International Journal of PSYCHOLOGICAL RESEARCH

RESEARCH ARTICLE

Vol 17, N° 2
https://revistas.usb.edu.co/index.php/IJPR
ISSN 2011-2084
E-ISSN 2011-7922

A Volumetric Deep Architecture to Discriminate Parkinsonian Patterns from Intermediate Pose Representations

Una arquitectura volumétrica profunda para discriminar patrones parkinsonianos desde representaciones de poses intermedias

Jean Portilla[1], Edgar Rangel[1], Luis Guayacán[1], Fabio Martínez[1,*]

[1] *BIVL2ab- Biomedical Imaging, Vision and Learning Laboratory. Universidad Industrial de Santander.*

Abstract.
Parkinson's disease (PD) is a common neurodegenerative disorder worldwide, with over 6.2 million registered cases. Gait analysis plays a fundamental role in evaluating motor abnormalities associated with this disease. However, current methods, such as marker-based systems, are intrusive and expert-dependent. Markerless alternatives, like video sequence analysis, have been proposed, but they tend to provide overall classification scores and lack the ability to interpret joint kinematics in detail. An innovative technique is presented using volumetric convolutional networks that can learn intermediate postural patterns and distinguish between Parkinson's patients and control subjects. This approach utilizes *OpenPose* activations and then applies hierarchical convolution to minimize classification. In tests conducted with 14 Parkinson's patients and 16 control subjects, this method achieved a classification accuracy of 98%.

Resumen.
La enfermedad de Parkinson (EP) es un trastorno neurodegenerativo común a nivel mundial, con más de 6.2 millones de casos registrados. El análisis de la marcha desempeña un papel fundamental en la evaluación de las anomalías motoras asociadas con esta enfermedad. Sin embargo, los métodos actuales, como sistemas basados en marcadores, son intrusivos y dependientes de expertos. Se han propuesto alternativas sin marcadores, como el análisis de secuencias de video, que tienden a proporcionar puntajes de clasificación globales y carecen de la capacidad de interpretar la cinemática articular detalladamente. Se presenta una técnica innovadora utilizando redes convolucionales volumétricas que pueden aprender patrones posturales intermedios y distinguir entre pacientes con Parkinson y sujetos control. Este enfoque utiliza activaciones de *OpenPose*, y luego aplica una convolución jerárquica para minimizar la clasificación. En pruebas realizadas con 14 pacientes Parkinson y 16 sujetos control, este método alcanzó una precisión del 98% en clasificación.

Keywords.
Parkinson's Disease, Posture, Artificial Neural Networks, Gait.
Palabras Clave.
Enfermedad de Parkinson, postura, redes neuronales artificiales, marcha.

OPEN ACCESS

Manuscript received: 30-10-2023
Revised: 19-03-2024
Accepted: 30-08-2024

***Corresponding author**:
Fabio Martínez
Email: famarcar@saber.uis.edu.co

Declaration of data availability: All relevant data are within the article, as well as the information support files.

Conflict of interests: The authors have declared that there is no conflict of interest.

How to Cite:
Portilla, J., Rangel, E., Guayacán, L., & Martínez, F. (2024). A Volumetric Deep Architecture to Discriminate Parkinsonian Patterns from Intermediate Pose Representations. *International Journal of Psychological Research*, *17*(2), 84–90. https://doi.org/10.21500/20112084.7405

1. Introduction

Parkinson's Disease (PD) is the second most common neurodegenerative disorder, affecting more than 6.2 million people worldwide (Dorsey et al., 2018, Feigin et al., 2021). Moreover, there has been a significant increase in prevalence over the last three decades, reaching up to 5 times people suffering PD (Tolosa et al., 2021). Currently, the diagnosis of PD is based on the observation and analysis of progressive gait motor disorders, such as rigidity, slowness of movement (bradykinesia), postural instability, among many other related symptoms (Rovini et al., 2017).

Nowadays, standard support for gait analysis characterization is based on marker-based systems, which capture dynamics of key joints by using invasive methodologies based on special markers placed on specific anatomical positions (Baker, 2006). This methodology is nonetheless invasive and alters the natural gesture of movements, which for Parkinson disease can include limitations on the normal development of locomotion. Besides, some of these protocols are strongly dedicated to capture lower-limb kinematics, losing important markers of PD, such as postural instability and coordination.

In the literature, these limitations have been tackled from video analysis alternatives that include markerless setups, achieving remarkable results on the characterization of Parkinson movements. Much of these strategies are based on the training and modeling of video descriptors to classify and differentiate Parkinson's from other motions (Lancet, 2017). A main limitation on these approaches is the poor adaptation in the clinical context, offering alternatives that are difficult to implement in observational setups. In fact, much of the support of these strategies are based on probability scores about malignancy, but losing regional information of affected regions. Hence, many of these strategies may biased for artifacts in the sequences, losing relevance to characterize anatomical and physiological during a locomotion process.

This work introduces a convolutional network that learns spatio-temporal patterns from intermediate postural representations. The proposed approach is based on markerless setups, avoiding additional artifacts to alter the patient's gestures. This work starts by adapting an *OpenPose* architecture to return the bank of intermediate activations related to knowledge about joint fields and probability joint maps. Later, this intermediate representation is projected to a convolutional network, which is there after minimized to discriminate between control and parkinsonian patterns. The results evidence sufficient support to characterize Parkinson from classification scores, but also the capability to explain results from postural information.

2. Proposed Approach

This work introduces a computational strategy for characterizing motor patterns associated with Parkinson's disease based on joint interest points calculated without the use of markers. Inspired in *OpenPose* (Cao et al., 2021), we generate the poses and identify key body points from video sequences. Then a spatio-temporal convolutional network is trained to discriminate these key points, regarding if the patient is control or Parkinson. This network was trained and adjusted from intermediate pose representations: the Joint Confidence Maps (JCM) and Part Affinity Fields (PAF) elements. The general pipeline of the proposed approach is illustrated in Figure 1.

2.1 A Deep Architecture for Pose Estimation (*OpenPose*)

A main contribution of this work is to use markerless setups to avoid the limitation of marker-based configurations. Hence, *OpenPose* (Cao et al., 2021) architecture was used as a pose estimator from gait videos to extract intermediate features. This network is widely used in the literature for estimating key joint points during movement and actions for one or several persons in a scene. Formally, *OpenPose* architecture use an image $I \in R^{w \times h}$ where a posture will be extracted ($P \in N^J$ and $J \in \{j_1, j_2, \ldots, j_n\}$ represents the set of n body joints). This network use only 18 articular points ($|J| = 18$), as showed in the Figure 2a. These joint points allow summarize the dynamics of a particular subject during locomotion. Specifically, each input image I is processed through convolutional layers $\Psi(I)$ to obtain a set of deep activations F, which are further processed through two branches: Part Affinity Fields (L) and Joint Confidence Maps (S). These maps and vector fields are processed through bipartite matching resulting in the association of body joints with articular locations giving us the skeleton shown in Figure 2a.

2.2 Part Affinity Fields

The Part Affinity Fields ($PAFs$) are sets of 2D vector maps used to model spatial and anatomical relationships between pairs of body joints. They are formally described as a set $L = \{L_i\}^C$, where $L_i \in R^{w \times h \times 2}$, and C is a hyperparameter determining the number of $PAFs$ to be found. Each pixel within the PAF contains a vector representing the direction and strength of the connection between the corresponding joint pair. The vector's direction indicates the orientation of the connection, and the vector's magnitude represents the confidence in that connection, as illustrated in Figure 2b.

Figure 1

Proposed Architecture for Generating OpenPose Activations, Featuring a 3D Convolutional Net-Work for Patient Classification Based on Gait

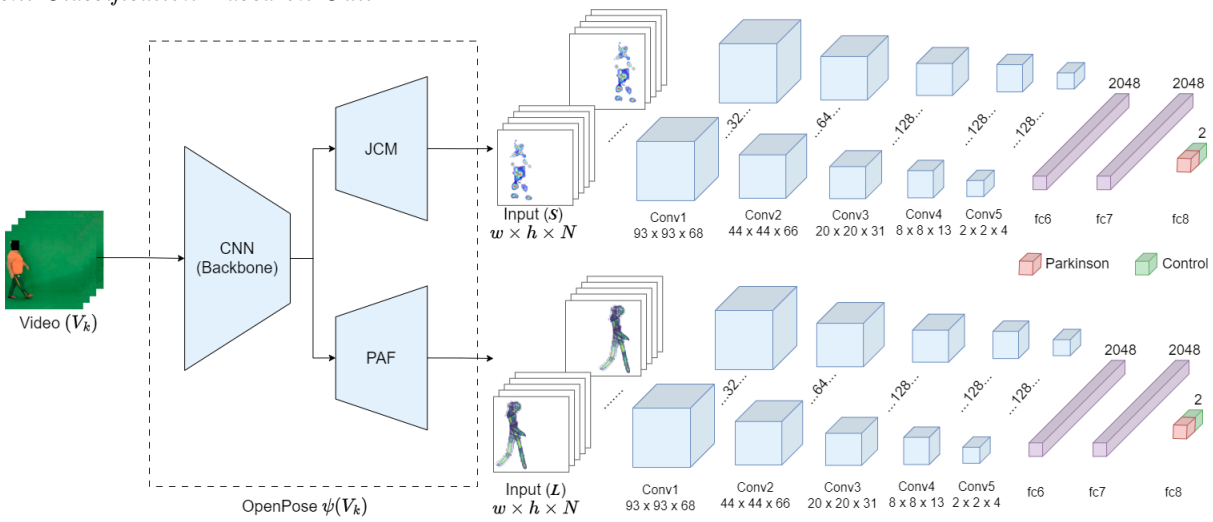
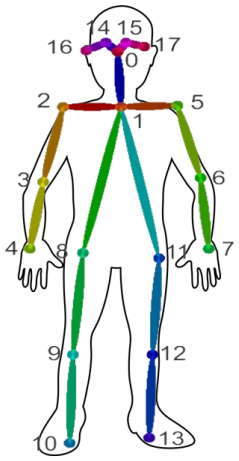


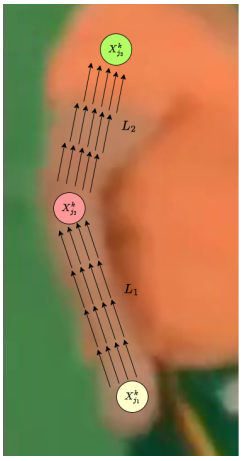
Figure 2

Operation of the Combination of PAF and JCM for the Generation of Poses through OpenPose



Index	Keypoint
0	Nose
1	Chest
2	Right Shoulder
3	Right Elbow
4	Right Hand
5	Left Shoulder
6	Left Elbow
7	Left Hand
8	Right Hip
9	Right Knee
10	Right Foot
11	Left Hip
12	Left Knee
13	Left Foot
14	Right Eye
15	Left Eye
16	Left Ear
17	Right Ear

(a) Skeleton obtained using *OpenPose*



(b) Graphical representation of an example PAF

Formally, *OpenPose* generates a set of PAF, represented as $L^{t=1} = \phi^{t=1}(F)$, where $\phi^{t=1}$ refers to the convolutional layers used for PAF calculation at $t = 1$. For each subsequent refinement stage, the predicted PAF from the previous stage, the original features F, and a set of joint confidence maps (S^{t-1}) are combined and used to generate refined predictions:

$$L^t = \phi^t(F, S^{t-1}, L^{t-1}), \text{ for } t \geq 2 \quad (1)$$

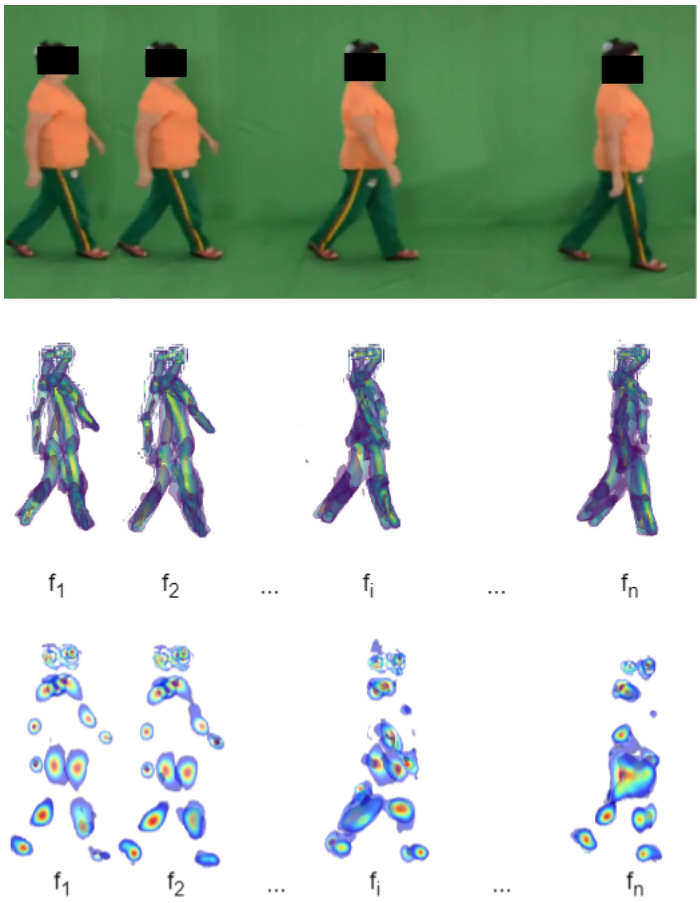
This approach allows progressive refinement of the PAF, contributing to the process of detecting and associating various body parts. In Figure 3, you can observe the resulting activations of these fields, which are potential intermediate representations of the locomotion process, including predominant directions during locomotion.

2.3 Joint Confidence Maps

In a parallel branch of processing, another bank of convolutional filters serves as input to generate a probabilistic representation of poses. Specifically, a Joint Confidence Map (JCM) is a two-dimensional representation that reflects the probability of a specific body part being located at a particular pixel. They are formally described as a set $S = \{S_j\}^J$, with $S_j \in R^{w \times h}$, where $J \in \{j_1, j_2, \ldots, j_n\}$ is the number of body parts (joints). Each pixel in S_j contains a value representing the probability that joint j is located at that position in the image. In other words, the $JCMs$ indicate how confident the model is that a specific joint is located at each pixel of the image.

Figure 3

Visualization of all Part Affinity Fields (PAF and all Joint Confidence Maps (JCMs) Obtained during the Gait of a Control Subject



Similar to the generation of $PAFs$, the network generates a set of $JCMs$, represented as $S^{t=1} = \rho^{t=1}(F)$, where $\rho^{t=1}$ refers to the convolutional module used for map calculation at $t = 1$. At each subsequent refinement stage, the $JCMs$ and $PAFs$ from the previous stage, along with the features F, are combined to generate refined predictions (see Figure 3):

$$S^t = \rho^t(F, S^{t-1}, L^{t-1}), \text{ for } t \geq 2 \quad (2)$$

2.4 3D Convolutional Network for Classifying Parkinsonian Patterns

Once the *OpenPose* architecture is fine-tuned with videos of both control subjects and Parkinson's patients, intermediate representations can be obtained: the PAF (L) and the Joint Confidence Maps (S). These intermediate representations are activations from *OpenPose* that contain relevant information about kinematics during locomotion (L) and joint importance in each frame (S).

In this work, we use these intermediate deep representations (L and S) for characterizing spatio-temporal motor patterns related to Parkinson's disease. To achieve this, a 3D convolutional network was designed and tuned, allowing for the learning of deep relationships while considering the volumetric nature of the information. The 3D convolutional architecture specializes in extracting spatiotemporal features from complete videos, capturing long-term temporal patterns in sequential data. This is crucial in the analysis of parkinsonian gait, as several cardinal symptoms of Parkinson's disease require prolonged observation throughout the gait cycle for accurate identification. In fact, such volumetric representations with 3D convolutions have been proposed in the literature for video analysis in various tasks, such as action recognition (Varol et al., 2017). In this work, the $PAFs$ L were obtained for each frame In of gait video during the last refinement stage ($t = T$), i.e., $\mathbf{L} = \{L^T(I_n)\}^N$, where N corresponds to the total number of video frames. Then, the set of $PAFs$ for all frames, denoted as L, was fed into the 3D convolutional architecture. This architecture incorporates spatiotemporal convolutions to identify different patterns that may occur during a gait process, determining the probability of whether these patterns correspond to a Parkinson's patient or a control subject, as shown in Figure 1. Formally, the network's operation can be expressed

as $P(\text{Parkinsons}|\mathbf{L}) = 1 - P(\text{Control}|\mathbf{L}) = \Psi(\mathbf{L})$, where Ψ represents the set of kernels, layers, functions, and operations that make up the model.

It is worth noting that the intermediate representation of the $PAFs$ contains directional information about the positioning of the joints. Therefore, through 3D convolutions, the architecture is expected to learn kinematics with greater discriminatory power between control subjects and parkinsonian patterns. On the other hand, from Joint Confidence Maps, it is expected to learn coherence between structural activations, which can also be discriminative.

3. Experimental Setup

3.1 Database Description

The database used in this work consists of a series of markerless RGB videos captured during a locomotion exercise. In this study, 30 subjects were invited to participate, including 16 control subjects and 14 who had been diagnosed with Parkinson's disease (PD). The PD patients were in stages of the disease ranging from 1.0 to 4.0 on the Hoehn and Yahr scale. In total, 8 patients were diagnosed with a score of less than or equal to 2.5, and 6 patients scored between 2.5 and 4.0. Each subject in the study was recorded on eight occasions while performing markerless natural walking, four times to the left and four times to the right, resulting in a total of 240 video sequences. This dataset is balanced by age, with an average age of 70.4±5.38 years for control patients and 73±7.45 years for PD patients. All videos were recorded indoors, with a static camera and a uniform background color. The average duration of the videos is 2 seconds. All participants provided informed consent and the research was approved by the ethics committee of the Industrial University of Santander.

3.2 Proposed Method Setup

From each video, we selected $N = 70$ intermediate frames (to cover approximately one gait cycle). Each video record ensures a complete gait cycle that fully exposes the kinematics during locomotion. The videos were resized to a size of 95×95 pixels ($w \times h$). Each frame was individually passed to the *OpenPose* network for pose estimation. The VGG19 net was used to compute convolutional features (Simonyan and Zisserman, 2014). These deep features were then used to generate $PAFs$ and $JCMs$, with the number of stages t set to 6.

The proposed convolutional architecture for classifying parkinsonian patterns from PAF and JCM sequences was fine-tuned considering different convolutional and embedding levels. In this particular work, the following configurations were validated: 5 Conv3D 3 dense layers, 5 Conv3D 1 dense layer, 3 Conv3D 3 dense layers, and 3 Conv3D 1 dense layer. For our models, we used 10 training epochs, a learning rate of 1×10^{-4}, an Adam optimizer, and a cross-entropy loss function. To evaluate each of the configurations, a leave-one-patient-out cross-validation scheme was followed, in which a model was trained for each patient, with the other samples used for model training. Additionally, classification metrics such as accuracy, precision, sensitivity, F2-score, and the area under the curve (AUC) were used for validation.

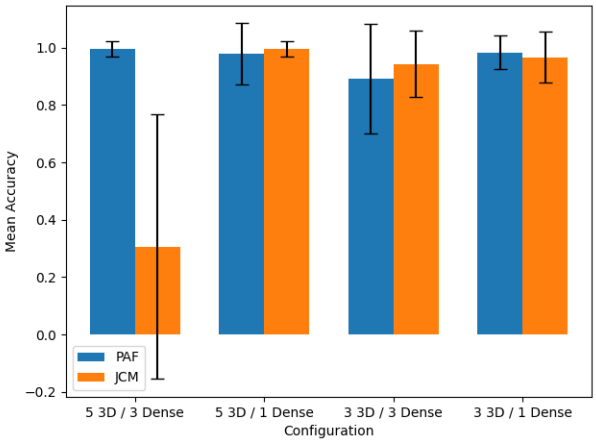
4. Evaluation and Results

To validate the capability of intermediate pose representations, we projected the PAF and JCM activations extracted from the *OpenPose* architecture. This block of activations was fed into a volumetric convolutional architecture to learn discriminative representations between Parkinson's disease and a control population. Simultaneously, we assessed the convolutional representation that yielded the best performance in the discrimination task, determining the various 3D convolutional layers and associated embedded vectors. Figure 4 shows the results obtained for the entire set of 22 patients, with accuracy as the measurement basis for the classification process.

As observed in Figure 4, the intermediate PAF representation yields consistent and robust results across different configurations of the trained architecture. PAF vector maps can influence the pose orientations, which may serve as a characteristic pattern allowing the network to discriminate between the two study populations. On the other hand, the JCM representation exhibits a limited performance in the architecture with 3 dense layers. This could be attributed to the limited training data, particularly in the case of JCM maps, which provide reduced information regarding attention maps around the joints.

Figure 4

Comparison of Patient Classification Accuracy for Control and Parkinson's Disease using Different Configurations, Using PAF vs JCM as Inputs



In a second experiment, we conducted a comparison with a state-of-the-art proposal that advo-cates vol-

Table 1

Comparison of Classification Metrics for Control and Parkinson's Patients between our Proposal and the State of the Art

Method	Accuracy	Precision	Sensitivity	F2-Score	AUC
Guayacán (RGB)	.949	.910	1.0	.780	.950
Guayacán (OF)	.847	.870	.780	.700	.910
Ours (PAF)	**.994**	**1.0**	.989	.991	**.999**
Ours (JCM)	**.994**	.989	**1.0**	**.997**	**1.0**

umetric representation but uses projections from raw videos or maps the response of an optical flow algorithm (Guayacán & Martínez, 2021). Table 1 summarizes the results obtained by the state-of-the-art method, employing an architecture with similar characteristics (3D convolutions) on both RGB and flow sequences (OF). We also included the projections using intermediate representations based on PAF and JCM maps.

As observed, in general, all the projections exhibit a notable performance in classification metrics. This result could be attributed to the limited dataset or the stages of Parkinson's patients within the population. It is worth noting that the intermediate projections provide a better representation of the information, correctly classifying the samples from the mapped videos (achieving perfect precision, sensitivity, and AUC in one of the two configurations). Furthermore, the reported AUC for the intermediate maps not only makes it robust for binary classification but also demonstrates a marked class separation. This can be crucial when extending the analysis to more comprehensive studies with additional cases. Additionally, these maps can offer greater explanatory power, breaking down the kinematic information into postural components.

5. Conclusions and Perspectives

This work introduced a novel markerless strategy to characterize spatiotemporal parkinsonian patterns from pose intermediate representations. In this work, firstly an *Openpose* architecture is tuned to learn locomotion from Parkinson disease and control subjects. From such pose generator is taken the intermedia bank of activations related to probability maps of joints and vector file maps of pose structure. These intermediate representations are mapped to a 3D convolutional net, adjusted to learn discriminative patterns among two considered populations. The results showed a high capacity in this task, with these indices being potential indicators of abnormalities associated with the disease during locomotion tasks. Future works include the analysis on extra datasets with a larger cohort of patients with different degrees of the disease.

6. Acknowledgment

To the Ministry of Science, Technology and Innovation of Colombia by the project: *Caracterización de movimientos anormales del parkinson desde patrones oculomotores, de marcha y enfoques multimodales basados en visión computacional*, with code 92694.

References

Baker, R. (2006). Gait analysis methods in rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, *3*(1), 1–10. https://doi.org/10.1186/1743-0003-3-1

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2021). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(1), 172–186. https://doi.org/10.1109/TPAMI.2019.2929257

Dorsey, E., Sherer, T., Okun, M. S., & Bloem, B. R. (2018). The emerging evidence of the parkinson pandemic. *Journal of Parkinson's Disease*, *8*(s1), S3–S8. https://doi.org/10.3233/JPD-181474

Feigin, V. L., Vos, T., Alahdab, F., Amit, A. M. L., Bärnighausen, T. W., Beghi, E., Beheshti, M., Chavan, P. P., Criqui, M. H., Desai, R., Dhamminda Dharmaratne, S., Dorsey, E. R., Wilder Eagan, A., Elgendy, I. Y., Filip, I., Giampaoli, S., Giussani, G., Hafezi-Nejad, N., Hole, M. K., & Murray, C. J. L. (2021). Burden of neurological disorders across the US from 19902017: A global burden of disease study. *JAMA Neurology*, *78*(2), 165–176. https://doi.org/10.1001/jamaneurol.2020.4152

Guayacán, L. C., & Martínez, F. (2021). Visualising and quantifying relevant parkinsonian gait patterns using 3d convolutional network. *Journal of Biomedical Informatics*, *123*, 103935. https://doi.org/10.1016/j.jbi.2021.103935

Rovini, E., Maremmani, C., & Cavallo, F. (2017). How wearable sensors can support parkinson's disease diagnosis and treatment: A systematic review. *Frontiers in Neuroscience*, *11*, 555. https://doi.org/10.3389/fnins.2017.00555

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv* preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

The Lancet. (2017). Artificial intelligence in health care: within touching distance. *The Lancet*, *390*(10114), 2739. https://doi.org/10.1016/S0140-6736(17)32846-5

Tolosa, E., Garrido, A., Scholz, S. W., & Poewe, W. (2021). Challenges in the diagnosis of parkinson's disease. *The Lancet Neurology*, *20*(5), 385–397. https://doi.org/10.1016/S1474-4422(21)00030-2

Varol, G., Laptev, I., & Schmid, C. (2017). Long-term temporal convolutions for action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(6), 1510–1517. https://doi.org/10.1109/TPAMI.2017.272304