Early-Stage Parkinson's Disease Detection Based on Action Unit Derivatives

Anas Filali Razzouki[1], Laetitia Jeancolas[2], Graziella Mangone[3], Sara Sambin[3], Alizé Chalançon[3], Manon Gomes[3], Stéphane Lehéricy[3], Jean-Christophe Corvol[3], Marie Vidailhet[3], Isabelle Arnulf[3], Mounim A. El-Yacoubi[1], Dijana Petrovska-Delacrétaz[1]

[1] Laboratoire SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, France

[2] Electrical & Computer Engineering Department, Concordia University, Montreal, Canada

[3] Sorbonne Université, Paris Brain Institute - ICM, Inserm, CNRS, APHP, Hôpital Pitié-Salpêtrière, Paris, France

Corresponding author: anas.filali_razzouki@telecom-sudparis.eu

Abstract

Background: Hypomimia is a symptom of Parkinson's disease (PD), characterized by a decrease in facial movements and loss of face emotional expressions. This study aims to detect hypomimia in participants with early-stage PD based on facial action units (AUs). Methods: A total of 299 video recordings were included, consisting of 208 PD subjects and 91 healthy controls (HC), asked to perform fast syllable repetitions. To distinguish typical facial muscle movements from PD subjects associated with hypomimia, we compute the AUs derivatives. Global features were extracted based on the AUs intensities and their derivatives, and XGBoost was used to classify PD vs. HC. Results: We obtain classification scores up to 73.00% in terms of balanced accuracy (BA) and an area under the curve (AUC) of 78.38% at video visit level. These results are promising for detecting hypomimia at an early stage of PD, and this work could potentially allow for continuous monitoring of hypomimia outside of hospitals through telemedicine.

Keywords— Hypomimia, Early Parkinson's Disease, Facial Action Units, XGBoost

1 Introduction

Context and motivation: Parkinson's disease (PD), first described by James Parkinson in 1817, is the second most common neurodegenerative disease, affecting 1% of people over 60 years [1]. PD affects the central nervous system by the destruction of dopaminergic neurons in the substantia nigra [2]. PD causes motor deficits like rigidity, bradykinesia and rest tremor, and non-motor ones like depression, anxiety and dysautonomia [3]. These symptoms occur years after disease onset [4] and by the time PD is diagnosed, 60% of the dopaminergic neurons are already lost [5]. Early-stage PD detection [6], therefore, is key to test treatments before large

irreversible brain damages occur, and to slow down, or even stop, its progression. A common, often-early stage, PD symptom [7] is hypomimia, or masked face, reflecting a decrease in facial movement and loss of face emotional expressions. This loss has negative social consequences on patients, who may face social rejection [8]. Hypomimia is also used in the MDS-UPDRS scale [9] by neurologists to track PD progression.

To quantify facial muscles' movement and detect hypomimia, participants are proposed different scenarios. Some evoke their emotional side by asking them to make facial expressions like smile, anger, surprise, disgust [10, 11] upon clinician's request, by imitating emotive faces shown on screen, or by eliciting spontaneous emotions as in [12] where the participants are shown a movie. Others evoke non-emotional facial movements by asking the subjects to make different facial gestures (close eyes, moving the eyebrows, look down) [13], and recording them talking about a positive and negative experience [14]. Hypomimia can be detected by single face image-based methods [15, 16], and face video-based approaches [17, 18]. Electromyography (EMG) signals can also be used by positioning small electrodes on the skin above the facial muscles to capture the electrical signals created by muscle movements [12].

Video-based state-of-the-art approaches can be split into two categories. The first extracts motion features from facial landmarks [10, 17, 19] or facial action units (AUs) intensities [14, 20] and uses them as input to shallow classifiers or statistical tests to discriminate PD from healthy control (HC) subjects. The second extracts the features by Convolutional Neural Networks (CNNs), that are then fed as input to shallow classifiers [16, 21], or directly to the CNN for classification [18, 22].

The first type of facial landmarks-based approaches compute changes in facial landmarks across frames. Concretely, global parameters quantifying the trajectory movement of these key points across the frames are extracted. On the other hand, facial action units-based approaches compute the intensity of AUs over time. First, each action unit representing a facial movement of a particular muscle or group of muscles is detected on a scale from 0 (absent) to 5 (maximal intensity). Then, features related to AU frequency, amplitude, variance, are extracted. These global parameters are then fed as input to machine learning algorithms such as KNN, Random forest, SVM or to statistical tests.

In the second type, the video RGB images are fed as input to CNNs to extract embedding features, which are provided as input to shallow classifiers or directly used by the CNNs. The RGB frames, however, represent only the video spatial component, and thus the dynamics is often lost. To consider dynamic information, Su *et al.* [22] combined the optical flow with spatial information to form a two-stream network (the optical flow and the spatial streams). The optical stream network captures the dynamic changes of facial muscles by taking as input a stack of optical flow images. The spatial stream network captures the spatial information by taking a sequence of RGB frames as input.

Our approach relies on facial AUs, which has yielded promising results in the literature. Moshkova *et al.* [23], for instance, computed the Euclidian distance between AUs of posed emotions and neutral states and found PD patients with significantly lower distances than HC subjects for all six emotions. [20] found that kinematic parameters of AUs, computed on smiling and eyebrow movement tests, found PD patients with decline in the frequency and speed of facial movements. Priebe *et al.* [24] calculated a composite score by multiplying the mean intensity with some selected AU frequency values and then averaged these product terms. They found that PD patients had a significantly lower pain-indicative composite score than the HC group in off state.

Langevin *et al.* [25] found a significant difference in the average intensity of AU4 (brow lowerer) between PD and HC during the "resting facial expression" task. Guan *et al.* [26] extracted 36 global features for each AU in time and frequency domains, selected relevant features using U-test, and applied five-fold CV. They achieved an AUC of 96.42%. However, feature selection was applied on the whole dataset, not the training set only. Additionally, CV was not stratified, implying different class distributions in the training and test sets. Gomez *et al.* [27] adapted a CNN model, originally designed for face recognition to detect AUs in the emotion net dataset, and fine-tuned it to classify PD vs. HC; they obtained an accuracy of 87.3%. Finally, Wu *et al.* [12] devised a scoring system to measure the change in AUs' intensities and their mean. This score, used to quantify the degree of facial expressivity and calculated as the difference between the scores obtained for disgust and neutral expressions, was shown to discriminate PD vs. HC subjects, and demonstrated that as PD severity increased, facial expressivity decreased. Despite their promising results, the studies above have some limitations. Except the study in [24], PD patients were not in their early stage. Furthermore, some studies perform feature selection on the whole dataset, instead of the training set only, which leads to biased results compared to the ones expected in a realistic test setting.

Paper contributions: We propose a study for hypomimia detection in early stage PD patients by analyzing their face videos. To characterize the facial muscles' movement from the videos, we calculate the facial AUs derivatives with order of k of the AUs intensities, that we call delta features. Since these derivatives are computed at the frame level, the resulting representation of each video is in the form of a time series. To obtain a global video representation, we compute the variance across the delta features. Additionally, we extract another global representation of the video by calculating the variance across the action unit intensities. We conducted three experiments, each of which involved evaluating an XGBoost model using nested cross validation on one of three different feature types: delta feature, variance of delta features, and variance of intensities.

The paper is structured as follows. Section 2 presents the database and our approach based on action unit derivatives. In Section 3, the results are presented along with their analysis. Finally, we conclude and discuss future work in Section 4.

2 Methods

2.1 Database

The dataset was collected as part of the ICEBERG protocol, a longitudinal study conducted at the Paris Brain Institute (ICM), aiming to identify and validate biomarkers of Parkinson's disease. The study enrolled 112 PD patients (73 men, 39 women) and 45 HC subjects (26 men, 19 women), according to the following inclusion criteria : (1) clinical diagnosis of idiopathic PD based on the United Kingdom Parkinson's Disease Society Brain Bank criteria [28], with a disease duration of less than 4 years and (2) controls without any neurological disorders. The participants come once a year to the Pitié Salpêtrière hospital for 5 years to undergo several examinations (neurological examination, motor and cognitive tests, biological sampling, and brain MRI) and audio visual recordings. PD subjects were pharmacologically treated and their faces were recorded while on state.

A total of 299 videos, denoted as ICEBERG Video-Feb2023 dataset, were included, consisting of 208 PD (129 males, 79 females) and 91 HC (58 males, 33 females). Information at visit level on age, Hoehn and Yahr stage, MDS-UPDRS III score (off state), MDS-UPDRS III face item, Moca is given in Table 1. The recording session lasts 15 to 20 minutes, where participants are asked to

perform 29 vocal tasks, explained through a user interface. We have considered three tasks where the participants repeat a set of syllables (/pataka/ (twice) /bagada/) as rapidly and continuously as possible, without taking pauses to breathe. These tasks can reveal articulation difficulties, enabling their visual capture. The recording device is a Webcam with integrated encoding and compression of the type 195 Logitech C922 Pro Stream Webcam, with a frame rate of 24 fps (frame per second) and a resolution of 1920 * 1080 pixels.

	PD		HC	
Biological sex	Male	Female	Male	Female
No. of videos	129	79	58	33
No. of subjects	73	39	26	19
Age (years)	64.23 ± 9.23	66.23 ± 8.17	63.46 ± 9.38	63.73 ± 8.46
Hoehn & yahr	1.92 ± 0.35	1.86 ± 0.55	-	-
MDS-UPDRS III total	32.6 ± 6.83	27.31 ± 9.32	4.24 ± 2.72	4.39 ± 2.98
MDS-UPDRS III face item	1.2 ± 0.55	0.85 ± 0.51	-	-
MoCA	26.59 ± 2.72	28.09 ± 1.77	27.52 ± 2.56	28.03 ± 1.67

Table 1: Information related to the ICEBERG Video-Feb2022 dataset recordings

2.2 Approach

2.2.1 Facial Action Units

Paul Ekman developed the Facial Action Coding System (FACS) [29], which identifies the basic movements of facial muscles known as facial action units (AUs). FACS categorizes the facial muscles into 44 action Units (AUs), each corresponding to a distinct movement pattern in the facial muscles. These AUs can be used to describe a wide range of facial expressions. OpenFace [30] is an open-source software tool that performs precise facial landmark detection, head pose estimation, facial action unit recognition, and eye-gaze estimation.

2.2.2 Feature Extraction

OpenFace was used to extract from each video 17 AUs (Figure 1) and providing information on the presence (0 or 1) and intensity level (ranging from 0 to 5) of each AU for every video frame.

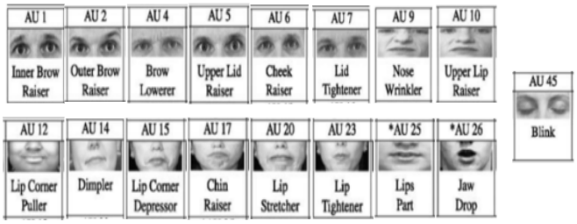

Figure 1: AU diagrams adapted from [31]

Delta features: To encode the video facial muscles' movement, we calculate for each action unit AU_j its delta feature Df_{AU_j}. For each frame, delta feature Df_{AU_j} at frame i is defined by computing the difference in intensities between frame $i+k$ and frame i of AU_j with fixed k (1).

$$Df_{AU_j}(F(i)) = In_{AU_j}(F(i+k)) - In_{AU_j}(F(i)) \quad (1)$$

where $In_{AU_j}(F(i))$ is the AU_j intensity at frame i.

Variance delta features: After computing the delta features of each video with step k, a video is represented as time series of delta AU features. To globally represent it, we calculate the variance of each delta feature Df_{AU_j} across the frames (2).

$$\mathrm{Var}(Df_{AU_j}) = \frac{1}{N}\sum_{i=1}^{N}(Df_{AU_j}(F(i)) - \bar{Df}_{AU_j})^2 \quad (2)$$

where $\bar{Df}_{AU_j} = \frac{1}{N}\sum_{i=1}^{N} Df_{AU_j}(F(i))$, and N is the number of frames in the video.

Baseline features: We extract other basic features by calculating for each AU_j the variance of the intensities across the frames (3). These features are used as reference to evaluate their predictive power in comparison to the variance delta features.

$$\mathrm{Var}(AU_j) = \frac{1}{N}\sum_{i=1}^{N}(In_{AU_j}(F(i)) - \bar{In}_{AU_j})^2 \quad (3)$$

where $\bar{In}_{AUj} = \frac{1}{N}\sum_{i=1}^{N} In_{AU_j}(F(i))$, and N is the number of frames in the video

2.2.3 Classification

We use XGBoost to classify PD vs. HC, for each task, where we conduct three experiments,

associated with our three feature types, i.e. delta features, variance delta features, and baseline features. We evaluate XGBoost using stratified nested cross validation (CV) with five folds in the inner and outer loops. Stratification is based on biological sex to ensure the training, validation and test sets have equal representation of male and female for each class (PD or HC). Note that same-subject videos appear in only one of these three sets. The XGBoost hyperparameters were optimized by performing an inner loop of CV with five folds in each training fold of the outer loop. The best hyperparameters and their corresponding estimated models were selected based on the highest average balanced accuracy (BA) over the validation folds. This ensures that although the number of final estimated models in the nested CV is $k_{outer} = 5$, the optimal hyperparameters are consistent across all estimated XGBoost models.

2.2.4 Training Phase

Training with delta features: For each task, we calculate the delta features with a step size k from 1 to 20. Then, we evaluate XGBoost for each training set of the nested cross-validation (2.2.3), which results in model's performance on the corresponding validation set in terms of BA at frame level for each task and k. As the three tasks involve the same phonetic aspect (/pataka/ (performed twice), /bagada/), we select a single optimal k for them, by computing the average BA across the three tasks for each value of k, and selecting the one that yields the highest average performance among the 20 estimated values.
XGBoost inference is performed at the frame level, resulting in a classification score (between 0 and 1) for PD and HC for each frame. The mean of the classification scores across the video frames is then used to predict the class (PD or HC). Finally, BA and AUC are used to assess model performance at the visit level and at the subject level.

By training each task separately, we obtain a classification score for each task in each visit. To determine the classification score for a specific visit, we average the individual task scores obtained for that visit across the tasks. To get a classification score per subject, we average the visit-based classification scores for that subject.

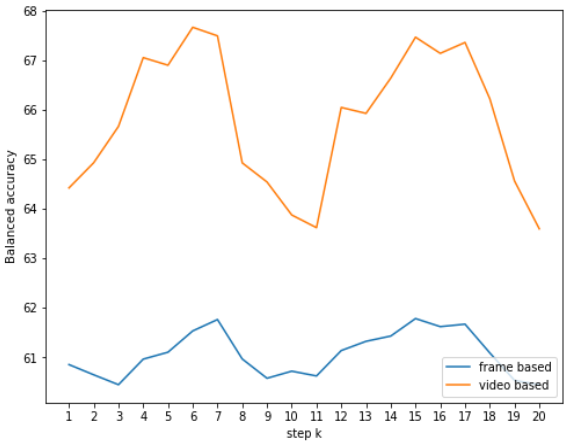

Figure 2: Mean balanced accuracy on validation sets across the three tasks

Training with global features: Our global features, which provide a global video representation, are either the variance delta features or the baseline features. For each task, the former are calculated using the optimal k obtained when performing the delta features experiments. We employ each of these two global feature representations as input to XGBoost to classify PD vs. HC through nested cross-validation.

3 Results and Discussion

3.1 Classification Results

3.1.1 Results with Delta Features

We have assessed the model's performance on validation sets for the three tasks and k values ranging from 1 to 20. We used BA score at both the frame and visit (video) levels to evaluate performance. Figure 4 displays the average BA across the three tasks for each k, at the frame and video visit level. The graphs exhibit a periodic pattern that is inherent to tasks involving rapid and continuous repetition of a set of syllables. The graph period, approximately 10, corresponds to

	Visit level		Subject level	
	BA	AUC	BA	AUC
Delta AU ($k = 7$) %	67.17	73.71	70.56	78.33
Variance Delta AU ($k = 7$) %	70.94	77.97	75.28	82.28
Variance AU %	69.36	74.74	74.62	80.33
Three approaches combined %	73.00	78.38	77.29	83.11

Table 2: Results of proposed approaches for the ICEBERG Video-Feb2022 dataset

the average number of frames that the expressions (/pataka/, /bagada/) last. The optimal k for frame-based evaluation was 7, with a BA of 61.76% at the frame level and 67.49% at the visit level on the validations sets. The results on the test sets using the models with the optimal k are as follows: 62.5% BA at the frame level, 67.17% BA at the video visit level and 70.56% at subject level.

3.1.2 Results with Global features

Using the optimal k (k=7) selected above, variance delta features were defined by calculating the variance across the frames of each delta feature. The PD vs. HC classification results show that the variance delta features have a discrimination BA of 70.94% at the visit level and 75.28% at the subject level, which are better than the results obtained with delta features. The classification using the variance across the action units intensities shows a BA of 69.36% at the visit level and 74.62% at the subject level. These results are similar to those obtained with the variance delta features.

3.1.3 Results of the Combined Approaches

By combining the scores provided by the three approaches, we obtain BA of 73% at visit level and 77.29% at subject level. Combining their complementary information thus improves overall performance. Table 2 shows the results at visit and subject levels. As shown, repeating a set of syllables (/pataka/ and /bagada/) through vocal tasks is effective in detecting hypomimia. This is consistent with existing literature on voice analysis, which suggests that these vocal tasks highlight PD related articulation issues [32]. This also suggests that hypomimia is one of the factors contributing to articulation problems.

To evaluate the statistical significance of these results, we conducted a McNemar's test [33], that compares the proportions of well-classified visits or subjects across the proposed approaches. This test revealed that, at the visit and subject levels, the improvements, w.r.t delta approach, of delta variance, variance and combination approaches are statistically significant (p-value = 0.003, 0.009, 3e-07 at visit level) and (p-value = 0.002, 0.005, 0.001 at subject level) respectively. However, no statistical significance was observed for the delta variance approach compared to the variance and combination approaches (p-value = 0.658, 0.576 at visit level) and (p-value = 1, 1 at subject level) respectively. Similarly, there was no significant difference observed between the variance approach and the combination of approaches (p-value = 0.351 at visit level and p-value = 1 at subject level). Globally, our results compare favorably with the state of the art ([15] [34] [35]) in terms of BA or AUC. However, this comparison must be approached with caution, as these works sometimes optimize some hyperparameters on the test set. Additionally, the datasets have varying sizes, and patients may have been recorded at different medication states, and are usually at later stages of the disease.

3.2 Model Interpretablity

SHAP (SHapley Additive exPlanations) [36] is a technique that assigns importance values to each feature to explain model prediction for each instance and globally. Figure 3 displays the SHAP values with XGBoost for one of the three tasks (/bagada/). We observe that features with lower values of the variance of intensity derivatives (blue dots) contribute more to PD prediction, whereas those with higher values (red dots) contribute more to HC prediction. This makes sense as hypomimia, characterized by the loss of facial muscles, is typically observed in PD patients and is likely to result in lower values of the variance of intensity derivatives as compared to HC. Moreover, the top

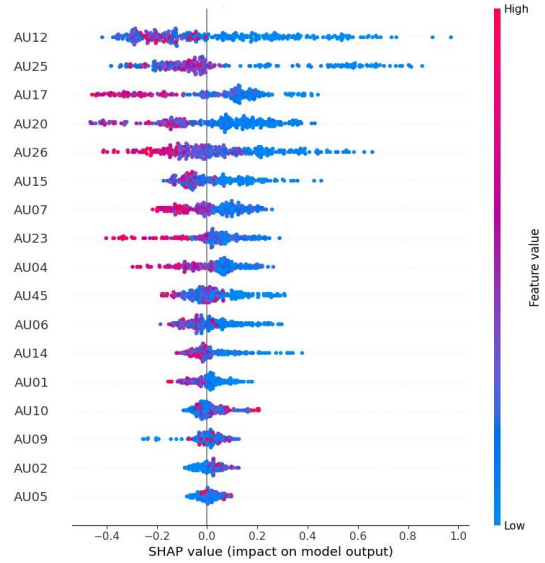

Figure 3: Shap values of variance delta features for the /bagada/ task

5 features with greater impact on discriminating PD vs. HC are those in the mouth region, namely AU12, AU25, AU17, AU20, and AU26. This finding is consistent with the fact that the subjects in this study were required to perform vocal tasks.

4 Conclusion

We have proposed a new method to detect hypomimia in early stage Parkinson's patients, that characterizes the facial muscles' movement by their facial action unit intensities' derivatives. By combining three tasks across visits for each subject, we obtain a BA of 77.29% and an AUC of 83.11% at subject level in PD vs. HC classification. In the future, we will combine optical-flow-based CNNs and AUs-based classifiers, and integrate facial and vocal features. Our work is limited by our insufficient dataset, although it is relatively large w.r.t to other datasets, and by the imbalance in the PD vs. HC and biological sex distributions.

Conflict of interest: None.

Acknowledgment: The authors thank V. Maheo for her contributions to the acquisitions, Dr Habib Benali for his support and all the subjects for their participation in this research.

Fundings: This work was supported by grants from ERA PerMed EU-wide project DIGIPD; ANR through "Investissements d'avenir" program (ANR-10-IAIHU-06 and ANR-11-INBS-0006); Fondation EDF; Fondation Planiol; Société Française de Médecine Estétique and Energipole.

References

[1] O.-B. Tysnes *et al.*, "Epidemiology of Parkinson's disease," *Neural Transm.*, 2017.

[2] L. Darden *et al.*, "Mechanisms and models," *Cambridge Companion to Philos. Biol.*, vol. 39, 2007.

[3] F. K. Salawu *et al.*, "Non-motor symptoms of Parkinson's disease: diagnosis and management.," *Nigerian J. Med*, 2010.

[4] C. H. Hawkes *et al.*, "A timeline for parkinson's disease," *Parkinsonism Relat. Disord*, vol. 16, 2010.

[5] B. R. Haas *et al.*, "Premotor biomarkers for parkinson's disease-a promising direction of research," *T. Neurodegener.*, vol. 1, 2012.

[6] H. Fröhlich *et al.*, "Leveraging the potential of digital technology for better individualized treatment of parkinson's disease," *FIN*, 2022.

[7] J. Jankovic *et al.*, "Parkinson's disease: Clinical features and diagnosis," *JNNP*, 2008.

[8] S. D. Gunnery *et al.*, "The relationship between the experience of hypomimia and social wellbeing in people with Parkinson's disease and their care partners," *Parkinsons Dis*, 2016.

[9] D. C. G. Goetz *et al.*, "State of the Art Review The Unified Parkinson ' s Disease Rating Scale (UPDRS): Status and Recommendations," *Mov. Disord*, vol. 18, no. 7, 2003.

[10] G. Su *et al.*, "Detection of hypomimia in patients with parkinson's disease via smile videos," *Ann Transl Med*, vol. 9, 2021.

[11] L. F. Gomez *et al.*, "Improving parkinson detection using dynamic features from evoked

expressions in video," in *CVPR*, 2021.
[12] P. Wu *et al.*, "Objectifying facial expressivity assessment of parkinson's patients: preliminary study," *Comput Math Methods Med*, 2014.
[13] K. Clawson *et al.*, "Automated representation of non-emotional expressivity to facilitate understanding of facial mobility: Preliminary findings," in *IntelliSys*, IEEE, 2017.
[14] A. Joshi *et al.*, "Context-sensitive prediction of facial expressivity using multimodal hierarchical bayesian neural networks," in *FG*, IEEE, 2018.
[15] A. Grammatikopoulou *et al.*, "Detecting hypomimia symptoms by selfie photo analysis," *ACM ICPS*, 2019.
[16] M. Rajnoha *et al.*, "Towards identification of hypomimia in parkinson's disease based on face recognition methods," *ICUMT*, 2019.
[17] A. Bandini *et al.*, "Analysis of facial expressions in parkinson's disease through video-based automatic methods," *J. Neurosci. Methods*, vol. 281, 2017.
[18] A. Abrami *et al.*, "Automated computer vision assessment of hypomimia in parkinson disease: Proof-of-principle pilot study," *J Med Internet Res*, vol. 23, 2021.
[19] B. Jin *et al.*, "Diagnosing parkinson disease through facial expression recognition: video analysis," *J Med Internet Res*, vol. 22, 2020.
[20] A. Moshkova *et al.*, "Studying facial activity in parkinson's disease patients using an automated method and video recording," in *FRUCT*, IEEE, 2021.
[21] L. F. Gomez-Gomez *et al.*, "Exploring facial expressions and affective domains for parkinson detection," *arXiv*, 2020.
[22] G. Su *et al.*, "Hypomimia recognition in parkinson's disease with semantic features," *TOMM*, vol. 17, 2021.
[23] A. Moshkova *et al.*, "Facial emotional expression assessment in parkinson's disease by automated algorithm based on action units," in *FRUCT*, IEEE, 2020.
[24] J. A. Priebe *et al.*, "Does parkinson's disease lead to alterations in the facial expression of pain?," *J. Neurol. Sci*, vol. 359, 2015.
[25] R. Langevin *et al.*, "The park framework for automated analysis of parkinson's disease characteristics," *PACM IMWUT*, 2019.
[26] Y. Guan *et al.*, "Application of logistic regression algorithm in the diagnosis of expression disorder in parkinson's disease," in *ICIBA*, vol. 2, IEEE, 2021.
[27] L. F. Gomez *et al.*, "Exploring facial expressions and action unit domains for parkinson detection," *PLoS ONE*, 2023.
[28] W. Gibb *et al.*, "The relevance of the lewy body to the pathogenesis of idiopathic parkinson's disease.," *JNNP*, 1988.
[29] P. Ekman *et al.*, "Facial action coding system," *Psychol. Nonverbal Behav*, 1978.
[30] T. Baltrušaitis *et al.*, "Openface: an open source facial behavior analysis toolkit," in *WACV*, IEEE, 2016.
[31] R. Zhi *et al.*, "A comprehensive survey on automatic facial action unit analysis," *Vis Comput*, vol. 36, 2020.
[32] L. Jeancolas *et al.*, "X-vectors: New quantitative biomarkers for early parkinson's disease detection from speech," *FIN*, 2021.
[33] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, 1947.
[34] W. S. Lim *et al.*, "An integrated biometric voice and facial features for early detection of parkinson's disease," *NPJ PD*, vol. 8, 2022.
[35] M. Novotny *et al.*, "Automated video-based assessment of facial bradykinesia in de-novo parkinson's disease," *NPJ DM*, vol. 5, 2022.
[36] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable ai for trees," *Nat. Mach. Intell*, 2020.