



Emotion analysis in Parkinson's patients using computer vision-based system during simulated driving scenarios

Ariel A. Lopez-Aguilar^a, Juan Felipe Medina-Lee^b, Jorge Villagra^{c,*}, Eduardo Rocon^c, Almudena Cerezo-Zarzuelo^{d,e,f}, Juan Pablo Romero^{d,e,g}, M. Rogelio Bustamante-Bello^a, Sergio A. Navarro-Tuch^a

^a School of Engineering and Sciences, Tecnológico de Monterrey, Anillo Periférico 6666, Tlalpan, Mexico City, 14380, Mexico

^b Computer Science and Engineering Department, Universidad de Puerto Rico, 259 Av., Alfonso Valdés Cobián, Mayagüez, 00680, Puerto Rico

^c Centre for Automation and Robotics (CSIC-UPM), Ctra. de Campo Real, KM 0,200, Arganda del Rey, Madrid, 28500, Spain

^d Brain Injury and Movement Disorders Neurorehabilitation Group (GINDAT), Institute of Life Sciences, Universidad Francisco de Vitoria, Ctra. Pozuelo-Majadahonda, KM 1.800, Pozuelo de Alarcón, Madrid, 28223, Spain

^e Brain Damage Unit, Beata Maria Ana Hospital, Dr. Esquerdo 83 St., Madrid, 28007, Spain

^f Escuela Internacional de Doctorado de la UNED (EIDUNED), Bravo Murillo 38 St., Madrid, 28015, Spain

^g Faculty of Experimental Sciences, Universidad Francisco de Vitoria, Av. de Majadahonda, 800, Pozuelo de Alarcón, Madrid, 28223, Spain

ARTICLE INFO

Keywords:

Parkinson's disease
Facial expression recognition
Computer vision
Driving simulator

ABSTRACT

Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects both motor and non-motor functions. These implementations can significantly interfere with daily activities, particularly those that require coordination, attention, and decision-making such as driving. For older adults, the ability to drive is closely related to autonomy and quality of life, as it provides access to essential resources, social interaction, and personal independence. Therefore, understanding how PD influences emotional behavior in driving contexts is essential. In this study, the emotional responses of people with PD and a control group of healthy older adults are analyzed while driving simulated scenarios using deep learning facial expression recognition. Emotional probabilities were extracted frame by frame and analyzed to identify group differences. The findings of these experiments suggest that although general emotional patterns may appear similar, specific high-demand situations can elicit different affective states in individuals with PD. This highlights the usefulness of facial expression recognition systems in studying human behavior during complex driving tasks and supports their potential role for the design of adaptive driver assistance technologies for neurologically vulnerable populations.

1. Introduction

Parkinson's disease (PD) is a neurodegenerative disorder that affects the central nervous system, affecting primarily brain cells that produce dopamine (Kalia & Lang, 2015), leading to motor impairments such as resting tremors, muscular rigidity, bradykinesia, and postural instability. In addition to those motor symptoms, patients experience cognitive decline, emotional disturbances, visual-perceptual deficits, and automatic dysfunction (Álvarez, 2016). Furthermore, variability in the efficacy and timing of PD medications in symptom relief, including side effects, has a severe impact on patients' ability to perform daily activities, including tasks that require coordination and cognitive flexibility, such as driving (Crizzle et al., 2012).

Driving is essential for employment, social activities, and maintaining a sense of independence; however, it demands sustained attention, quick decision-making, environmental awareness, and motor coordination to respond to the dynamic changes in the road (Phokaewvarangkul et al., 2019). When these capacities are diminished, individuals are often forced to cease driving, this transition leads to reduced mobility, social isolation, and a decline in overall quality of life (Jacobs et al., 2017). Drivers with PD tend to show a lower driving performance compared to healthy individuals due to the compounded effects of motor and non-motor impairments (Gotardi et al., 2022). Simulator-based studies have reported that drivers with PD are 2.63 times more likely to be involved in crashes than their healthy counterparts (Thompson et al., 2018). However, driving performance does not

* Corresponding author.

E-mail address: jorge.villagra@csic.es (J. Villagra).

<https://doi.org/10.1016/j.chbr.2026.101293>

Received 26 August 2025; Received in revised form 31 August 2026; Accepted 1 September 2026

Available online 5 September 2026

2451-9588/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

solely rely on motor-cognitive ability; emotional behavior also plays a critical role. The uncertainty, frustration, and fear triggered by PD symptoms can exacerbate performance difficulties. The role of emotions in driving performance has gained major relevance. Emotions influence attention, perception, and decision-making that are expressed through measurable physiological signals. Advanced Driver Assistance Systems (ADAS) have emerged as a critical solution to improve the safety for drivers and passengers. While these systems traditionally monitored the vehicle's exterior to detect the road conditions, lighting, and proximity to obstacles, they have evolved to analyze the internal state of the driver as well (Izquierdo-Reyes et al., 2018) using technologies such as electroencephalography (EEG), electrocardiography (ECG), galvanic skin response (GSR), and facial expression recognition (FER). Among these, FER is specially suited for intelligent systems to adapt accordingly, making it a key component for ADAS by helping to detect fatigue, attention levels, and emotional load of drivers in a non-invasive way (Jeong & Ko, 2018; Lopez-Aguilar et al., 2021).

Despite the growth of research into FER applications for ADAS, significant challenges remain, such as variability in lighting and head poses, as many existing datasets are generated under highly controlled conditions (Chen et al., 2024). Beyond these environmental factors, a critical clinical challenge in neurodegenerative contexts is hypomimia. Often considered a motor symptom of PD, hypomimia is characterized by the reduction or loss of spontaneous facial expressions, typically manifesting symmetrically as the disease progresses diagnosis (Cossini et al., 2024). While recent advancements in Deep Learning (DL) and Machine Learning (ML) have significantly improved FER accuracy to healthy populations, their application in individuals with PD remains constrained by this facial presentation (Skaramagkas et al., 2025).

This creates a clear research gap regarding the reliability of FER models for PD drivers in dynamic environments. Therefore, the primary objective of this study is to evaluate the effectiveness of a computer vision-based system in detecting and categorizing the emotional states of drivers with PD compared to healthy controls during specific simulated driving maneuvers. To achieve this, the following specific objectives were established:

- To determine differences in emotional valence (positive vs. negative emotion probabilities) between PD and control groups across varying levels of task demand.
- To identify high demand scenarios or road complexities that trigger prominent negative emotional responses in the PD group.
- To analyze the influence of task load on the temporal resolution of facial expressions, considering the modulation of executive control and attention.

Based on these objectives, this research addresses the following question: In which scenarios do individuals with PD exhibit a higher propensity for negative emotions, such as fear and anger? Addressing this question is crucial for developing affective-adaptive ADAS tailored to neurologically vulnerable populations, potentially enhancing their independence in daily activities. While driving simulators provide a safe, controlled, and replicable environment, these findings aim to bridge the gap between affective computing and clinical PD research. The null hypothesis is that the FER system will detect significantly lower intensities of emotional expression in the PD group due to hypomimia, regardless of the task difficulty. Meanwhile, the main hypothesis is that the high-demand maneuvers will elicit measurable, detectable changes in the emotional state of both groups, through the system's sensitivity may vary between populations.

2. Background

PD is a complex progressive neurodegenerative disorder characterized by motor symptoms such as resting tremors, muscular rigidity, bradykinesia, and postural instability. It is the second most common

neurodegenerative disease after Alzheimer's disease. The number of people over 50 years affected by PD is expected to reach between 8.7 and 9.3 million by 2023 (Dorsey et al., 2007), this represents a growing public health concern due to its social and economic implications.

Pathologically, PD is characterized by loss of nigrostriatal dopaminergic innervation, along with neurodegeneration of cells located in other regions of the neural network (Macphee, 2018). Clinically, the diagnosis of PD is based on the presence of bradykinesia in combination with resting tremors or muscular rigidity. Early symptoms present asymmetrically and without atypical features. An asymmetric onset of motor symptoms and a good therapeutic response to levodopa are considered two of the most important criteria to distinguish PD from other forms of Parkinsonism (Williams-Gray & Worth, 2016).

2.1. Motor and non-motor symptoms of Parkinson's disease

Resting tremors are typically the most recognizable motor feature of PD. They often begin unilaterally and distally, usually affecting the limbs, and are characterized by a frequency of 4 to 6 Hz. A classic manifestation is the "pill-rolling" tremor of the hand; however, resting tremors can also present in the jaw or tongue (orofacial tremors). In addition, asymmetric postural tremors can occur, presenting at a higher frequency of 6 to 7 Hz when the arms are extended (Massano & Bhatia, 2012). Other motor symptoms, which often emerge as the disease progresses, include muscle rigidity, postural instability, gait disturbances, and dysphagia.

Another symptom of PD is bradykinesia, characterized by a progressive slowness of movement and a reduction in amplitude during repetitive tasks. Bradykinesia manifests as a decrease in facial expressivity (hypomimia) and softened speech (hypophonia). Hypomimia is particularly significant for automated monitoring systems, as the facial expressions are frequently misidentified as depression or apathy, even in non-depressed patients, due to impaired motor control of facial musculature (Abrami et al., 2021). Clinical assessment of hypomimia utilizes the Unified Parkinson Disease Rating Scale (UPDRS) tool (Fahn & Elton, 1987), where neurologists rate expressions on a 5-point scale: 0 for normal, 1 for minimal hypomimia, 2 for slight but abnormal diminution of facial expression, 3 for moderate hypomimia, and 4 for severe or complete loss of facial expressions.

Although PD is typically diagnosed based on its characteristic motor symptoms, it involves a wide range of non-motor symptoms that significantly contribute to disability and early placement in nursing homes (Poewe et al., 2017). These include sleep disturbances, autonomic dysfunction, and sensory abnormalities. While pharmacological treatments can mitigate motor issues, these non-motor manifestations are often less responsive to dopaminergic therapies and present a major challenge for comprehensive care (Carvalho et al., 2018).

Cognitive impairment related to frontostriatal dysfunction often appears in early stages, potentially progressing to dementia as the disease advances. The prevalence of dementia is estimated at 15-20% after five years, rising to 46% after ten years (Postuma & Berg, 2016). Furthermore, neuropsychiatric symptoms are highly prevalent; clinical depression affects approximately 35% of patients (Reijnders et al., 2008), while anxiety affects up to 60%. It often occurs with depression and can manifest generalized anxiety, panic attacks, or social phobia (Sieber, 2018). It should be noted that anxiety symptoms tend to fluctuate with motor status, intensifying during "off" periods when dopaminergic medication effects are diminished. These affective states are often exacerbated during high-cognitive-load tasks, such as driving, where the intrinsic fatigue of PD can impair performance regardless of motor preservation.

2.2. Parkinson's disease and driving

Driving is a complex daily activity related to an individual's autonomy, mobility, and quality of life. The presence of motor and non-motor

symptoms in PD patients can compromise these functions, increasing the risk of traffic violations, near misses, and crashes (Madeley et al., 1990). As a result of these impairments, many patients with PD adopt reduce their driving time or permanently quit driving; although these behaviors are intended to preserve safety, they can have unintended consequences such as limited social participation, and dependence on caregivers (Brock et al., 2022). The loss of driving privileges is associated with depression, isolation and poor quality of life. Bradykinesia and rigidity are linked to the avoidance of complex driving conditions, whereas cognitive decline and emotional disturbances are associated with diminished self-confidence decline and emotional disturbances contribute to diminished self-confidence and attention control (Chang et al., 2024). Although many PD patients adjust their driving habits by reducing frequency, shortening routes, and modifying vehicle setups to accommodate their capabilities (Brock et al., 2022).

Simulator-based studies have offered insights into how fluctuations in symptom severity, especially during “wearing-off” phases when medication efficacy decreases, and negativity affect driving performance. Interventions such as opicapone have been shown to improve adaptability in driving scenarios (Marano et al., 2024). Epidemiological data from Swedish national traffic registries indicate that PD patients are at higher risk of single-vehicle crashes and severe injuries compared to a general driver population (Forsman et al., 2021). Driving simulators provide a unique platform for safe studying the interactions between PD symptoms and driving behavior. These controlled and repeatable scenarios allow researchers to manipulate task complexity, environmental stressors, and time constraints without exposing patients to real-world risks. When integrated with physiological monitoring systems, simulators allow the identification of emotional responses such as stress, anxiety, or frustration that may not be evident by behavioral observation (Classen et al., 2022; Tozman et al., 2015). These applications can support early detection of emotional disturbances and cognitive fatigue, ultimately contributing to the development of monitoring and assistive technologies for drivers at risk.

2.3. Emotions and facial expression analysis

Emotions are complex psychophysiological responses to events, individuals, or environmental changes. These responses can be conscious or unconscious and play a critical role in everyday tasks, influencing decision-making, reasoning, interpersonal communication, and quality of life (Clark et al., 2020). Emotions manifest through multiple channels, including speech, body posture, physiological signals, and facial expressions, widely regarded as the most immediate means of emotional communication (Taghi Zadeh et al., 2019; Videla et al., 2019).

To understand the relationship between emotions and facial expressions (Ekman, 1992), and his colleagues identified a core set of basic emotions: joy, anger, contempt, disgust, fear, sadness, and surprise. Expressions are innate and consistent between individuals regardless of age, ethnicity, or culture. To systematically analyze facial expressions Ekman and Friesen developed Facial Action Coding System (FACS), a standardized tool for classifying facial muscle based on the Action Units (AUs). FACS enables the identification and interpretation of individual

or grouped muscle activations that correspond to specific emotions (Ekman & Rosenberg, 2012). However, FACS originally required manual coding by trained experts; this process requires a lot of time making it unsuitable for large-scale analysis or real-time applications. To address these limitations, researchers began to integrate computer vision algorithms to automate facial expression analysis. With the advancement of ML and DL algorithms, automated FER systems have been applied in domains such as psychology, education, marketing, and affective computing (Canedo & Neves, 2019; Clark et al., 2020).

Common automated FER pipeline consists of the following stages: image acquisition, pre-processing, feature extraction, and prediction. As illustrated in Fig. 1, this process begins with the capture of a frame (input image) from a video stream. During the pre-processing stage, the system identifies and extracts regions of interest (ROIs), in this case, faces, to reduce computational workload and focus on relevant features. This step may involve face detection, cropping, resizing, and color filtering depending on the specific architecture of the model. When the ROI is prepared, the image is passed to a trained model that classifies emotions. The most used algorithms in FER systems are Convolutional Neural Networks (CNN) due to their performance on image-based tasks, and Support Vector Machines (SVMs) known for their effectiveness in binary and multiclass classification problems (Liu et al., 2017). The output of these models is a label that indicates the emotion with the highest probability.

As previously mentioned, PD patients also experience non-motor, including emotional-processing impairments that lead to difficulty in recognizing emotion from faces (Argaud et al., 2018). Research indicates that these deficits are not uniform; for instance, PD patients with mild cognitive impairment may show selective impairments in recognizing specific emotions like disgust, suggesting that cognitive status directly modulates socio-emotional decoding. Consequently, examining FER-derived emotional signatures in ecologically valid contexts, such as simulated driving, provides a crucial clinical rationale for understanding how PD impacts complex daily activities.

While PD impairs driving performance and increases crash risk, the internal emotional states that contribute to these risks remain difficult to monitor objectively. FER offers a non-invasive solution for real-time monitoring; however, its application in PD is challenged by hypomimia, which blunts the very signals these systems are trained to detect.

By deploying a DL FER pipeline within a high-fidelity driving simulator, this research evaluates the system's ability to decode the emotional responses under varying levels of task demand. This approach moves beyond the simple application of known techniques by investigating the reliability of affective computing when applied to neurologically vulnerable populations. Ultimately, these findings are intended to inform the development of adaptive ADAS that can account for the unique emotional signatures of drivers with PD.

3. Methods

The experiment was designed to collect data from individuals diagnosed with PD and from an age-matched healthy control group, while they performed a series of simulated driving tasks. The aim was to

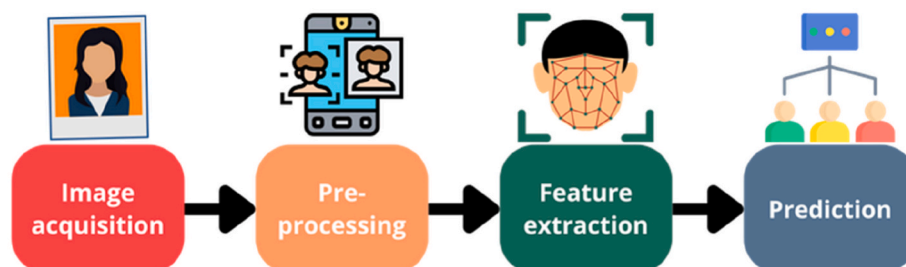


Fig. 1. Common automated FER pipeline (Lopez-aguilar et al., 2023).

observe and compare emotional response under controlled driving conditions.

3.1. Participants

A sample of eight drivers with PD and seven age-matched healthy controls with no history of neurological disorders was recruited. All PD participants were clinically diagnosed by a neurologist at Hospital Beata María Ana. To ensure safety and capture functional performance representative of daily-life conditions, PD participants maintained their regular medication schedule for the 30 days prior to the study and were tested in their “ON-medication” state, approximately 60 min after their last oral dose of Levodopa. All participants held active driving licenses. The PD group reported an average of 43.57 ± 8.84 years of driving experience, a daily distance of 20.42 ± 18.54 km, and a driving frequency of 3.43 ± 2.23 days per week. The control group reported 44.86 ± 8.70 years of experience, 42.86 ± 21.57 km of daily distance, and a frequency of 5 ± 1.77 days per week.

Meanwhile, the exclusion criteria consisted of: (1) took more than 180 s on the Trail Making Test B (TMT-B); (2) scored more than 7 s on the Rapid Pace Walk; (3) scored greater to 3 on the modified Hoehn & Yahr scale; (4) scored a punctuation lower to 24 on the Montreal Cognitive Assessment (MoCA); (5) had a use field of view greater than 3 (range 1 – 5, 5 indicating high risk for crashes due to visual perceptual impairment); or (6) were older than 75 years.

All participants provided written informed consent in accordance with the Declaration of Helsinki. The study protocol was approved by the Ethics Committee of the 12 de Octubre Hospital (Code: 23/603, approved 6th February 2024). Demographics from participants are shown in Table 1.

3.2. Experiment setup

To ensure environmental control and reproducibility, all tests were conducted using a fixed-based driving simulator. The visual interface consisted of three 27-inch monitors providing a panoramic field of view, with the participant seated approximately 0.9 m from the central screen. An additional monitor was used by the researcher to control the simulation. The virtual environment was generated using SCANer Studio 1.9 (AVSimulation, France), which provided validated models for vehicle dynamics, road infrastructure, and autonomous traffic agents (Medina-Lee et al., 2023). The simulation was executed on a workstation equipped with an Intel Core i7-7700 (3.6 GHz) processor and 16 GB of RAM. Participant interaction was facilitated through a Logitech G29 kit which featured a force-feedback steering wheel and pedal set with progressive braking resistance.

A custom software application was developed to synchronize and record participant interactions with the simulator data. A Logitech C920

HD Pro Webcam was used for video acquisition, capable of capturing at 30 frames per second with 1080p resolution. The camera was positioned at approximately a 35° yaw angle relative to the participant's face, ensuring optimal facial visibility for subsequent emotion analysis. The overall simulator configuration is illustrated in Fig. 2.

3.3. Experimental procedure and tasks

The experiment consisted of six driving tasks performed in a fixed, incremental order of complexity. The sequence was established to mitigate the learning curve associated with the driving simulator, ensuring all participants reached a baseline of motor proficiency before engaging in the most cognitive demanding scenarios. Each task was performed four times: the first repetition served as a trial of familiarization, while the subsequent three were recorded for analysis without supervisor intervention. An exception was made for Task 2, which was performed only twice due to its extended duration. The tasks were defined as follows:

- Task 1: Participants were required to start the vehicle and maintain a constant speed in a straight line.
- Task 2: Participants were required to start the vehicle, accelerate and decelerate to maintain a safe distance behind a lead vehicle in a straight-line scenario.
- Task 3: Participants were required to perform a 90° right turn at an intersection, ensuring the vehicle remained within lane boundaries, followed by parking at a specific spot.
- Task 4: Participants were required to perform a 90° left turn at an intersection, ensuring the vehicle remained within lane boundaries, followed by parking at a specific spot.
- Task 5: Participants were required to enter and exit a roundabout with different circumstances:
 - o A: There was no traffic.
 - o B: Another vehicle was entering the roundabout.
 - o C: Another vehicle was already in the roundabout.
- Task 6: Participants were required to reverse the vehicle for 10 m and come to a controlled stop.

3.4. FER pipeline

A total of 30 videos per participant were recorded, resulting in 450 videos across both PD and control groups. These data were processed using a FER algorithm based on the VGG16 architecture, a DL model established for large-scale image classification. The architecture consists of 13 convolutional layers for feature extraction and three fully connected layers for classification (Aksoy & Sayin, 2021, pp. 705–714)



Fig. 2. Driving simulator setup.

Table 1
Demographic characteristics of participants.

Participant	Gender	Age (years)	Years since diagnosis
Patient 1	Masculine	71	1
Patient 2	Masculine	69	5
Patient 3	Masculine	66	7
Patient 4	Masculine	56	3
Patient 5	Masculine	70	11
Patient 6	Masculine	54	3
Patient 7	Masculine	63	9
Patient 8	Masculine	58	5
Control 1	Masculine	60	-
Control 2	Masculine	74	-
Control 3	Masculine	60	-
Control 4	Masculine	63	-
Control 5	Masculine	66	-
Control 6	Masculine	59	-
Control 7	Masculine	74	-

A custom training dataset was used for the FER model. From an initial pool of 240,346 frames, a balanced subset of approximately 4000 frames per emotion was curated to mitigate the bias typically associated with the overrepresentation of neutral expressions. This dataset included 97 individuals (male and female, aged 18-65) to ensure demographic diversity. The model was trained to estimate the probabilities of the neutral expression and six basic emotions (joy, anger, disgust, fear, sadness, and surprise). The resulting model achieved an overall accuracy of 82.66% (López Aguilar, 2023). Table 2 details the final data distribution used for training. It should be noted that fear and surprise had a lower number of frames due to the spontaneous nature of these emotions.

The model was trained on a computer with an Intel Core i7-7700 (3.6 GHz) processor, 32 GB of RAM, and an Nvidia RTX 2080 GPU (6 GB VRAM). The software framework utilizes Python 3.9, TensorFlow 2.8.0, and Keras 2.8.0.

The MediaPipe face detection framework (Lugaresi et al., 2019) was utilized to localize the facial region within each video frame. This facilitated the pre-processing pipeline in which the detected face was cropped, resized to 224×224 pixels, and converted to the BGR color format to maintain compatibility with the VGG16 input layer.

One challenge faced during the experiment was the occurrence of extreme head poses (e.g. when participants check side monitors or perform wide maneuvers) which can lead to self-occlusion and classification errors. To ensure the reliability of the emotional signatures, a filtering protocol was implemented based on Euler rotation angles (pitch, yaw, and roll). A frame was considered valid for analysis only if the face orientation met the following criteria: 45° to 135° for pitch and yaw, and -30° to 30° for roll. These thresholds ensure that the model only processes frames where facial landmarks are sufficiently visible for accurate feature extraction. Fig. 3 illustrates these acceptable head pose ranges.

The output of the processing stage was exported as a series of CSV files containing frame-by-frame probabilities for the six basic emotions and the neutral state. To maintain the integrity of the emotional signatures and prevent the artificial dilution of intensity scores, frames that did not meet the head-pose inclusion criteria were flagged and excluded from the statistical analysis. The reported task-level means reflect only periods of valid facial tracking. The percentage of valid frames analyzed per task, relative to the total captured, is summarized in Table 3. Following the data processing stage, time-series visualizations were generated to illustrate the participants' emotional responses during the driving maneuvers. These graphs depict average emotion probabilities over time, facilitating a comparative analysis between the PD and control groups. Fig. 4 provides representative samples of these visualizations, comparing an individual with PD to a healthy control.

3.5. Statistical analysis approach

To compare the emotional responses between the PD group and the control group during each driving task, a statistical analysis was performed based on the mean emotion probabilities extracted from the processed videos. For each participant, four repetitions per task were conducted (except for Task 2), and the mean probability per emotion

(including the neutral expression) was calculated across valid frames. These averages were then used to obtain a single representative value per emotion and task, per participant.

It is important to clarify that the reported emotion probability values in the results tables do not sum to one. This is because each value represents the mean probability of a given emotion across all valid frames, calculated independently for each emotion category. While the FER model generates a probability distribution summing to one for each frame, these values are averaged per emotion across time, thus the resulting summary values are not constrained to form a unified distribution. This approach enables a more robust comparison of the prevalence of specific emotional states between groups.

3.5.1. Normality testing

Prior to the hypothesis testing, the Shapiro-Wilk test was used to determine whether the data distributions for each group-emotion-task combination followed a normal distribution. The null hypothesis of the test assumes normality. A p-value greater than 0.05 indicates that the data does not significantly deviate from a normal distribution, thereby validating the assumption required for parametric tests.

If the normality assumption was not met for either group, the non-parametric Mann-Whitney *U* test was used to compare their distributions without assuming normality.

All statistical tests were two-tailed with a significance threshold of $\alpha = 0.05$. Tests resulting in a p-value below 0.05 were considered statistically significant.

3.5.2. Student's T-test

The independent samples *t*-test compares the means of two independent groups (PD vs. control) and is calculated as:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (1)$$

Where:

- $\bar{X}_1, \bar{X}_2$: means of groups 1 and 2.
- s_1^2, s_2^2 : variances of groups 1 and 2.
- n_1, n_2 : samples sizes of group 1 and 2.

3.5.3. Mann-Whitney *U* test

The Mann-Whitney *U* test compares the distributions of two independent groups by ranking all values and evaluating the sum of ranks. The *U* statistic is calculated as:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \quad (2)$$

Where:

- n_1, n_2 : sample sizes of group 1 and group 2.
- R_1 : sum of the ranks of group 1.

The smaller *U* value between both groups is used for significance testing. If the resulting p-value is less than 0.05, the difference is considered statistically significant.

3.5.4. Power analysis

Power analysis is a statistical technique that determines the probability that a study will detect an effect (a significant difference between groups) if that effect truly exists in the population, for example: rejecting the null hypothesis when the alternative hypothesis is true. This probability, known as statistical power ($1 - \beta$), is conventionally set at 0.8 (or 80%) for research designs.

In this study, a post hoc analysis (or retrospective power analysis) was performed. Unlike the priori analysis, which is performed before

Table 2
Data distribution for custom VGG16 architecture training.

Emotion	Frames
Anger	4000
Disgust	4000
Fear	3934
Joy	4000
Sadness	4000
Surprise	3724
Neutral	4000

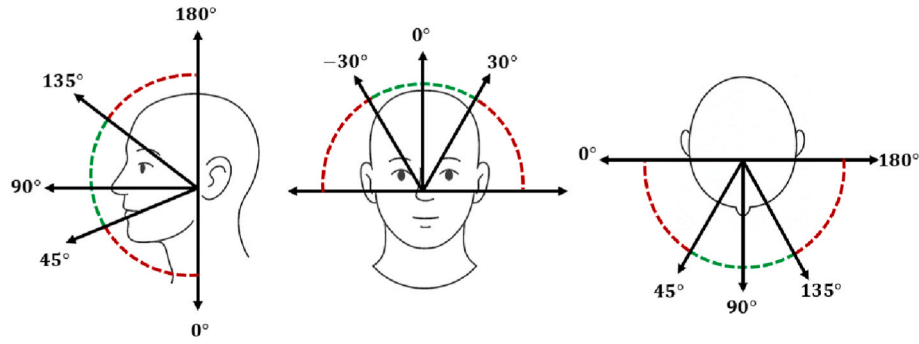


Fig. 3. Head pose estimation criteria for valid FER analysis. Only frames within the specified Euler angle ranges (pitch/yaw = $90^\circ \pm 45^\circ$; roll = $0^\circ \pm 30^\circ$) were included to ensure landmark visibility and classification accuracy.

Table 3

Percentage of valid frames analyzed per task (total number of captured frames is in parentheses).

Participant	Task 1	Task 2	Task 3	Task 4	Task 5-A	Task 5-B	Task 5-C	Task 6
Patient 1	93.57% (2676)	99.65% (8158)	98.93% (6388)	86.23% (6358)	96.86% (7264)	97.81% (8248)	99.31% (6964)	86.59% (9308)
Patient 2	100% (9308)	97.25% (9612)	100% (7524)	99.01% (5648)	99.09% (10,100)	92.52% (4496)	99.09% (6648)	89.58% (6376)
Patient 3	84.46% (3168)	86.10% (9392)	98.83% (4812)	83.58% (3792)	88.21% (5872)	77% (8540)	67.85% (6620)	75.39% (7284)
Patient 4	96.13% (3312)	83.15% (9140)	42.95% (4784)	75.72% (3088)	47.03% (4652)	75.96% (7424)	91.18% (5944)	99.34% (3692)
Patient 5	100% (3268)	99.58% (9528)	97% (7076)	71.48% (4432)	83.24% (6328)	90.13% (4664)	99.66% (6048)	87.28% (9724)
Patient 6	90.51% (1940)	97.22% (10,298)	99.03% (5784)	76.49% (5768)	97.63% (6088)	94.18% (9076)	94.22% (7000)	93.71% (5720)
Patient 7	53.28% (2880)	46.89% (9474)	51.07% (5224)	83.04% (4184)	45.18% (7466)	59.38% (7820)	52.53% (5604)	50.72% (6008)
Patient 8	40.15% (2260)	38.49% (9468)	99.04% (4604)	41.07% (4908)	94.28% (6084)	87.18% (5244)	97.39% (6304)	92.74% (7164)
Control 1	99.87% (3204)	99.97% (9254)	99.29% (5356)	93.06% (4324)	99.75% (6544)	97.92% (8100)	99.81% (6644)	99.89% (3688)
Control 2	96.76% (2472)	100% (9514)	100% (4956)	80.14% (4352)	99.11% (6316)	97.57% (5444)	97.49% (7992)	98.24% (5232)
Control 3	95.51% (2764)	100% (9500)	100% (4640)	99.79% (3836)	100% (5468)	99.52% (5088)	100% (3776)	100% (4116)
Control 4	92.14% (1476)	99.42% (9732)	97.31% (3720)	73.57% (3800)	97.41% (4332)	92.63% (4288)	99.72% (5908)	94.73% (3264)
Control 5	100% (2408)	100% (8848)	99.55% (6264)	80.75% (5632)	98.78% (6888)	100% (8792)	99.32% (6996)	100% (4576)
Control 6	100% (1744)	98.97% (9576)	99.45% (3656)	98.99% (3180)	100% (5200)	99.26% (4348)	100% (5620)	100% (4272)
Control 7	76.75% (3200)	79.60% (9522)	96.87% (4100)	81.75% (3824)	99.23% (5208)	96.22% (7096)	87.50% (3368)	95.26% (11,572)

data collection to determine the required sample size, the post hoc analysis is performed after obtaining the results and serves two main objectives with the final sample size ($N = 15$). The first objective is the validation of significant findings ($p\text{-value} < 0.05$); the post hoc analysis evaluates the achieved power to confirm that the sample size was sufficient to robustly detect the observed effect size (Cohen's d). High power (>0.8) validates the robustness of the findings. The second objective is the interpretation of non-significant findings; in contrast, for comparison that were not statistically significant ($p\text{-value} > 0.05$), the analysis is used to determine the study's achieved power to detect a hypothetical medium effect size (conventionally set at $d = 0.5$). Low power (<0.8) in this scenario suggests that the lack of significance may be due to insufficient statistical power (a high risk of Type II error) rather than the absence of a true effect on the population.

To calculate Cohen's d for observed differences, the pooled standard deviation (S_{pooled}) was first calculated:

$$S_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \quad (3)$$

Once the pooled standard deviation was obtained, Cohen's d was calculated as follows:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{S_{pooled}} \quad (4)$$

This entire analysis process was automated using a Python script that carried out the following tasks:

1. Loaded the csv files containing emotion probabilities.
2. Conducted normality checks.
3. Applied the appropriate statistical test based on the results.

4. Generated summary tables including mean, p -values, Cohen's d and significance flags.

4. Results

In our sample, most patients with PD recorded a score of 2/4 on the facial expression item of the UPDRS. Despite this, no differential features were detected between groups that would suggest overt reduced facial expression, though it is possible the study's sensitivity was influenced by this factor. Furthermore, all patients were in the ON state, which may have mitigated hypomimia; future studies must consider that more advanced PD stages might limit the effectiveness of FER systems. The tables below present the mean probability of each emotion for the PD and control groups. The "PD Group Mean" and "Control Group Mean" columns present the mean probability calculated across valid frames. The t -test and Mann-Whitney U were selected based on the results of the Shapiro-Wilk normality test.

4.1. Task 1 – straight-line driving

The initial driving task, maintaining a constant speed on a straight path, served as a baseline condition of low motor and cognitive demand. As shown in Table 4, no statically significant differences ($p > 0.05$) were observed between the PD and control groups across any emotional categories, including neutral expressions.

To further evaluate these findings, effect sizes (Cohen's d) were calculated to assess the magnitude of group differences beyond binary significance. Fear ($d = 0.5008$) and Neutrality ($d = -0.5327$) yielded medium effect sizes, suggesting emerging patterns in the PD group despite the lack of statistical significance.



Fig. 4. Comparison of PD participant and healthy participant – Experiment 1 - i.

Table 4

Statistical comparison of mean emotion probabilities between PD group and control group during baseline linear driving (Task 1). Metric probabilities represent frame-by-frame temporal averages calculated across the total analyzed frames, rather than normalized distributions sum to 1. Statistical comparisons were conducted using independent samples *t*-test or Mann-Whitney *U* test based on data normality distribution.

Emotion	PD Group Mean	Control Group Mean	Test	p-value comparison	Cohen's d	Significant difference
Anger	0.3881	0.0352	Mann-Whitney U	0.7789	0.1879	No
Disgust	0.0002	0.3485	Mann-Whitney U	0.2538	0.2538	No
Fear	0.0578	0.0001	Mann-Whitney U	0.3357	0.5008	No
Joy	0.0250	0.0024	Mann-Whitney U	0.4634	−0.3954	No
Neutral	0.0256	0.0178	Mann-Whitney U	0.3357	−0.5327	No
Sadness	0.1440	0.4047	T-test	0.8619	0.0918	No
Surprise	0.1526	0.1493	T-test	0.9365	−0.0420	No

Table 5

Statistical comparison of mean emotion probabilities between PD group and control group when following a vehicle (Task 2). Metric probabilities represent frame-by-frame temporal averages calculated across the total analyzed frames, rather than normalized distributions sum to 1. Statistical comparisons were conducted using independent samples *t*-test or Mann-Whitney *U* test based on data normality distribution.

Emotion	PD Group Mean	Control Group Mean	Test	p-value comparison	Cohen's d	Significant difference
Anger	0.4444	0.0664	T-test	0.3149	0.5410	No
Disgust	3.55E-05	0.3629	Mann-Whitney U	0.1893	0.0329	No
Fear	0.0404	0.0068	Mann-Whitney U	0.8665	0.4889	No
Joy	6.67E-05	0.0002	Mann-Whitney U	0.2810	−0.2489	No
Neutral	0.0004	0.4265	Mann-Whitney U	0.3357	−0.5222	No
Sadness	0.0365	0.5527	Mann-Whitney U	0.3357	−0.1297	No
Surprise	0.0769	0.0162	Mann-Whitney U	0.8665	−0.3925	No

4.2. Task 2 – following a vehicle in a straight line

Task 2 evaluated driver response while following a lead vehicle at variable speeds. Normality was assessed using Shapiro-Wilk test; consequently, a parametric *t*-test was employed for the “Anger” category, while the non-parametric Mann-Whitney *U* test was used for all other emotional categories. As shown in Table 5, there were no statistically significant differences ($p > 0.05$) in mean emotional probabilities between the PD and control groups for this maneuver.

Despite the lack of *p*-value significance, multiple medium-magnitude effect sizes ($|d| \approx 0.5$) were identified. Specifically, Anger ($d = 0.5410$), Neutrality ($d = -0.5222$), and Fear ($d = 0.4889$) exhibited clinical relevance, indicating divergent emotional trends between the groups that were not captured by binary significance testing alone.

4.3. Task 3 – turn 90° to the right and park

Task 3, compromising a 90° right turn followed by a precision parking maneuver, represented a high-demand scenario requiring complex decision-making and fine motor control. As detailed in Table 6, no statistically significant differences ($p > 0.05$) were observed between the PD and control groups in mean emotional probabilities.

However, effect size analysis revealed a large magnitude effect for Joy ($d = -0.8170$), despite the lack of *p*-value significance. The negative Cohen's *d* indicates a substantial suppression of joy in the PD group compared to the control group during the successful completion of the maneuver.

4.4. Task 4 – turn 90° to the left and park

Task 4, consisting of a 90° left turn followed by a parking maneuver, mirrored the high motor and spatial demands of Task 3. However, data acquisition for this scenario was affected by a technical constraint involving the spatial configuration of the simulator's left-side monitor. This setup required participants to perform head rotations that occasionally exceed the camera's optimal field of view, leading to frame loss. To maintain data integrity, a rigorous inclusion criterion was applied: only videos where valid frames accounted for at least 85% of the total duration were included in the analysis. Consequently, the results of Task were interpreted with caution, acknowledging the potential bias of the reduced dataset.

As detailed in Table 7, no statically significant differences ($p > 0.05$) were observed in mean emotional probabilities. However, effect size analysis identified substantial trends. Notably, Fear exhibited its strongest effect across all experimental tasks ($d = 0.5487$), while a moderate suppression of Joy ($d = -0.6153$) was also observed.

4.5. Task 5-A – enter a roundabout and exit (no traffic)

Task 5-A evaluated driver responses while entering and exiting a roundabout in the absence of surrounding traffic. This maneuver represents a scenario of moderate-to-high traffic. Although the absence of surrounding vehicles reduces immediate external stressors, this

maneuver still requires precise trajectory planning, speed estimation, and continuous coordination. As detailed in Table 8, no statistically significant differences ($p > 0.05$) were found in mean emotional probabilities between the PD and control groups.

Despite the lack of *p*-value significance, effect size analysis revealed a medium magnitude effect for Fear ($d = 0.5520$). Additionally, Disgust ($d = 0.4870$) approached the threshold for a medium effect. These results suggest that the internal demands of executing multi-step maneuvers trigger distinct emotional trends in the PD population, even when external traffic stressors are removed.

4.6. Task 5-B – enter a roundabout and exit (another vehicle is entering to the roundabout)

Task 5-B, involving the entry and exit of a roundabout with a vehicle entering, represented the most cognitively and motorically demanding scenario in this study. This task required simultaneous risk perception, conflict management, and rapid decision-making.

As detailed in Table 9, this experiment yielded the most critical finding: a statistically significant difference in the probability of Fear between the PD and control groups ($p = 0.0177$).

Notably, this significance was accompanied by an exceptionally large effect size ($d = 1.4830$), the highest observed across all maneuvers. While other categories, such as Disgust ($d = 0.4408$) and Joy ($d = -0.4135$), showed moderate effect magnitudes, they did not reach statistical significance, likely due to the study's restricted statistical power (14.59% for $d = 0.5$). These results confirm that emotional disparities in PD drivers become statistically manifest primarily under high-pressure conditions that challenge executive control.

4.7. Task 5-C – enter a roundabout and exit (another vehicle is already in the roundabout)

Task 5-C involved entering and exiting a roundabout with another vehicle already present, requiring the driver to yield and precisely estimate entry and exit timing. This maneuver presented the highest level of external interference in the study. As detailed in Table 10, no statically significant differences ($p > 0.05$) were observed between the PD and control groups in mean probabilities.

However, effect size analysis revealed the strongest evidence of group disparities in the study, with four emotional categories reaching medium-to-large magnitudes ($|d| \geq 0.5$). Notably, Joy ($d = -0.9998$) exhibited a large effect size, while Neutrality ($d = -0.6369$), Surprise ($d = -0.6459$), and Anger ($d = 0.5183$) all maintained medium-to-large effects. The simultaneous occurrence of these four effects highlights the clinical relevance of emotional signatures in high-conflict driving scenarios, despite the study being underpowered for formal significance.

4.8. Task 6 – reversing

Task 6, compromising a 10 m reverse driving maneuver, presented a significant technical challenge for FER-based monitoring. The nature of this task required participants to rotate their heads to check side and

Table 6

Statistical comparison of mean emotion probabilities between PD group and control group while making a turn of 90° to the right and park (Task 3). Metric probabilities represent frame-by-frame temporal averages calculated across the total analyzed frames, rather than normalized distributions sum to 1. Statistical comparisons were conducted using independent samples *t*-test or Mann-Whitney *U* test based on data normality distribution.

Emotion	PD Group Mean	Control Group Mean	Test	p-value comparison	Cohen's d	Significant difference
Anger	0.0756	0.0097	Mann-Whitney U	0.3969	0.2839	No
Disgust	0.0002	0.0129	Mann-Whitney U	0.3969	0.0456	No
Fear	0.0799	2.83E-05	Mann-Whitney U	0.4634	-0.0717	No
Joy	8.28E-05	0.0442	Mann-Whitney U	0.1520	-0.8170	No
Neutral	0.0150	0.4239	Mann-Whitney U	0.4634	-0.2772	No
Sadness	0.3460	0.5205	T-test	0.7287	0.1834	No
Surprise	0.3846	0.0938	T-test	0.3499	-0.5019	No

Table 7

Statistical comparison of mean emotion probabilities between PD group and control group during making a turn of 90° to the left and park (Task 4). Metric probabilities represent frame-by-frame temporal averages calculated across the total analyzed frames, rather than normalized distributions sum to 1. Statistical comparisons were conducted using independent samples *t*-test or Mann-Whitney *U* test based on data normality distribution.

Emotion	PD Group Mean	Control Group Mean	Test	p-value comparison	Cohen's d	Significant difference
Anger	0.8052	0.0263	Mann-Whitney U	0.3969	0.2069	No
Disgust	2.24E-05	0.8642	Mann-Whitney U	0.1893	0.1123	No
Fear	0.0852	5.76E-05	Mann-Whitney U	0.6126	0.5487	No
Joy	0.0001	0.0055	Mann-Whitney U	0.1520	−0.6153	No
Neutral	0.0016	0.6318	Mann-Whitney U	0.3969	−0.2228	No
Sadness	0.1970	0.2698	T-test	0.9445	−0.0367	No
Surprise	0.1166	0.4700	T-test	0.9753	0.0164	No

Table 8

Statistical comparison of mean emotion probabilities between PD group and control group during entering a roundabout with no traffic and exiting (Task 5-A). Metric probabilities represent frame-by-frame temporal averages calculated across the total analyzed frames, rather than normalized distributions sum to 1. Statistical comparisons were conducted using independent samples *t*-test or Mann-Whitney *U* test based on data normality distribution.

Emotion	PD Group Mean	Control Group Mean	Test	p-value comparison	Cohen's d	Significant difference
Anger	0.3779	0.0229	Mann-Whitney U	0.3829	0.0257	No
Disgust	0.0001	0.3524	Mann-Whitney U	0.8048	0.4870	No
Fear	0.8202	0.0006	Mann-Whitney U	0.1282	0.5520	No
Joy	0.0106	0.0367	Mann-Whitney U	0.6200	−0.1873	No
Neutral	0.0003	0.8348	Mann-Whitney U	0.5350	−0.3152	No
Sadness	0.3085	0.4683	T-test	0.9633	−0.0251	No
Surprise	0.0597	0.3615	T-test	0.7203	0.1960	No

Table 9

Statistical comparison of mean emotion probabilities between PD group and control group during entering and exiting a roundabout while another vehicle is entering (Task 5-B). Metric probabilities represent frame-by-frame temporal averages calculated across the total analyzed frames, rather than normalized distributions sum to 1. Statistical comparisons were conducted using independent samples *t*-test or Mann-Whitney *U* test based on data normality distribution.

Emotion	PD Group Mean	Control Group Mean	Test	p-value comparison	Cohen's d	Significant difference
Anger	0.7135	0.1304	T-test	0.6153	0.2786	No
Disgust	0.0006	0.9526	Mann-Whitney U	0.9497	0.4408	No
Fear	0.7687	0.3791	T-test	0.0177	1.4830	Yes
Joy	0.0004	0.4843	Mann-Whitney U	0.5728	−0.4135	No
Neutral	0.0003	0.3537	Mann-Whitney U	0.4908	−0.3508	No
Sadness	0.1753	0.7115	T-test	0.8481	−0.1057	No
Surprise	0.1398	0.0079	Mann-Whitney U	0.7546	−0.2720	No

Table 10

Statistical comparison of mean emotion probabilities between PD group and control group during entering and exiting a roundabout while another vehicle is already in the roundabout (Task 5-C). Metric probabilities represent frame-by-frame temporal averages calculated across the total analyzed frames, rather than normalized distributions sum to 1. Statistical comparisons were conducted using independent samples *t*-test or Mann-Whitney *U* test based on data normality distribution.

Emotion	PD Group Mean	Control Group Mean	Test	p-value comparison	Cohen's d	Significant difference
Anger	0.9027	0.0854	T-test	0.3349	0.5183	No
Disgust	0.0016	0.4987	Mann-Whitney U	0.6126	0.4127	No
Fear	0.2342	0.0002	Mann-Whitney U	0.5358	−0.0301	No
Joy	0.0116	0.1875	Mann-Whitney U	0.2810	−0.9998	No
Neutral	6.08E-05	0.2838	Mann-Whitney U	0.4634	−0.6369	No
Sadness	0.3252	0.7409	T-test	0.5669	0.3941	No
Surprise	0.1633	0.0961	T-test	0.2340	−0.6459	No

Table 11

Statistical comparison of mean emotion probabilities between PD group and control group for a reversing maneuver (Task 6). Metric probabilities represent frame-by-frame temporal averages calculated across the total analyzed frames, rather than normalized distributions sum to 1. Statistical comparisons were conducted using independent samples *t*-test or Mann-Whitney *U* test based on data normality distribution.

Emotion	PD Group Mean	Control Group Mean	Test	p-value comparison	Cohen's d	Significant difference
Anger	0.8688	0.0584	T-test	0.8038	0.1358	No
Disgust	0.0392	0.0012	Mann-Whitney U	0.8048	0.1128	No
Fear	0.1176	0.0015	Mann-Whitney U	0.4557	0.1337	No
Joy	0.0279	0.0002	Mann-Whitney U	0.2086	−0.5865	No
Neutral	0.0160	0.3782	Mann-Whitney U	0.3829	−0.4138	No
Sadness	0.2592	0.9310	T-test	0.5140	0.3595	No
Surprise	0.3630	0.1842	T-test	0.3505	−0.5193	No

rearview mirrors, frequently moving their faces outside the camera's observable field of view. Consequently, only videos meeting a strictly defined threshold of valid frames were included, and the following results are based on this restricted dataset.

As detailed in Table 11, statistical analysis of the valid data revealed no significant differences ($p > 0.05$) between PD and control groups. However, effect size analysis confirmed the persistence of emotional patterns observed in other high-demand maneuvers. Joy ($d = -0.5865$) exhibited medium-magnitude effects, while Neutrality ($d = -0.4138$) approached the medium threshold. Distinctively, Fear ($d = 0.1337$) and Anger ($d = 0.1358$) showed only trivial effects in this specific scenario.

5. Discussion

The experiment was designed to collect data from individuals diagnosed with PD and from an age-matched healthy control group, while they performed a series of simulated driving tasks. The aim was to observe and compare emotional response under controlled driving conditions.

5.1. Addressing statistical power and type II error

A critical takeaway across these tasks is the prevalence of medium-to-large effect sizes despite non-significant p -values. Given the restricted sample size ($N = 15$) and low statistical power (14.59% for $d = 0.5$), the group comparisons reported herein, except for Task 5-B, should be interpreted as preliminary exploratory trends rather than conclusive evidence of population-level differences. The limited statistical power increases the risk of Type II errors, which may have obscured underlying differences between the groups that could become apparent in larger cohorts. Therefore, the observed effect sizes should be considered indicative of potential group differences that warrant further investigation with larger sample sizes.

5.2. Neurological and clinical interpretations

The observed reduction in Joy magnitude ($d = -0.8170$ in Task 3; $d = -0.9998$ in Task 5-C) may suggest an interaction between facial hypomimia and dysfunction within dopaminergic reward pathways, which are known to modulate positive affect processing in PD (Bianchini et al., 2024; Rodríguez-Antigüedad et al., 2024). This attenuated reward for successful maneuver completion could contribute to driving-related fatigue over time; however, this mechanism remains speculative and warrants further analysis.

Elevated Anger and Fear effect sizes across demanding tasks (Task 2, 4, and 5-B) align with previous clinical literature reporting heightened emotional stress and executive overload when PD patients navigate high-complexity environments (Ranchet et al., 2016, 2020). While these findings highlight the potential of FER-derived measures as candidate digital biomarkers for cognitive-emotional stress, this framing should be interpreted cautiously as a preliminary trend given the exploratory nature of our dataset.

The lower surprise probabilities observed during high-conflict maneuvers (Task 5-C) might reflect underlying deficits in attentional set-shifting and orienting responses, which have been documented in PD cohorts (Mandal & Khan, 2023; Sawada et al., 2012; Stefanova et al., 2014). However, whether this pattern stems from impaired cognitive flexibility or facial motor restriction remains an open question for future experiments.

5.3. Implications for affective-adaptive ADAS

While these preliminary results suggest that automated FER tools could eventually complement traditional vehicle-telemetry metrics by tracking internal affective load, further validation in real-world driving conditions and larger clinical cohorts is required before deploying such

affective-adaptive ADAS framework.

5.4. Technical limitations

Despite the promising results, this study has several limitations that must be acknowledged to contextualize the findings:

- **Statistical power and sample size:** The small sample size resulted in a post-hoc power analysis of approximately 15%, which is a major limitation for detecting subtle emotional effects. Consequently, this study should be considered exploratory, and results should be interpreted with caution, as they may not be generalizable beyond this specific cohort.
- **Hypomimia in PD drivers:** A significant challenge in FER-based research with patients with PD is hypomimia, which may persist even in "ON-medication" states. This clinical feature could act as a bias, potentially leading the FER system to underestimate the intensity of the PD driver's emotional state.
- **Data aggregation:** The use of averaged emotion probabilities across tasks may have obscured transient, high-frequency emotional dynamics that occur during rapid driving maneuvers.
- **Validation in real scenarios:** While the driving simulator offers a controlled and safe environment, the findings may not fully translate to real-world driving conditions where the perceived risk and consequences are high-frequency emotional dynamics that occur during rapid driving maneuvers.

6. Conclusion

The study implemented a computer vision-based FER system as a non-intrusive monitoring tool to assess the emotional responses of patients with PD during a series of simulated driving scenarios. Statistical analysis revealed distinct emotional patterns between the PD group and the healthy control group, with the most notable differences observed in high-demand scenarios involving the introduction of other vehicles.

The hypothesis that emotional differences would manifest under high-stress conditions was validated by the activation of fear; the only difference that reached statistical significance was the probability of fear in task 5-B (enter in a roundabout and exit when another vehicle is entering). This massive effect, with an exceptionally large effect size ($d = 1.438$), demonstrated that the PD group experienced significantly greater anxiety or risk perception when faced with rapid decision-making and traffic conflict.

Although most comparisons were not statistically significant, post hoc analysis revealed that the study's low statistical power (only 14.59% to detect a medium-magnitude effect). This low power consistently masked large-magnitude effects that indicate the existence of relevant differences such as affective suppression and chronic negative activation. In the case of affective suppression, the largest effects were observed for joy and neutrality (with a peak of $d = -0.9998$ and $d = -0.4138$ in task 5-C, respectively). This pattern suggests that the PD group experienced marked suppression or attenuation of the achievement or satisfaction response upon successfully completing a difficult maneuver, in addition to maintaining a persistent state of arousal (lower neutrality). In the case of chronic negative activation, the medium and consistent effects on fear and anger on moderate and high demand tasks reinforce that motor difficulty and sustained cognitive effort translate into increased frustration and perceived anxiety in the PD group.

The findings of this study demonstrate the feasibility of using non-invasive FER systems to monitor the affective state of drivers with neurodegenerative diseases. However, there are limitations that must be addressed for future research due to the dynamic nature of driving. In tasks requiring large head movements, substantial data loss occurred as participants' faces moved out of the camera's field of view. This introduces potential bias and may not have covered moments of peak emotional arousal in the sample analyzed. Intrinsically, FER systems

may be limited when assessing patients with PD due to the hypomimia associated with the disease. The reduction in the amplitude of facial expressions in the PD group may result in an underestimation of the true emotional intensity, thus reinforcing the validity of the large effects observed, as these were detected despite the facial motor restriction.

Finally, the main statistical limitation was the small sample size ($N = 15$), which resulted in a low power and a high risk of Type II error. Future research should focus on replicating these findings with a significantly larger sample size to formally validate the large-magnitude patterns observed.

Despite the limitations, the results of this study highlight the relevance of affective monitoring for future ADAS. An ADAS optimized for drivers with PD should be able to detect spikes in fear or a decrease in neutrality in order to intervene preventively, improving safety and the driving experience.

CRedit authorship contribution statement

Ariel A. Lopez-Aguilar: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Juan Felipe Medina-Lee:** Software, Methodology, Investigation. **Jorge Villagra:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition. **Eduardo Rocon:** Supervision, Resources, Methodology, Investigation. **Almudena Cerezo-Zaruelo:** Supervision, Resources, Methodology, Investigation. **Juan Pablo Romero:** Supervision, Resources, Methodology, Investigation. **M. Rogelio Bustamante-Bello:** Writing – review & editing, Supervision. **Sergio A. Navarro-Tuch:** Writing – review & editing, Validation, Supervision, Conceptualization.

Ethics declaration

Written informed consent to take part in the study and to publish the article has been obtained from all participants or their legal representatives. The privacy rights of participants have been observed.

This study was performed in compliance with relevant laws, regulatory frameworks and guidelines where the research took place. This study was approved by the Comité de Ética de la Investigación con Medicamentos del Hospital Universitario 12 de Octubre. (Approval No. 23/603)

Funding

This work was supported by Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), Mexico under grant number A00399843, and by the DrivABLE-CM project (Neurorehabilitation and Autonomous Driving Integration for Guidance and Assessment Tools), funded by the Comunidad de Madrid under grant SYG-2024/TEC-634, within the 2024 call for Synergistic R&D Projects.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Hospital Beata María Ana for their collaboration and for providing access to Parkinson's patients. Special thanks to all participants. We also thank Centre for Automation and Robotics (CAR), CSIC – Universidad Politécnica de Madrid, and Tecnológico de Monterrey – Mexico City Campus.

Data availability

Data will be made available on request.

References

- Abrami, A., Gunzler, S., Kilbane, C., Ostrand, R., Ho, B., & Cecchi, G. (2021). Automated computer vision assessment of hypomimia in parkinson disease: Proof-of-Principle pilot Study. *Journal of Medical Internet Research*, 23(2), Article e21037. <https://doi.org/10.2196/21037>
- Aksoy, B., & Sayin, İ. (2021). *Emotion analysis using deep learning methods*. https://doi.org/10.1007/978-3-030-79357-9_65
- Álvarez, F. J. (2016). Parkinson's disease, antiparkinson medicines, and driving. *Expert Review of Neurotherapeutics*, 16(9), 1023–1032. <https://doi.org/10.1080/14737175.2016.1218278>
- Argaud, S., Vérin, M., Sauleau, P., & Grandjean, D. (2018). Facial emotion recognition in Parkinson's disease: A review and new hypotheses. *Movement Disorders*, 33(4), 554–567. <https://doi.org/10.1002/mds.27305>
- Bianchini, E., Rinaldi, D., Alborghetti, M., Simonelli, M., D'Audino, F., Onelli, C., Pegolo, E., & Pontieri, F. E. (2024). The story behind the mask: A narrative review on Hypomimia in parkinson's disease. *Brain Sciences*, 14(1), 109. <https://doi.org/10.3390/brainsci14010109>
- Brock, P., Oates, L. L., Gray, W. K., Henderson, E. J., Mann, H., Haunton, V. J., Skelly, R., Hand, A., Davies, M. L., & Walker, R. W. (2022). Driving and Parkinson's Disease: A Survey of the patient's perspective. *Journal of Parkinson's Disease*, 12(1), 465–471. <https://doi.org/10.3233/JPD-212686>
- Canedo, D., & Neves, A. J. R. (2019). Facial expression recognition using computer vision: A systematic review. *Applied Sciences (Switzerland)*, 9(21), 4678. <https://doi.org/10.3390/app9214678>
- Carvalho, V., Cunha, C. V., & Massano, J. (2018). Parkinson's disease: Contemporary concepts and clinical management. In *Neurodegenerative diseases* (pp. 349–378). Springer International Publishing. https://doi.org/10.1007/978-3-319-72938-1_17
- Chang, J. H., Bhatti, D., Uc, E. Y., Rizzo, M., & Merickel, J. (2024). Digital driving data can track driving exposure and quality of life in Parkinson's disease. *Traffic Injury Prevention*, 25(1), 20–26. <https://doi.org/10.1080/15389588.2023.2247110>
- Chen, J., Dey, S., Wang, L., Bi, N., & Liu, P. (2024). Attention-Based multi-modal multi-view fusion approach for driver facial expression recognition. *IEEE Access*, 12 (August), 137203–137221. <https://doi.org/10.1109/ACCESS.2024.3462352>
- Clark, E. A., Kessinger, J., Duncan, S. E., Ann Bell, M., Lahne, J., Gallagher, D. L., & O'Keefe, S. F. (2020). The facial action coding system for characterization of human affective response to consumer product-based stimuli: A systematic review. *Frontiers in Psychology*, 11(May), 1–21. <https://doi.org/10.3389/fpsyg.2020.00920>
- Classen, S., Winter, S., Wei, J., Jeghers, M., Rogers, J., & Giang, W. (2022). Feasibility of automated in-vehicle technologies on volunteers' driving performance. *Technology and Disability*, 34(4), 233–246. <https://doi.org/10.3233/TAD-220374>
- Cossini, F., Cuesta, C., Román, C., Zambrano, S., Rubinstein, W., & Politis, D. (2024). Relationship between severity of hypomimia and basic emotion recognition in Parkinson's disease. *Revista de Neurología*, 79(3), 71–76. <https://doi.org/10.33588/rn.7903.2024169>
- Crizzle, A. M., Classen, S., & Uc, E. Y. (2012). Parkinson's disease and driving. *Neurology*, 79(20), 2067–2074. <https://doi.org/10.1212/WNL.0b013e3182749e95>
- Dorsey, E. R., Constantinescu, R., Thompson, J. P., Biglan, K. M., Holloway, R. G., Kieburtz, K., Marshall, F. J., Ravina, B. M., Schifitto, G., Siderowf, A., & Tanner, C. M. (2007). Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology*, 68(5), 384–386. <https://doi.org/10.1212/01.wnl.0000247740.47667.03>
- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3–4), 169–200. <https://doi.org/10.1080/02699939208411068>
- Ekman, P., & Rosenberg, E. L. (2012). What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). In *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195179644.001.0001>
- Fahn, S., & Elton, R. L. (1987). Unified parkinson's disease rating Scale. In S. Fahn, C. D. Marsden, D. Calne, & M. Goldstein (Eds.), *Recent developments in parkinson's disease* (pp. 153–163). Macmillan Health Care Information.
- Forsman, Å., Anund, A., Skyving, M., & Filtneis, A. J. (2021). Injury crashes and the relationship with disease causing excessive daytime sleepiness. *Traffic Injury Prevention*, 22(4), 272–277. <https://doi.org/10.1080/15389588.2021.1894639>
- Gotardi, G. C., Barbieri, F. A., Simão, R. O., Pereira, V. A., Baptista, A. M., Imaizumi, L. F., Moretto, G., Navarro, M., Polastri, P. F., & Rodrigues, S. T. (2022). Parkinson's disease affects gaze behaviour and performance of drivers. *Ergonomics*, 65(9), 1302–1311. <https://doi.org/10.1080/00140139.2022.2028901>
- Izquierdo-Reyes, J., Ramirez-Mendoza, R. A., Bustamante-Bello, M. R., Navarro-Tuch, S., & Avila-Vazquez, R. (2018). Advanced driver monitoring for assistance system (ADMAS). *International Journal on Interactive Design and Manufacturing*, 12(1), 187–197. <https://doi.org/10.1007/s12008-016-0349-9>
- Jacobs, M., Hart, E. P., & Roos, R. A. C. (2017). Driving with a neurodegenerative disorder: An overview of the current literature. *Journal of Neurology*, 264(8), 1678–1696. <https://doi.org/10.1007/s00415-017-8489-9>
- Jeong, M., & Ko, B. C. (2018). Driver's facial expression recognition in real-time for safe driving. *Sensors*, 18(12), 4270. <https://doi.org/10.3390/s18124270>
- Kalia, L. V., & Lang, A. E. (2015). Parkinson's disease. *The Lancet*, 386(9996), 896–912. [https://doi.org/10.1016/S0140-6736\(14\)61393-3](https://doi.org/10.1016/S0140-6736(14)61393-3)

- Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, 234(November 2016), 11–26. <https://doi.org/10.1016/j.neucom.2016.12.038>
- López Aguilar, A. A. (2023). *Emotional Domotics: Development of a Methodology for an Objective Evaluation of User Experience in Inhabitable Spaces* [Instituto Tecnológico y de Estudios Superiores de Monterrey]. <https://hdl.handle.net/11285/703453>.
- Lopez-aguilar, A. A., Bustamante-bello, M. R., Navarro-tuch, S. A., & Molina, A. (2023). Development of a framework for the communication System based on KNX for an interactive System for UX evaluation. *Sensors*, 23(23), 9570. <https://doi.org/10.3390/s23239570>
- Lopez-Aguilar, A. A., Bustamante-Bello, R., & Navarro-Tuch, S. A. (2021). Advanced system to measure UX in online learning environments. In *IEEE global engineering education conference, EDUCON* (pp. 774–777). <https://doi.org/10.1109/EDUCON46332.2021.9454083>, 2021-April(April).
- Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., Chang, W.-T., Hua, W., Georg, M., & Grundmann, M. (2019). MediaPipe: A framework for perceiving and processing reality. *Google Research*, 1–4. <https://github.com/google/mediapipe>.
- Macphee, G. (2018). Parkinson's disease: Pathogenesis and clinical aspects. In T. B. Stoker, & J. C. Greenland (Eds.), *Parkinson's disease in the older patient* (2nd ed.). Codon Publications. <https://doi.org/10.15586/codonpublications.parkinsonsdisease.2018>.
- Madeley, P., Hulley, J. L., Wildgust, H., & Mindham, R. H. (1990). Parkinson's disease and driving ability. *Journal of Neurology, Neurosurgery & Psychiatry*, 53(7), 580–582. <https://doi.org/10.1136/jnnp.53.7.580>
- Mandal, M., & Khan, A. (2023). Attention switching deficit in patients of Parkinson's disease who experience freezing of gait. *Applied Neuropsychology: Adult*, 30(4), 389–400. <https://doi.org/10.1080/23279095.2021.1951268>
- Marano, M., Esposito, M., Sergi, G., Proietti, F., Bonura, A., Toro, S., Magliozzi, A., Anzini, G., & Di Lazzaro, V. (2024). Driving abilities and wearing-off in parkinson's disease: A driving simulation Study. *Brain Sciences*, 14(11). <https://doi.org/10.3390/brainsci14111072>
- Massano, J., & Bhatia, K. P. (2012). Clinical approach to parkinson's disease: Features, diagnosis, and principles of management. *Cold Spring Harbor Perspectives in Medicine*, 2(6), Article a008870. <https://doi.org/10.1101/cshperspect.a008870>. a008870.
- Medina-Lee, J. F., Godoy, J., Artunedo, A., & Villagra, J. (2023). Speed profile generation strategy for efficient merging of automated vehicles on roundabouts with realistic traffic. *IEEE Transactions on Intelligent Vehicles*, 8(3), 2448–2462. <https://doi.org/10.1109/TIV.2022.3211350>
- Phokaewvarangkul, O., Krootjohn, S., Yanthitirat, P., Anan, C., & Bhidayasiri, R. (2019). Objective monitoring of driving behavior in parkinson's disease: The utility of the chula parkinson car®. *European Neurology*, 81(3–4), 128–138. <https://doi.org/10.1159/000500953>
- Poewe, W., Seppi, K., Tanner, C. M., Halliday, G. M., Brundin, P., Volkman, J., Schrag, A.-E., & Lang, A. E. (2017). Parkinson disease. *Nature Reviews Disease Primers*, 3(1), Article 17013. <https://doi.org/10.1038/nrdp.2017.13>
- Postuma, R. B., & Berg, D. (2016). Advances in markers of prodromal Parkinson disease. *Nature Reviews Neurology*, 12(11), 622–634. <https://doi.org/10.1038/nrneurol.2016.152>
- Ranchet, M., Broussolle, E., & Paire-Ficout, L. (2016). Longitudinal executive changes in drivers with parkinson's disease: Study using neuropsychological and driving simulator tasks. *European Neurology*, 76(3–4), 143–150. <https://doi.org/10.1159/000448992>
- Ranchet, M., Devos, H., & Uc, E. Y. (2020). Driving in parkinson disease. *Clinics in Geriatric Medicine*, 36(1), 141–148. <https://doi.org/10.1016/j.cger.2019.09.007>
- Reijnders, J. S. A. M., Ehrt, U., Weber, W. E. J., Aarsland, D., & Leentjens, A. F. G. (2008). A systematic review of prevalence studies of depression in Parkinson's disease. *Movement Disorders*, 23(2), 183–189. <https://doi.org/10.1002/mds.21803>
- Rodríguez-Antigüedad, J., Martínez-Horta, S., Horta-Barba, A., Puig-Davi, A., Campolongo, A., Sampedro, F., Bejr-Kasem, H., Marín-Lahoz, J., Pagonabarraga, J., & Kulisevsky, J. (2024). Facial emotion recognition deficits are associated with hypomimia and related brain correlates in Parkinson's disease. *Journal of Neural Transmission*, 131(12), 1463–1469. <https://doi.org/10.1007/s00702-023-02725-3>
- Sawada, Y., Nishio, Y., Suzuki, K., Hirayama, K., Takeda, A., Hosokai, Y., Ishioka, T., Itoyama, Y., Takahashi, S., Fukuda, H., & Mori, E. (2012). Attentional set-shifting deficit in parkinson's disease is associated with prefrontal dysfunction: An FDG-PET Study. *PLoS One*, 7(6), Article e38498. <https://doi.org/10.1371/journal.pone.0038498>
- Siebert, C. (2018). Neurodegenerative diseases. In D. Galimberti, & E. Scarpini (Eds.), *Innere medizin (Germany)* (Vol. 64). Springer International Publishing. <https://doi.org/10.1007/978-3-319-72938-1>. Number 2.
- Skaramagkas, V., Boura, I., Karamanis, G., Kyprakis, I., Fotiadis, D. I., Kefalopoulou, Z., Spanaki, C., & Tsiknakis, M. (2025). Dual stream transformer for medication state classification in Parkinson's disease patients using facial videos. *npj Digital Medicine*, 8(1), 1–11. <https://doi.org/10.1038/s41746-025-01630-1>
- Stefanova, E., Ječmenica Lukić, M., Žiropadja, L., Marković, V., Stojković, T., Tomić, A., Dragasović Mišković, N., & Kostić, V. (2014). Attentional set-shifting in parkinson's disease patients with freezing of gait-acquisition and discrimination set learning deficits at the background? *Journal of the International Neuropsychological Society*, 20(9), 929–936. <https://doi.org/10.1017/S1355617714000769>
- Taghi Zadeh, M. M., Imani, M., & Majidi, B. (2019). Fast facial emotion recognition using convolutional neural networks and gabor filters. In *2019 5th conference on knowledge based engineering and innovation (KBEI)* (pp. 577–581). <https://doi.org/10.1109/KBEI.2019.8734943>
- Thompson, T., Poulter, D., Miles, C., Solmi, M., Veronese, N., Carvalho, A. F., Stubbs, B., & Uc, E. Y. (2018). Driving impairment and crash risk in Parkinson disease. *Neurology*, 91(10), e906–e916. <https://doi.org/10.1212/WNL.0000000000006132>
- Tozman, T., Magdas, E. S., MacDougall, H. G., & Vollmeyer, R. (2015). Understanding the psychophysiology of flow: A driving simulator experiment to investigate the relationship between flow and heart rate variability. *Computers in Human Behavior*, 52, 408–418. <https://doi.org/10.1016/j.chb.2015.06.023>
- Videla, L. S., Rao, M. R. N., Anand, D., Vankayalapati, H. D., & Razia, S. (2019). Deformable facial fitting using active appearance model for emotion recognition. In S. C. Satapathy, V. Bhateja, & S. Das (Eds.), *Smart innovation, systems and technologies* (Vol 104, pp. 135–144). Singapore: Springer. https://doi.org/10.1007/978-981-13-1921-1_13
- Williams-Gray, C. H., & Worth, P. F. (2016). Parkinson's disease. *Medicine*, 44(9), 542–546. <https://doi.org/10.1016/j.mpmed.2016.06.001>