives will emerge to address this expectation and to be ready to use it in practice. In fact, in an autonomous car, Because the driver is now free of responsibility, it does Not necessarily have to keep his hands focused on the steering wheel or gear level to drive [6]. Although there will actually be a need to quickly intervene in traffic or to control the vehicle, most of that time could be saved and used for other businesses. This newly emerged traffic and travel-free time that can be evaluated in many different ways, such as consuming various services and using them for entertainment, has been called 'other things'.

A significant number of research articles that have grown exponentially in the driver state monitoring area and the development of autonomous vehicles have simply studied driver emotion because it affects driver behavior [7]. Because the driver emotion must be tracked, one of the most important human-vehicle interaction (HVI) elements is the positioning of emotion recognition in autonomous vehicles (AVs). AVs that can efficiently understand and comprehend the emotional states of human passengers can build trust between them and, thereby, a safe and secure human-automobile relationship. As a result, researchers have concentrated on ways to enable affective AVs such that they can recognize human emotions by observing their facial expressions, speech, physiological markers or other behavioural patterns and adjust their actions to accommodate their emotional condition. It is fairly evident that with the proliferation and adoption of AVs, the environment around the passengers will perhaps take an even more central role – offering seamless, personalized and more focused emotional support experiences. In line with this reasoning, several recent studies have been conducted to anticipate what scenarios the future vehicle interiors will be used for by discovering new research areas around the use of autonomous vehicles.

1.3. Scope and Significance

Nowadays, autonomous vehicles and other automotive manufacturers are constantly integrating connected technologies for assisted and autonomous vehicles to offer mentilementive systems for safe driver and passenger travel. According to an announcement by Forbes contributors, the global emotion detection and recognition market size is forecasted to grow at a robust CAGR (Compound Annual Growth Rate) till 2025. For safety purposes, automotive manufacturers integrate one of the trillion Emotion AI and machine learning algorithms to contentment information and detect joy, sadness, or detect dangerous driving states in less than few milliseconds [5].

In the era of autonomous vehicles that interact with and host humans in their cabins, it is essential for service-providers, vehicle original equipment manufacturers (OEMs), and researchers to analyze human physiological and emotional responses to ensure they provide seamless and safe interaction between humans. Recent approaches of detecting and recognizing emotions from various signals include using physiological responses, including electroencephalography (EEG) and electroencephalography (ECG), voice signals, and facial expressions, all of which have shown high accuracies in detecting human emotions [8].

2. Fundamentals of Emotion Recognition

The proposed on-board human-vehicle interaction (HVI) model has proposed a new emotion recognition approach using deep learning which relies on inbuilt vehicle-camera-captured images. Most of the prior in-vehicle facial attribute recognition research used facial landmarks detection before formulating the identification model. In contrast, we proposed successful integration of the vision foundation model with machine learning to solve an automatic driver emotion recognition multi-task problem without the need for any facial landmark localization [9]. It is shown that our well-trained deep learned model on is capable of tracking facial features with a higher level of accuracy than many other publicly available facial attribute recognition/deep learned models.

The study of emotions and their impact on driving behavior and performance has gained considerable attention due to the recent trends toward the development of self-driving vehicles, aggressive control systems, and driving simulators [10]. Emotion recognition has the potential to facilitate a more advanced model of intelligent vehicle control that can ensure comfort and safety of the driver and passengers. It also has the potential to modify the vehicle interaction and services by understanding the real emotions of human beings. Emotion recognition can be performed by using various psychophysiological sensors and deep learning-based machine vision approaches. For example, recent advancements in facial expression recognition have enabled computer vision methods to recognize the real-time emotions of drivers from captured video or picture frames [11].

2.1. Psychological Theories of Emotions

A guidebook for training and teaching people who strive to live and flourish in ACs thus has been established. This leads us to our research questions: (Q1) What are the best ways for an AC to recognize emotions portrayed by human drivers and passengers? (Q2) Can the facial recognition system (FRS), developed to observe emotional signs on human drivers, be implemented through detecting these signs illustrated by human passengers? We understand that the task of recognizing emotions is more difficult in a car than in normal life because of car vibrations, noise, limited light or light fluctuations, potential differences in human facial characteristics, and human driving movements. Any techniques that can be applied offline can only be partially effective. For example, we found that drivers' facial emotions changed most often immediately after their vehicles moved or slowed down, probably because they were happy to observe rapid movements, when positive to other human beings, or were afraid of observation of a severe reduction of speed.

Emotions are critical factors in the daily lives and interpersonal relationships of humans [12]. Besides, the emotion of a human driver is essential for their driving performance. A happy driver, for instance, can stop quickly in an emergency condition compared to a bored driver. For an automated car (AC) with no driver, emotion recognition provides context for better autonomous decision-making and coordination of the AC with its human passengers. This section represents psychological theories based on this background. Cicchetti and Ackerman

(1995) analyzed facial realization patterns in three emotion categories and proposed separate systems for distinct emotions [13]. Newton-Wellesley Hospital nursing theory, training, and research confirmed "the convergence of caregiver and patient affect," "the relevance of expressions of fear to patient accommodation to an AC," and "the relevance of expressions of depression to self-adjustments of the holistic AC [14].

2.2. Techniques for Emotion Detection

An increasing field of Artificial Intelligence (AI) that has become popular by media, industry and research community only in the last 5 years is the domain of Emotion Recognition from audio signals such as speech events or musical analysis. Studies in and also address more multimodal methods combining more than one non-textual data source like speech and facial information recognition. Especially for human-computer interaction showing the visual adaptation of the personal assistance device may help to sustain cognitive and emotional aspects in the interaction. Follows that, the next section presents an overview of psychological, social and technical works that have been done to achieve the recognition of affect expressions and emotions in human-human and human-computer interaction, especially performed for different languages, cultures and contexts.

As part of the endeavor that includes the analysis and recognition of the affective state of the human partner, many approaches to emotion recognition using different features are being discussed in the literature [2]. Commonly used features related to the affective state recognition are speech, facial expressions, and multimodal recognition combining both. For a long time, there was no standardized and publicly available speech-based emotion recognition in the research community. As a result, numerous databases were developed and compared with each other [1]. With regard to the car environment, a lot of work has been done in the past 20 years on the recognition of occupancy states including recognition of the driver's fatigue, drowsy, and vigilance level. Popular methods are based on physiological signals such as heart rate, skin conductance, eye-tracking, or lane departure shocks. Emotions on the other hand have long been explored in human-computer interaction and are more relevant for naturalistic human-vehicle interaction in autonomous vehicles [7].

3. Human-Vehicle Interaction in Autonomous Vehicles

Hence, the development of an effective emotion monitoring technology for human passengers has much potential to be extended, due to a novel preference on comfort, passenger satisfaction, driving safety and so on of those kind of vehicles. However, the change of passengers may be easier and faster than the change of vehicle design in a long period service, it requires the emotional recognition and monitoring systems to be designed to be general as possible. In the process of reviewing the key articles related to affect and autonomous driving, I noticed that the interest is strongly directed towards the analysis of unimodal data. Even though there is research available based on multimodal studies, the tools were rarely adapted to AV driving contexts. So far, there is a lack of coherent knowledge on how to model various aspects of a passenger in a vehicle, including their comfort, mental state, health, travel behavior, and social interactions [15].

Emotion recognition systems in human-vehicle interaction (HVI) in the driver's seat of conventional and autonomous vehicles are interesting but challenging at the same time. In the early stages the system was mainly focusing on the driver, however, with the development of shared mobility and the trend of autonomous driving are more and more vehicles able to provide transportation services to drivers as well as to passengers. As a result, any passenger (henceforth referred to as human) can come into contact with the vehicle. In particular, there are many autonomous vehicle projects emerging now with the aim of removing the human driver from the vehicle and the driving tasks have been undertaken by the vehicle system. While these technologies evolve, it is important to maintain the balance in HVI in autonomous vehicles to ensure occupant safety and vehicle-fleet operation [2].

3.1. Challenges and Opportunities

A potential challenge in a vehicle intelligence system is identifying more than one human occupant simultaneously interacting in a visually occluded multi-physical manner in the vehicle environment. It becomes even more difficult when the environment is restricted inside the vehicle. Scenarios like this have rarely been addressed. In such a scenario, intelligent agents cannot take proper actions because their inputs are, to a significant extent, hidden and, furthermore, they should be acknowledged and preferred in an autonomous vehicle [9]. However, this type of visual analysis problem is still open, primarily because of the following reasons: first, there is a need to detect non-frontal facial features; second, the agents may be performing various tasks in covert space; and finally, visual occlusions exist partially because of physical barriers they are causing (e.g. magazine, table), and sometimes simply due to other agents' views.

The interior of autonomous vehicles provides a unique and intimate environment for their occupants, which pose several challenges and opportunities for human-vehicle interaction systems. The spatial constraints within the vehicles provide ample opportunities to use interaction technology such as touch, gesture, and voice to communicate with their occupants. Personalized interaction can be tailored to suit the individual preferences of the driver and passengers, linked to their emotional states, and also monitor the passenger's physiological state to identify and manage stress and annoyance [16]. An occupant emotion sense system that is designed to convey information to the driver about the emotional states of passengers has the potential to improve vehicle safety.

3.2. Importance of Emotion Recognition

There is little research on human-vehicle interaction in autonomous vehicles. The studies we've found are about human-vehicle system interaction without an autonomous feature. Therefore, numbers of usages except for not drowsiness detection, for instance, are limited. We consider it's now just on the cusp of an area in which studies with a mature state will be conducted [1]. The above research gap constitute the reason for subjecting the present research in which we investigate emotion recognition of a driver while driving a fully autonomous vehicle, considering which the possibilities of repeated applications for feelings and understanding another mental state will be expanded.

We are on the cusp of the an era of vehicles with semi-autonomous and fully-autonomous systems, and thus, designer are considering how the passengers and driver can interact with this vehicle and how can it be safer by understand the driver emotions [11]. Human-vehicle interaction (HVI) is a topic of increasing interest in the era of autonomous driving. Passengers and drivers must be able to summon the vehicle, input a destination, monitor the vehicle and receive notifications about the outside environment. The functions identified in this paper that need to be realized include, but are not limited to, the following: recognizing the drowsiness or tiredness of the driver, notifying the driver about his/her health condition, responding to the facial expression of the driver and providing information about a place, a business, traffic, etc. When a driver is sad or happy, the content shown on the screen of a vehicle may be changed [15]. Additionally, the personal artificial intelligence (AI) of an autonomous vehicle should also detect his/her emotions, discuss it with the driver and thus make decisions in proposals that do not make the driver unhappy.

4. Deep Learning for Emotion Recognition

[14] [3] Emotion recognition in an autonomous vehicle are merely a subset of emotion recognition because the robot–human and robot–robot interactions exist in many other situation [5]. The emotion recognition in an autonomous vehicle has drawn attention among researchers from all over the world. Emotion recognition can be classified into three types: one is based on facial expression recognition, the other one is sound-based emotion recognition and the last one is based on physiological signals. Deep learning has been widely used in the field of computer vision. Specifically, convolutional neural networks (CNNs) can exploit multiple layers of abstraction. For emotion recognition from physiological signals, a wide range of machine learning and supervisory techniques have been developed. The ability to recognize the rider's emotional state in an automated vehicle (AV) can bring various benefits to the overall AV system. Emotion recognition is essential in order to offer a more empathetic driving experience tailored to individual riders with a goal to reduce stress and increase safety. Based on audio and visual rider inputs from various multimodal sensors in the interior of the vehicle, a mixture of deep neural networks has been proposed in this paper to classify a rider's emotional state. The considered scenarios are characterized by diverse riding configurations such as displaying content on a screen inside the vehicle. Ultimately, a multistage 4-layered long short-term memory network, a gated recurrent unit, and a fully connected network demonstrate a weighted average accuracy of 90.75% on the kaggle and 91.89% on a novel dataset, obtained during driving tests using a complete AV (including an interior and an exterior look).

4.1. Neural Networks

Convolutional neural network (CNN), a kind of feed-forward artificial neural networks that has been successful in their application for analyzing visual inputs. One-central characteristic of CNN is relying on its composition of neurons into irreducible hierarchical layers. The inputs to each neuron in the visual cortex are assembled into feature maps. Multiple feature maps are then assembled into a multidimensional output. CNN is sensitive to some type of spatial or temporal patterns deeply, which flares deep and enormous impact in face expression recognition domain. The successive levels are called associative multitask learning, aim at dealing with high-level representation characteristics like the emotion rich. CNN end-toend structure includes an input layer, several hidden layers (for instance, convolutional layers, pooling layers and affine transformation layers) and an output layer [11].

Neural networks are the information processing systems that can be used to improve the system performances in human-vehicle interaction in a variety of domains including computer vision, perception, speech recognition and natural language processing field. Deep learning techniques, have recently drawn great attention in the research, as the model could effectively handle big complex data. Multi-modal data includes speech expression, voice and gesture. Neural networks are the information processing systems that can be used to improve the system performances in human-vehicle interaction in a variety of domains including computer vision, perception, speech recognition and natural language processing. Deep learning techniques, have recently drawn great attention in the research, as the model could effectively handle big complex data. Multi-modal data includes speech expression, voice and gesture [17].

4.2. Convolutional Neural Networks (CNNs)

The sliding window method is a well-known method for detecting facial expressions in deep learning. And apart from the method selection, the base CNN is significant since it can capture the co-occurrence relations among features in different layers. Generally, after an input image is processed by convolution layer, pooling layer, and fully connected layer, the final layer is the softmax layer, which is used to output the likelihood for each class. CNN models can learn the complex information of different expression emotions on different layers of the network. For example, lower layers of CNNs model low-level representations, such as local gradients at different scales, local edge statistics, texture radiance, and orientation, while higher layers use the vector representations of previous convolutional layers and learn cross-channel input features to represent expressions and classify them. Also, CNNs can learn a proper regularization of an activation pattern. Zhang et al. utilized CNN models for learning a proper regularization of the activation pattern. They found the learned filter was similar to known pattern units, enhancing frowny or smiley faces, and occluding eyes or mouth regions to infer expression labels [18].

The idea of Convolutional Neural Networks (CNNs) used for learning task-specific features has proved highly beneficial in the computer vision domain. The main idea behind CNNs is that the input image is convolved with a set of learned filters to obtain feature representations. After applying the convolutional layer, it is common to further compress the spatial feature dimensions using pooling, followed by fully connected layers and a probability layer for mapping the input to probability values. CNN architectures have been used in both facial expression recognition and emotion recognition domains, the two of which are similar tasks and can be interwoven [19]. For example, the two-dimensional inputs (brightness and color) of the captured images can be used as frames of the temporal facial expression data. CNNs can extract impressive features for facial expression recognition tasks compared to classical non-deep convolutional learning methods. CNNs can also be used as feature extractors for the detection of mid-level and high-level expressions, gender and age classification, and other related tasks. Similarly, in [20], the first two or three layers of the CNN were used to extract appropriate high-level abstractions and then, an additional deep neural network was utilized to learn a proper transformation from the intermediate CNN representation to the final output space.

4.3. Recurrent Neural Networks (RNNs)

In the field of contextual words, CNN mainly recognizes the context features that share some words with the current word, while the RNN model participates in learning and updating the features of the current text according to the context features of the text. Specifically, at time t, an expression At emerges, which will first go through a position-independent linear convolution module to extract the feature map of At, and then be employed for batch normalization and non-linear activation function to obtain the expression x t . This is the input state of RNN, which passes through the multilayer RNN model to update the sequential state, and finally outputs the predicted value Y t . Temporarily speaking, the RNN process constitutes a feedback loop, allowing the predictive results at a previous point in time updated upon receipt of new information. In addition, the convolutional neural network (CNN)+RNN model has been demonstrated to be conducive to enhancing the prediction performance of MRI data, which can make the model have the ability to perform global characteristics extraction for MRI data. The experiment proves that the deep learning model will greatly improve the performance of the characteristics extraction task. The research demonstrates that the PBA-RNN deep learning model developed using the MRI data of individuals could effectively predict the valence of the individuals, contributing to the better supplementary of the research progress in emotion recognition [13].

With the advent of the deep learning era, Recurrent Neural Networks (RNNs), proposed by Mikolov et al. [21], have emerged to artificially process natural languages. The basic idea is to make the network able to learn the relationship between the context of different time points of the expression sequence and establish a state equation to achieve the final prediction. A typical application is machine translation with certain time constraints. In fact, RNN has unique advantages in processing the features of expression, especially the features with time information, in the field of feature extraction. In text, RNN is widely used in the field of natural language processing because it can utilize word features with time information to extract the deep features of context semantics.

4.4. Transfer Learning

In recent years, TL has gained much popularity in the development of deep learning-based architectures for natural language processing, including recognition of emotions from speech or text as well as facial expressions. With the development of speech emotion recognition in the call center, speech emotion recognition technology is used for emotion recognition in the field of human-vehicle interaction [22]. In order to overcome the influence of limited data and environmental noise on speech emotion recognition accuracy, researchers have improved accuracy of speech emotion recognition using transfer learning coupled with pre-processing techniques for feature selection and or extraction.

[23] The use of transfer learning (TL) has been successfully applied in the emotion recognition field. TL refers to common techniques used to apply knowledge gained from the training of a model to another task, where a highly correlated and limited amount of labelled data is available. Emotion recognition is a common field where previous knowledge can be reused, since emotions are evolutionarily inherited and can be recognized in various contexts [14]. TL can cope with many challenges in emotion recognition, including pose variations, noise, limited training set, and many complex conditions restrained in the development of autonomous vehicles.

5. Datasets and Preprocessing

Due to the participants really do the low level of emotion as a reaction, could be using the very small percentage (T% = 0.014) of the original captured videos that belong to Sobbing data, in please of happiness case, only to contribute ourselves final model. An off-the-shelf

OpenCV-based Haar-cascade multi-view face detector recognizer detection model was used to detect the people's face cropped-bmwaedited dataset together with half of the original MAD and the PCs faces. The frontal face only was cropped in our dataset images in way of 90x 90 pixels and then fed them into the model to get the XML file and other procedure as explained in the article in details. The mobile i2R In-Car Analytics contain behind training test across selected traffic scenarios our predictive environments in which the emotion and context related probabilistic densities are captured automatically which are then used to build the not online separated driving models towards. The focus of this article is mainly on results, where the emotion prediction classification is employed by adding one Nearest Neighbour predictor and combining three machine learning models. and six new algorithms (four shown in video) in up as two maker are designed to drive comfort and other discomfort face perception value. [cf: 651aea8c-a2ed-4b34-a383-488335ecc926]

The obtained EEG and the immobile ergo laboratory VR scenarios were used to inform the initial study called the LAELab (Fear Inducing VR walking study VPS) based on the twin brothers dataset, which was collected as a subset of the LAEL-Book for hazard warnings and it was used in both low arousal training and high arousal testing hyper-scenarios. So we had to slightly update the dataset and re-evaluate it (e.g. it was different in data presentation sequences). This new study is going to be firstly experimented in VR, called the collision avoidance (C-Avoid-D) dataset, and then it will be increasingly from rest-vp of 109 to 70 km/h in the real car of 120-160 km/h (C-Avoid 2) for simulate the cars environment. At both cases, the EEG data was recorded with Emotiv Epoc 3 and iMotionNeuro eye-tracker head band. For C-Avoid Task, the driving simulation VR data collected for healthy male subjects both under condition of two kinds of classic ECAP experimental scenario: fear inducing and fear suppressing (used both for training and investigating, for details Ref cited). The first performance level in the experiment was validated the prediction of the emotion minivalence by the physiology features of the human. His experiment evisually detected the major fear response with the eye movement trajectory and the fear-eye events preference of the trajectory, both of which were the best prediction of the Presented vPPG accuracy (0.7148) and the EEG accuracy (0.6743) respectively, while not being treated them as the noise property of the input and they showed without noise effects.

We have collected a custom dataset called 'In-Car Database' (ICD) for autonomous cars and executed a human driver vehicle interaction (HD-V2X) experiment to capture the variety of emotions of the passengers (see Ref. [24] for more details). For the ICD dataset, three actual human-vehicle interaction experiments were conducted. The main objective of Experiment 1 was to capture the fear emotion naturally (i.e., without inducing any external emotional video/audio stimulus) from the passengers, which makes long duration driving simulation experiments into account. Data were recorded after a driving scenario in a purpose-built motion-based driving simulator. As a result, six healthy participants were recruited expecting a fear-based driving scenario, and a total of 7.5 h facial expressions data (with 25 fps) were collected for each victim. It was seen that the emotional activity of some of the persons and therefore amount of labelled data could be low. Therefore, we wanted to increase the labelled data by compelling emotions experiment, where they watched videos that induce different emotions. If we look to the gender of *EmoInduced* database, male count is threefold of the females, therefore to balance this issue we have to get two separate datasets as submission for the *In-Car Database*. Therefore, data were recorded from 17 university students (male: 9, female: 8 and aged 19-27) in a simulated drive environment where they actually being seated on an actual car seat (lacking only the car's all main mechanical parts) and going through a decided same test route for every participant. The onboard iS10 eye tracker provided the feature representation change of their facial gestures when their numbed data started to be captured. The details of the experimental procedure adopted for this purpose along with the presented dataset, is explained in the next part. To improve the generalization, we planned for the third experiment where the ICD dataset should be recorded from a real car driving environment on a planned with planned same route within a day as it is expected that the result would be good to test in the driver alertness and in-car interaction also outside in the traffic. Therefore, facial action and then in the second level the emotion labels were recorded continuously with iMotionNeuro system inside the car. A careful designed idle and sad emotional video clips and evaluated comfortable music videos were also used to improve the labelled data through the emotional facial expression stimulus implementation. A total of 57 students partook in the experiment joining the dataset to generaliseas shown in the figure. The dataset is balanced in relation to the age data, is as well as the gender distribution. Here, however, the facial action and test protocol have been designed considering each user’s physiological resting condition. Because of this, we could not get any quality data from the senior citizen and handycap mand SSD dataset by EmoInduced experiment and succusssfully collected only one database for each as the fear and happy. Therefore, since the number of data collected for some options are below 1-2, finally all these captured data classified as happy emotion but presented in separate textual files. Ultimately, the naïve and spontaneous physiological face reactions for the real environment in "In-Car Database" DITEN/ VN Lab can be separately downloaded from here by unzipping these EmoPhysioNavSeq_age_gender_ ID.7z. First, exact physiology facial gestures we will see in the dynamic display. It is noted that the most (said approximately 80–85%) of happiness recordings of all user's data in *EmoInduced* are well labelled with grabbing some data to make preprocessing. Therefore, around 17, 500 frames were extracted as selected EEG suffer data from male for training and rest of all 50, 496 frames are used as a totally separate dataset for testing. It is also noted that we could not any dataset and neither used likewise as an evaluation of *EmoPhysioNavSeq*, so incerely we say we could not any other eference to compare the dataset in this study.

5.1. Commonly Used Datasets

In our use case, a core goal is to influence the operation of autonomous vehicles by accommodating the emotional state of passengers [4]. We demonstrate that superior emotion recognition performance is needed, i.e. the capacity of discriminating a more fine-grained range of emotions. We first illustrate an analysis of the uncertainty in confusion matrices with a focus on the emotions sadness, anger, happiness, and fear, as these are the most prominent expressions in the usage scenarios in in-car settings. We discovered that confusions of prominent emotions indeed cause the core issue of wrong decisions when a car adapts to the emotions.

Several standard databases are available in the community of computer vision and machine learning for emotion recognition [25]. A commonly used one is the CK+ database, which provides seven types of emotion corresponding to seven basic expressions (anger, contempt, disgust, fear, happiness, sadness, and surprise). MEEI is another database that is publicly accessible which has the same emotion expressions as CK+, and one additional expression, neutral. In both Yin and Real-World AFEW databases, eight expressions including valence and arousal are provided for training and testing.

5.2. Data Augmentation Techniques

Other Affective Information (MOSH, FACS, Action Unit Recognition) facilitate predicting drivers' or passengers' emotions / internal states from faces for 'human-like' personalization, effect strategies in situations such as drowsiness or anxiety, e.g., which are of primary importance in most H2P interaction scenarios. Unfortunately, it is currently technically difficult to measure 'real-time' affect reliably in vehicles, using nonintrusive or intrusive, not limited, expensive, multi-modal sensor systems. Consequently, we focus on tailoring the HCI (AMIGO – Affective Man In the GOlden Dragon) to sensitive stimuli in an autonomous vehicle using, in general, FER [26]. We sort people's emotions according to the famous 22 EMOTIONS (e.g., twelve basic, e.g., joy happiness excited (happily surprised), disgust, sadness sadness disappointed (sadly surprised), fear fear anxious (fearfully surprised), anger, surprise surprised amazed (amused), and ten complex emotions (e.g., boredom, interest, and cognitive engagement). Gadgets HUMILIATE FER and decode sample images of these twelve basic emotions in Figure 2, and samples of the Lexical and Confusion Emotions in the Appendix [27].

[classif_keyword: #5dq4g6of] A variety of data augmentation techniques and oracle-based training were used during the network tuning stage to build a more balanced training dataset with emotions represented by a similar number of images. Not only did that provide varieties for certain classes by rotating images, but it also allowed the network to generalize well on the problem by complementing the Transformer architecture with larger amounts of training data [28]. These data augmentation methods in neural network training were also applied to emotion representation on FER (Facial Expression Recognition) in the neural network and developed better classification performance. Use of strategies that could restore the unbalanced class while training is also expected to significantly improve the performance in terms of data-driven machine learning. Thus, training with an increased amount of datasets convergence on local minima that generalizes well on the problem.

6. Training and Evaluation

In the work, we explored an ensemble learning-based multi-task deep model, known as Argous-Net that maps the driver's observations and the AUs to continuous and discrete emotions. Since this model tends to overfit on each other's tasks, therefore, we proposed the ArgousX-Net to reduce overfitting while improving overall performance [9]. The proposed model is sophisticated in which (i) different representation of features are learned separately, (ii) complementary AUs which are considered to contextualize the emotions are exploited in parallel. Therefore, ArgousX-Net can be used in these emotion-rich settings like in the autonomous vehicles where this network could be utilized to learn the driver's emotional state, and therefore a better interaction could be made.

Emotion recognition can be regarded as a crucial investigating topic in the autonomous driving system to understand the emotional state of a driver. In this context, the interaction could therefore be improved and vehicles could adapt to the driver's behaviors and emotional states [12]. Here, we followed Transfer Learning methodology for emotion-rich dataset i.e. image-based AVEC-2019-Workshop that can be used to model driver's behavior [29]. The video dataset consists of driver's facial expressions, driver's observations and driver's annotations like valence-arousal scores. We followed a deep learning-based multi-task learning strategy ArgousX-Net which means driving observations i.e. road maps, and facial action units are estimated as a byproduct.

6.1. Model Training

Deep learning-based feature extractors together are powerful enough to achieve robust emotion state recognition in driver CVT datasets. Additionally, the fusion models we studied have an intermediate fusion algorithm and deep emotion feature extractors achieve robust multi-modal emotion state recognition results for the driver CVT dataset, with greater accuracy than individual unimodal which is occluded. We recommend the DEN model for the design of the fusion-based method for driver CVT datasets; however, in CVT, the DEN model has a higher improvement than VGG-F compared to the unimodal model. We also recommend the high efficiency of the VOT model to perform the driver's real emotion recognition. The problem, therefore, remains an approach to generalizing CVTs for multi-modal data fusion emotion recognition, and it may be the optimal selection so that our unfused R-CNN model can be used as such.

We trained and evaluated three individual models (R-CNN, Deep Emotion Network (DEN), and VGG-F), two fusion methods (mid-, late-level), and a VOT model in unimodal multi-modal and multi-modal emotion data with emotion explicit learning. Prior to training, we used MAC and mean-shift clustering to perform unsupervised training to optimize the CNN part of R-CNN for multi-modal emotion features. The results show the best performance in unimodal and multimodal emotion data for VGG-F, VOT, and DEN, with only midfusion yielding slightly better results. Our model outperforms state-of-the-art emotion recognition models and traditional models based on pre-engineered features for both training subdatasets. Our deep learning scheme is relatively sensitive and helps improve the robustness and accuracy of emotion state recognition in the future.

[30] [14]

6.2. Performance Metrics

An in-depth performance analysis is presented in terms of accuracy to exhibit the capabilities in emotion recognition and their adaptability to the present scenario. Furthermore, this comprehensive investigation exposes the reasons for the most and least successful recognitions, and shows the effectiveness of the transfer learning approach [14]. A variety of statistical measures, such as precision, recall, F-measure, area under the ROC curve, and accuracy, have been computed with a well-known dataset and with newly collected in-vehicle corpora. Experiments firstly use SEWA (Multimodal performance-based emotion & attraction award) German database in order to perform a fair comparison with other existing work and assess our computational approach. Then, attracted by the lack of series of real emotion corpus in the recognition of emotions by autonomous vehicles, the authors decided to collect two different corpora—one collected in laboratory and one during the real use of the vehicle—both within the ERC funding 2connect 2050 project [31].

This section presents the evaluation of the D.EmoNet (Deep Learning-based Emotion Recognition for Human-Vehicle Interaction in Autonomous Vehicles) models, aimed at the recognition of passengers' emotional states during in-vehicle interactions, supporting the development of reliable safety features, and advanced human-computer interaction services. Emotional interaction is interpreted as a multimodal event, as emotions can be conveyed through speech, expressed through facial expressions, and are recognized from the semantic information content of speech [1].

7. Applications of Emotion Recognition in Autonomous Vehicles

Consequently, helping the system to respond to the emotional cues of the passengers in a smart way, regulating its own behavior according to the passengers' emotions, would be beneficial. [5] Specifically, passenger emotions are modelled by a visual-semantic coupled model in the vehicle vision (in addition to a vehicle's sensory system): Emotional Discriminant Convolutional Neural Network (VCN) occurs on one of the views. Output of the VCN is then associated with the vehicle's behaviour, stopping the autonomous vehicle in case a passenger's state-of-mind corresponds to fear. Our aim is to provide unobtrusive emotion analysis, which could be applied to advanced driving assistance systems and autonomous vehicles. The SVM used for classification was trained to recognize basic forced-choice emotions without the use of AUs features (confusion or neutral). With just poses, partial poses or AUs, the classification results decrease in percentage with respect to VGGFace in Table 2. Modifications cause a decrease in the F1-Scores of the figures. Understanding and interpreting few poses or AUs, haven't allow the SVM to develop any effective classifiers.

[9] [32]Emotions invariably play a role in human behavior and affect not only our mental states but also our physical performance in different tasks. The emotions convey important information about how drivers are feeling and can be extremely useful for cooperative human–vehicle interaction. The ability to recognize emotions accurately in the driving environment can provide useful information for road safety and the design of appropriate recommendation systems. Providing autonomous vehicles with the cognitive abilities to recognize passenger emotions can allow such systems to respond dynamically with emotional intelligence for more cooperative vehicle–passenger interactions. An example of an application for emotion recognition in the context of cooperative in-vehicle environments is proposed, where we go beyond the modelling of arousal that can drive an autonomous vehicle to the design of a multi-view system capable of recognising a rich set of discrete emotions, as previously detailed.

7.1. Adaptive Driving Systems

According to the results produced in these usages, a quite good agreement for different emotions is verified by the Emotion Classification Table in article [article_id: ae588867-a404-485d-82fe-41e1e06086ee]. Compared with the other datasets used for driver emotion classification, our drivers emotions dataset is verified to be special because of its rare features. Although the total number of features used for My Driver emotion analysis in our model is less than you, Better prediction for four emotion classes is got by the 2-ge AMDNN. Case 7 in this paper is better processed by the convolutional neural networks. Even better, all AMDNNs obtain nice agreement for all emotion classes. Any of four AMDNN models learns the features perfectly. Since driver emotions in the driving scene are more special than facial expressions, voice, hands processing, body gestures, and so on our better driver emotion recognition can be considered as quite actual problem in this research.

Emotion, including happiness, sadness, anger, and fear, typically influence a human's behavior in driving. Driving performance and driving safety can be both influenced by the emotions of a driver in car. Researchers have studied how to monitor driver emotions dynamically by applying our drivers emotions approach and several versions of the IEMOCAP database. Different kinds of signals have been used as the sources to predict driver emotions. This includes actions of drivers when they drive or deal with driving scene and its dynamic Images, drowsiness Features, Electroencephalogram (EEG), and Electroencephalogram (EEG), Electroencephalogram (EEG).

7.2. In-Car Personalization

Self-learning vehicles were first only able to capture emotions by the driver and to a certain level emotions of the passengers. Spontaneous automatic AER and mood recognition in-car might be fundamentally advanced in several future scenarios. If a car would, in first instance, gather the mood from the direct surrounding of a car as expert system to support the prediction of secondary mood states in faster time scale could even be served. Completely autonomously predicts the secondary mood states in next future, without any third party input, and executes the change of the mood ambient to prevent nervousness or even stress. Keep in mind that there is a latency to consider between the observed trigger and the predicted, relevant mood state ambient at the human-vehicle interface.

With mild intervention, a car could support mood management. For instance, an alarming car could reduce mood by advising to relax and not be governed by the bad driving style of someone behind you, even not allowing the mood of the behind-driver to further irritate [4]. Girls can't be the hero of every car simulator game. The car is learning, in theory, from the movements that bring and maintain the driver in a specific level of arousal. Fair game playing guidelines could also be learned. Anxiety can be reduced by programing an ambient change based on the driver's mood. Summarized, the surrounding of the car could be informed about the mood of the driver and passengers to inform the car of the surrounding, proper ambiance adaptations in-car, via light, sound, smells, temperature and ambient functionalities of the infotainment or in high levels of automation by car support in order to influence the emotions of the human-vehicle interface.

Although the earlier generation of in-car natural language understanding and speech systems have already demonstrated the potential for enhancing in-car task engagement, by nudging the driver to use the infotainment system, satisfaction with the system has further room for improvement [14]. Perceptual inferences could, however, be a vehicle for developing intelligent ambient systems, aware of the emotions and well-being of humans in the car. Ideally, particular items of in-car wellness would be monitored and processed to derive a notion of the actual state. Traditional sensors, positioned on the headrest, on the steering wheel or inside the car, have been used to infer the driver's stress, fatigue or activity level. Supporting the coronavirus, in-car health will probably be even more important. An interesting direction is mood recognition based on facial expressions. Automatic emotion recognition (AER) has reached a reasonable state of maturity, and in-car monitoring is within its scope. AER using facial expressions is practical and available through video sensors in every vehicle. Predicted emotions are meaningful for autonomous vehicles, where the user may display different motivational attitudes and an active or selective learning could be carried out [7].

8. Ethical and Privacy Considerations

On the side of ensuring the accuracy and precision of the recognition tasks, several challenges are posed too. One of the most important is the generalization to different populations and their ethical implications. In fact, how will the system be customized towards each apparent age, gender, cultural nuances, educational level, and so on, of the passengers/users so that the recognition should be similarly accurate in the paradigms under study conditions? This drawback is mainly related to the current absence of any "emotional database" of facial expressions for thousands of people from different cultural backgrounds. But the potential existence of this database poses severe privacy risks [8]. Indeed, similar to the racial bias sometimes observed in automated facial recognition systems interpret the feeling of people with a lighter skin more accurately than for other ethnic groups, a general emotional database might have a bias in favor of some populations, which is not only "inclusivity" but also a threat to "privacy." Therefore, to enhance the benefits of the emotional recognition technologies while limiting the risks, the establishing of an AI ethical rating system will support the full addressing of these issues.

One of the most critical aspects of technology development is ethical consideration. In particular, emotional recognition technologies raise privacy concerns [33]. In fact, one of the added values brought by the adoption of AI in autonomous vehicles will be related to obtaining a solid prediction of the emotional state of the passengers. However, to deliver this, driving activity is often monitored and tracked, particularly emotions of passengers, which can lead to privacy issues due to the potential association of emotional states (an inherently private aspect) with identity. The simple act of letting an AI system recognize a person's emotional states can be a potential invasion of privacy as people have an intuitional discomfort when artificial agents interpret deeply human characteristics. Ethical principles will need to be established in this respect. Furthermore, both ethical and legal implications are related to the data ownership. Who is the owner of the emotional data, the supplier or the manufacturer?

8.1. Data Collection and Consent

None of the participants agree and approve their photos and labels in our databases. The collected databases will be used as the external databases for testing purposes, but not for training. Even when we have bought the pictures of human facial emotions from the companies providing the data usufruct contracts [24], as described in the section 6.2, on every person appearing in photographs, a data label on facial emotions is introduced. This problem occurred because it is not possible to apply to obtain ethical approval retrospectively.

By now we have collected several databases with labeled pictures of the emotional faces [5]. At the beginning of collecting the data for this initiative we have been signed a data usufruct contract with all the participants of this initiative. All the participants have to approve their data and their labelling directly by the user, sending a document or using a mobile application to express their consensus [7]. The problem of contacting people for us was underestimated. The collected pictures of faces were bought from websites providing the labeled pictures of faces for data usufruct contracts. Unfortunately, we cannot contact and ask people for permission to be labeled or not. We have decided that it is very difficult to find and receive a new one, and we have received ethics approval because we cannot undo one's label anymore.

8.2. Bias and Fairness

Materials and methods - The platform required has already been implemented in the laboratory: it is simply an autonomous vehicle, designed for future use in various automotive functions, equipped with many cameras (RGB, depth, NVPMIR, thermal, IR and RGB-Depth) and various sensors (radar, ultrasound and lidar). In the lab we also have simulator-based platforms for testing the driving dynamics and virtual humans, who appear as passengers on the vehicle, to test emotions and other cognitive states. To exploit the actual presence of an autonomous vehicle, we started by analyzing a usual scenario for the detection and understanding of the state in which the driver finds himself: verifying the passenger on an autonomous vehicle. This is the context we want to analyze in this work. We developed a visual and verbal artificial intelligence that belongs to the autonomous vehicle, to reason and act on the user in a solo adaptation of the vehicle respectively similar to a chameleon and to a conversation. The visual AI was specifically focused on analyzing the user's expressions, while the verbal one was dedicated to understanding his intentions. The visual AI is capable of recognizing the emotional state of the user, of pointing out dropouts and of suggesting three new way of treating the user. The verbal AI contemplates the chameleon phase, with the user's response to the AI, and a second phase of action, through some advisable replies. For simplicity the same AI has been developed for a man and for other displays developed, currently, limited to melancholy and jealous women [34].

In particular, this emotional field has not yet been extensively covered, also because the techniques already present to study emotions in an ecosystem such as that of vehicle and driving are quite limited. It is true that the level of autonomy of vehicles has a significant impact on the type of HVI that can be implemented, but it is essential to consider also the gender, age, culture and habits of the drivers. We have tried to create an environment, based on visual artificial intelligence, which is capable of recognizing the emotional situation in which the user of the vehicle is situated by suggesting culturally appropriate initiatives [32] even modifying the behavior of the vehicle in order to meet the emotional needs of the user. Various machine learning techniques have been experimented, comparing these with recent techniques of deep learning aimed at the analysis of the user's expressions (such as YOLO and OpenPose). In order to enrich the entire mechanism with a human-like intelligence to enrich the non-verbal language of communication with the user of the autonomy [35].

Introduction - Positive interaction between the human driver and an autonomous vehicle (AV) is a keystone task for the development of future personal mobility. The work carried out over the years has identified different elements that are at the basis of these interactions. However, literature often focuses on the customer acceptance of AVs, gradually placing human-vehicle interaction (HVI) in the background. We discuss the most important components from our point of view, evidencing how some of these could represent an anthropocentric vision of HVI. In our view it is of primary importance to maintain the driver's feelings and emotions, including those linked to traditionally masculine traits such as competition and challenge, in order to maintain the identity of the driver at the bases of the HVI.

9. Challenges and Future Directions

The classification of driver and passenger emotions data using an emotional deep attentive-DAN in the AVEE system is done using ethnic group data using a real RGB and thermal data fusion system control demands and measures emotion changes [24]. Moreover, if only a driver face is regarded in the face detection process, the accuracy of real-time emotion recognition would be decreased because they might not be able to manage the information of a passenger's face. Depending on the emotion of a passenger, objects inside the autonomous vehicle may demand attention and understanding of passengers' emotional states. We add the real air container data (Temperature, Pressure, Wind velocity, Noise, Humidity, Smoke) as a representative example of these objects. As a result, passengers' gestures can be modeled when they feel cold, hot, fine comfortable, uncomfortable, sleepy exhausted due to high CO2 etc. Moreover, passengers' facial expressions can be simulated and captured with respect to the texture and material info. Our system performs facial attribute recognition on synthetic colors of MS COCO and CelebA-HQ data from BUPT-HCI, UTKFace demographical and real representative front, and side thermal data of TID29 data. Our realistic product exhibits very close performance to the original face identification.

[9] An effort must be made to gather ground truth human emotions that resemble practical driving conditions in order to produce usable datasets [5]. In addition to the risk of a small but noticeable sampling bias, the publicly available dataset can not be expanded effectively without repeatedly sampling from the same passengers for the training and test data because there is a limited number of data samples. This means that, in practice, it is necessary to make the most of the available data to avoid re-using the same individual data samples for different purposes and optimally using all available data. This is due to the different reaction times and subjective variations in emotion among people in various environments and their corresponding individual physiological health conditions. In addition, in the case of using head rotation detection among facial expressions, precautions must be taken for driving conditions such as sudden braking, lane deviation, and strong side-to-side steering, in order

to maintain sound and stable conditioned data. Given the recent increase in road accidents caused by the movement of visiting drivers, it is essential to handle sleeping driver detection as only part of the driving state, not as the sole driver's detection method because the recording of only driver's facial expression data by visual sensing in vehicles with Level 3 or lower autonomous driving is not sufficient. These inconsistencies between the recognition quality of simulated data and the real data may affect the safety benefits from Level 3 or lower autonomous vehicles during a safety-critical phase in the human-vehicle transportation system when the driver must take over driving manually (e.g., driver and passenger arousal, different emotion, very short reaction time, driver/convoyplex fatigue caused by prolonged driving, psychological drivers fatigue, etc.). The solution for these issues and the theoretical framework of equivalent-distribution training strategies such as reinforcement learning-based realistic scaling factor calculation, deep Q-learning network-based scaling factor learning, and meta domain adaptation are necessary.

9.1. Real-time Processing

Audio and video data in particular present challenges for real-time processing due to the often large amounts of data that need to be processed. Video data can be especially challenging for processing because it often comes in high-resolution colors, each frame consisting of millions of pixels. Consequently, such data needs to be passed through several bottleneck analyses before the relevant high-level information can be extracted. Traditionally, audio and video data are processed in multiple steps. We describe each of the major steps for audio and video, using emotion recognition as an example scenario. In real-time operation, the data flow between the computer-assisted recordings, the memory, and the individual stages of processing must be scheduled such that appropriate data items are sent to each stage of processing in a timely manner, so that the entire audio or video stream of interest can be endowed with high-level semantic labels in real-time.

An AV should be able to process a large variety of sensor data in real time, including visual content and surround sound [36]. In the case of AVs, the massive amount of data that needs to be processed for many aspects of the system, such as perception, planning, control, monitoring, and human-vehicle interaction, has to be processed from many different sensors. Although, to process and visualize this data, modern multi-core computers offer a great deal of processing power using parallel and distributed computing frameworks, some tasks within the AD stack are crucial, and small latencies are essential for road safety [9]. For most of the subsystems' components of an AV, information on the world and on the status of the car can be collected and used multiple times (e.g., for perception, mapping, object recognition, and sensor fusion) before the vehicle needs to perform an action. In some crucial situations, such as collision avoidance or monitoring of the driver and the passengers, small latencies in the response of the system can make the difference between safety and danger.

9.2. Interpretable Models

In an independent study and development of autonomous vehicles, deep neural networks (DNN) are used in end-to-end approaches. The magnificence of DNN is manifested in its ability to automatically learn and discover exciting patterns and knowledge, and its capability to maintain scale-invariant recognition features owned by humans. These are the inner reasons that drive the "explosive" growth of deep learning in computer vision. Because of weaknesses in rules and transparency, influencing human factors and ethical concerns lie heavy upon society's decision making and have become obstacles to the widespread use and trust of such AI-based systems in our autonomous and ubiquitous future experience. Explaining the intelligent behavior of autonomous vehicles is considered a crucial prerequisite for improving the transparency of the intelligent transportation system [37].

Emphasizing transparency and ethical concerns, the autonomy and responsibility of the onboard artificial intelligent (AI) assistants need to be addressed. The in- vehicle interaction between drivers and vehicles is accompanied by cognitive activities, such as emotional clues, for the driver to communicate with and understand the vehicle environment [9]. According to the service series that is provided by the human- oriented—smart— connected—autonomous vehicle, AI speech representatives are established as AI-driving robot users to facilitate the human-connected relationship. In autonomous vehicles, the human conversation with AI-driving robots, emotional cues, and the feedbacks are concerned [13].

9.3. Multi-modal Emotion Recognition

An analysis of driving context derived from passengers' traits as possible background for affect recognition algorithms was published by Soleymani et al.. Predicted emotions are derived from perceived situation and are also training by passengers' behavior. So, they provide a driving strategy that is enhancing user experience. Transition control, as common application, was applied on modelling perceived events. Kefalidou et al. propose developing an emotional contagion machine learning enhanced model for passenger comfort when sharing driving decisions to reduce any potential sources of discomfort. They showed that the belief that lateral control is shared with the system reduces the level of fear users express in their ratings [16]. Convergent with this, the analysis of one concept-user group and the user-study steering experiment showed that user-centred HCI concepts of the system-controlled trajectory minimize the physical effect size. Emotional contagion modelling was extended with an integration of existing personality and physiology-based HMI methods to develop our multi-modal, improved in-cabin passenger comfort enhancement.

Emotions in interaction can play a crucial role in achieving successful experiences in autonomous vehicles (AVs) for all passengers. AVs will need to detect and adjust to the emotional state of a passenger to guarantee safe driving and successful user experiences. Interactions with AVs in urban environments can be particularly emotional because of the uncertainties caused by the behavior of other traffic participants and other, sometimes unexpected, events [3]. Despite slower driving and less traffic-pressure, the frustration and anger at traffic jams or all the dangerous situations due to other traffic participants are scenarios that can stress a user even more. Beyond recognizing the stress and perhaps frustration scenarios, the detection of passenger emotions such as anger, fear, depression, confusion, disgust, surprise, and joy can be important. Some of these emotions have often not been considered in investigated AV scenarios and were often not observed through HMI affective signals. These have been typically overlooked in tests conducted with infrequent driving maneuvers, which mainly affected motion sickness. Areas like coping with navigating an unknown area and adjusting to personalized driving style were often not included as causes for frustration, stress, or fear [13].

10. Conclusion

We find that the choice of the Convolutional Neural Network improvements the accuracy to identify emotions based on the facial expressions, but the best approach to solve those issues will depend on further testing of the intentions and abilities of the end users. In our work, our system achieved classification accuracy of 85.2%, 96%, 90%, 85%, and 80% for Angry, Fear, Disgust, Happy and Sad, respectively. Our work from the analysis of the relaxation time in bipartite random walks with a limited number of steps interprets and explains how the joint action of backward and forward jumping behavior determines the bow-tie relaxation time distribution, which diverges for incomplete networks. Thus, we can claim that the method is now ready for implementation in an in-car analytics system for automatic vehicles. A driver could be prompted to provide the system with the truth in their emotional expression, as it is investigated how to use the emotion regulation capacity of individuals to set emotional responses on expressions that are true.

Our deep learning-based emotion recognition method for human-vehicle interaction has been implemented and tested in an automated vehicle [24]. Our system consists of three components: face and head detection, facial landmark localisation, and emotion prediction. We use the YOLO version 3 algorithm, which can recognise faces of different sizes, with different positions and angles. The 5-point facial landmark localiser we used determined the points of the eyes and the mouth to the coordinates of the eyes and the mouth. The emotion prediction part distributes the pixels of the emotions from the position of the facial landmarks to the counters used for 5 different emotions (angry, disgusted, fearful, happy, sad) [14]. The obtained images, with a facial expression on the 5-point facial landmarks, were for the model input. The DRER sensor fusion algorithm integrates CNN-based facial expression recognition (FER) with bio-physiology signals acquired by electrocardiography and electrodermal activity. The FER model CNNs are pre-trained on multi-view facial emotion recognition in the wild, and we fine-tune the CNNs on a dataset of drivers with differing emotional representations [3].

10.1. Summary of Key Findings

1. We tested different combinations of two data modalities, voice and face, for VAS prediction. Their performance is analyzed in detail. We used mel-frequency cesptras for voice and hog features for face. Those extracted feature are integrated by simple linear combination. We find that the GS1 performs better than single modality and GS2 which uses both modality as in-put performs better than GS1. Further, our integrated model using a more complex fusion mechanism as- compared to simple GLASSO, i.e., two level transformer network DHS performs best [14]. It is thus concluded that the two data modalities have some complementary information which can be successfully aggregated to give a better result. 2. We demonstrated the performance of developed emotion prediction model in a human-vehicle interaction intro scenario based simulation of autonomous vehicle [15]. We proposed

a real world interpretation on how our emotion recognition model could be used for the human-vehicle interaction case study. Moreover, their contemporarily drivers could be classified into different categories like relaxation, boredom, enjoying, frustrated, anxious, impassive, etc., considering his valence and arousal as fully scaled output. Our model can be used to impact the future operation of the car, say in level-5 vehicle, in a real world scenario.

The study proposed an emotion recognition model for human-vehicle interaction in the inner loop of autonomous vehicle platform [24]. A word-embedding based two-level transformer network was used to evaluate, confer, and predict a 12-dimensional emotion state defined by the Circumplex model with valence and arousal dimensions. We recorded data on the human face and voice for the emotion recognition task. The developed emotion recognition model can be integrated in in-vehicle analytics system for level-5, so that it can potentially impact operation of vehicle according to the inferred emotion and state of vehicle. Key findings are summarized below:

10.2. Future Research Directions

In the connected vehicle environment, offers to the driver may also emerge in different ways such as responding to different road participants' stressful faces or interacting with the driver in a way that will positively affect the facial expression of the driver. Moreover, a variety of applications can be developed for the driver and automotive user by approaching human-vehicle interaction through the emotion recognition capability of the vehicle, with Internet of Things (IoT)-based approaches and human-centered vehicle development processes, especially from an affective computing point of view. In autonomous vehicles, an approach in which the vehicle is affected by the emotional state of all participants in a vehicle, including passengers, is also expected to open new doors in the near future. In this way, the automotive industry has the potential to become an ecosystem that develops by learning from even the weak signals of emotional communication that have not been necessarily expressed before [14].

Deep learning-based methods are also paving the way for the development of real-time emotion recognition engines that capture driver emotions while considering attempts to change lanes, following other cars, and other factors. Such a development can be expanded to a human-centered approach that considers even the emotions of other road users. For example, a vehicle that captures a driver's smile and acts as a result, reacts positively to an angry honking of a driver behind it, continuously prevails its peace-of-mind state and drives with a safe speed-distance policy [15].

Reference:

1. Tatineni, Sumanth, and Venkat Raviteja Boppana. "AI-Powered DevOps and MLOps Frameworks: Enhancing Collaboration, Automation, and Scalability in Machine Learning Pipelines." *Journal of Artificial Intelligence Research and Applications* 1.2 (2021): 58-88.
2. Ponnusamy, Sivakumar, and Dinesh Eswararaj. "Navigating the Modernization of Legacy Applications and Data: Effective Strategies and Best Practices." Asian Journal of Research in Computer Science 16.4 (2023): 239-256.
3. Shahane, Vishal. "Security Considerations and Risk Mitigation Strategies in Multi-Tenant Serverless Computing Environments." *Internet of Things and Edge Computing Journal* 1.2 (2021): 11-28.
4. Abouelyazid, Mahmoud. "Forecasting Resource Usage in Cloud Environments Using Temporal Convolutional Networks." *Applied Research in Artificial Intelligence and Cloud Computing* 5.1 (2022): 179-194.
5. Prabhod, Kummaragunta Joel. "Utilizing Foundation Models and Reinforcement Learning for Intelligent Robotics: Enhancing Autonomous Task Performance in Dynamic Environments." *Journal of Artificial Intelligence Research* 2.2 (2022): 1-20.
6. Tatineni, Sumanth, and Anirudh Mustyala. "AI-Powered Automation in DevOps for Intelligent Release Management: Techniques for Reducing Deployment Failures and Improving Software Quality." Advances in Deep Learning Techniques 1.1 (2021): 74-110.