Spatio-Temporal Hypomimic Deep Descriptor to Discriminate Parkinsonian Patients

WSSFN 2025 Interim Meeting. Abstract 0166.

William Omar Contreras López,[1] Brayan Valenzuela,[2] Jhon Arevalo, [3] Fabio Martínez.[2]
[1] International Neuromodulation Center NEMOD. Colombia.
[2] Universidad Industrial de Santader. Colombia.
[3] Researcher With Machine Learning Analysis and Computer Vision (MLACV). Colombia.

Corresponding author: William Omar Contreras López email: wcontreras127@unab.edu.co

How to Cite: Contreras López WO, Valenzuela B, Arevalo J, Martínez F. Spatio-Temporal Hypomimic Deep Descriptor to Discriminate Parkinsonian Patients: WSSFN 2025 Interim Meeting. Abstract 0166. NeuroTarget. 2025;19(2):144-5.

Abstract

Introduction: The hypomimia is a main clinical sign of Parkinson disease that describes motor patterns associated with the reduction and progressive loss of facial expression. This clinical sign constitutes a main biomarker to support diagnosis, even at early stages, and to establish progression and description of the disease. In clinical routine, the evaluation of such signs remains subjective or limited to the description of some landmarks that poorly describe little expressions correlated with the disease. This work introduces a new digital biomarker, expressed as a spatio-temporal convolutional representation that learns facial movement patterns to discriminate between Parkinson and control patients.

Clinical description: This paper implements a 3D convolutional representation, inspired by the net scheme of I3D. This representation was here trained to automatically identify Parkinsonian facial patterns, using a supervised end-to-end scheme. From such representation it is possible to discriminate patterns but also to recover spatio-temporal activations that enhance spatial regions with major correlation with the disease. The trained volumetric network demonstrated distinct activation patterns between groups. In Parkinson's patients, the model concentrated on restricted facial regions (eyes, nose, mouth), reflecting reduced expressivity and limited dynamic movement. In contrast, control participants elicited broader and more diverse activation across facial areas. The ablation analysis varying temporal resolution showed that a configuration using 14 frames per video yielded the highest classification performance. Compared with a standard 2D ResNet-50 model and previously reported static-feature approaches, the proposed 3D spatio-temporal descriptor achieved superior discrimination. The learned embedding space further showed clear separation between patients and controls.

Discussion: These findings suggest that the reduced activation spread in Parkinson's patients corresponds to clinical hypomimia characterized by facial rigidity and diminished micro-movement dynamics. The improved performance at lower temporal resolution indicates that moderate temporal subsampling facilitates the detection of subtle motion deficits inherent in hypomimic expression. The superiority of the spatio-temporal method over static models supports the relevance of capturing dynamic facial patterns rather than isolated frames or manually selected landmarks. The embedding distribution also suggests potential alignment with clinical severity markers, indicating that this approach may serve as a complementary digital biomarker for disease characterization and monitoring.

Conclusions:This work introduced an end-to-end method for learning a spatio-temporal representation of facial patterns in video. The proposed method outperformed similar approaches for Parkinson detection in video sequences. The analysis of attention maps in intermediate layers highlighted facial regions associated with the neural network's prediction. These regions could provide support to clinicians'

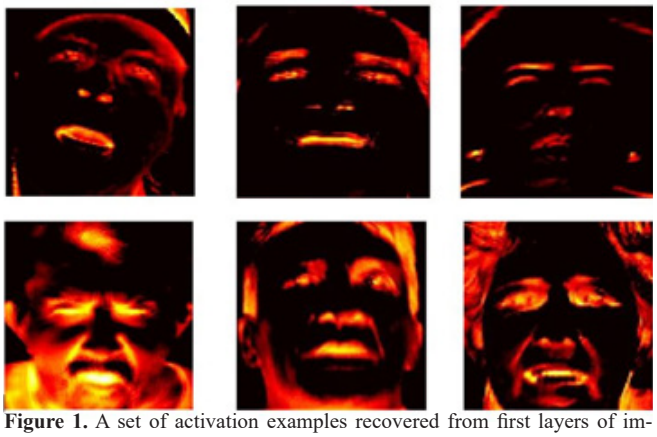

Figure 1. A set of activation examples recovered from first layers of implemented deep representation. In top, it is represented PD patients, with isolated and marked facial regions. In button it is shown a set of activations for control subjects, with larger spatio-temporal patterns.

evaluation and further patient's monitoring. In this research we also collected, to the best of our knowledge, the first audiovisual dataset of Parkinson's patients and control subjects with diagnosis stage annotations. This dataset will enable the exploration of computational methods for multi-modal analysis.

References

1. Cattaneo L, Pavesi G. The facial motor system. Neurosci Biobehav Rev. 2014. Available from: http://dx.doi.org/10.1016/j.neubiorev.2013.11.002
2. Katsikitis M, Pilowsky I. A study of facial expression in Parkinson's disease using a novel microcomputer-based method. J Neurol Neurosurg Psychiatry. 1988;51(3):362–6. Available from: https://pubmed.ncbi.nlm.nih.gov/3361329/
3. Carreira J, Zisserman A. Quo Vadis, action recognition? A new model and the Kinetics dataset. 2017. Available from: http://arxiv.org/abs/1705.07750
4. Abrami A, Gunzler S, Kilbane C, Ostrand R, Ho B, Cecchi G. Automated computer vision assessment of hypomimia in Parkinson disease: Proof-of-principle pilot study. J Med Internet Res. 2021 Feb 22;23(2):e21037. Available from: https://pubmed.ncbi.nlm.nih.gov/33616535/
5. Gomez LF, Morales A, Orozco-Arroyave JR, Daza R, Fierrez J. Improving Parkinson detection using dynamic features from evoked expressions in video. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) [Internet]. IEEE; 2021. p. 1562–70. Available from: http://dx.doi.org/10.1109/CVPRW53098.2021.00172