Parkinson's disease detection on DaTscan images using Random Forest and Graph Deep Learning

Urien Hélène
Isep, LISITE
Issy-Les-Moulineaux, France
helene.urien@isep.fr

***Abstract*—Parkinson's disease is characterized by motor disorders caused by the progressive death of neurons producing dopamine. DaTscan imaging, namely Single-Photon Emission Computed Tomography (SPECT) with ioflupane (^{123}I), allows to quantify this loss of neurodopaminergic neurons, more particularly observed in the striatum. In this article, a two-step classification method is proposed to detect Parkinson's disease on DaTscan images focusing on this key structure. A graph is first created from the segmentation of the striatum, relying on the creation of the Max-Tree of the DaTscan image and a prior segmentation. Then, a Random Forest algorithm is applied to classify the nodes of the striatal graph, computing features extracted locally or related to subject's properties. Finally, the classification is made on the whole graph using Deep Learning, and adding the Random Forest predictions as new features. The combination of these two approaches offers promising results on the Parkinson's Progression Markers Initiative (PPMI) database, achieving an average balanced accuracy of 91.1% using a nested cross-validation strategy to increase the robustness of the results.**

***Index Terms*—Parkinson's disease, DaTscan, Graph Deep Learning, Random Forest, Max-Tree**

I. Introduction

Parkinson's is the second most common neurodegenerative disease in the world [1]. It results in the progressive death of neurons producing dopamine, a hormone involved in the control of motor movements. The symptoms, including tremors, muscular rigidity and slowness of movement, occur when 50 to 70% of these neurons are dead. Thus, the disease, and more especially at an early stage or when symptoms appear non-simultaneously or at different intensity levels, may be difficult to diagnose.

Single-Photon Emission Computed Tomography (SPECT), a nuclear imaging technique, can be used to confirm diagnosis by quantifying dopamine level in the brain when using ioflupane (^{123}I) as a radiotracer [2]. The resulting images, also called DaTscans, showing a striatum with an unusual shape and lower intensity values (see Fig. 1), can be suggestive of Parkinson's disease, as the radiotracer uptake is then expected to be reduced in this region because of the loss of dopaminergic neurons.

Hence, numerous classification methods for Parkinson's disease focus on the striatum, for example restricting the DaTscan images to one or multiple slices [3]–[5] or regions [6] encompassing this key structure. Features extracted from the striatum can also feed the model [7]. Moreover, regions outside the striatum and not affected by the disease can be used for intensity normalization [4]. If most of these methods apply Deep Learning models, and more especially Convolutional Neural Network (CNN) [8], on the image focused on the striatum, [6] further refines the search domain by masking regions potentially outside the striatum using isosurfaces associated with different thresholds. This method also offers promising and robust performances using a nested cross-validation strategy. Thus, focusing on the region really delimited by the striatum on DaTscan images can increase the performances to detect Parkinson's disease.

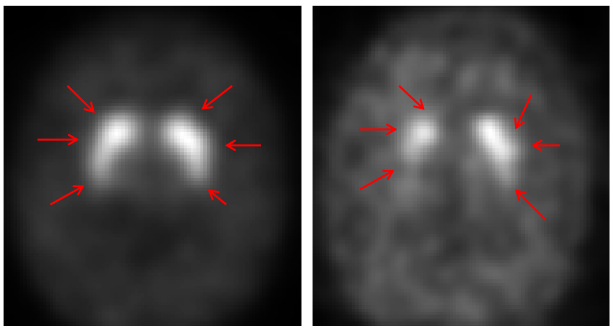

Fig. 1: DaTscan images of a healthy subject (left) and a patient with Parkinson's disease (right). The striatum is indicated by red arrows.

In this context, this article proposes an original Graph Deep Learning-based approach to diagnose Parkinson's disease on DaTscan images. First, a subject specific segmentation of the striatum is automatically performed relying on an atlas of the brain and Max-Trees. Then, a graph is created from this segmentation, each node being a connected component of the binary segmentation of the striatum on each axial slice. Finally, the classification is divided in two steps: a single node classification using Random Forest, followed by the whole graph classification using Deep Learning. More precisely, the Graph Deep Learning model uses the node predictions provided by the Random Forest algorithm, and both classification tasks benefit from subject and striatal properties such as shape, intensity or symmetry. The associated features, as well as the whole detailed procedure, are further described.

II. Material and method

A. Dataset

The dataset used in this article is made of DaTscan images from 1012 subjects, 810 patients with Parkinson's disease and 202 healthy subjects, downloaded in the Parkinson's Progression Markers Initiative (PPMI) database during october 2024. More particularly, the collected images include data from 525 male patients and 88 male control subjects. The minimum, maximum and median age, namely 30, 85 and 64 years, are similar for the chosen patients and control subjects. The images were downloaded from subjects who did not have any pathogenic genetic variant, diagnostic change over the study or withdrawal from the study. Then, for each selected subject, the first available reconstructed and pre-processed DaTscan image was downloaded. The pre-processing consisted in aligning the data into the MNI space, leading to images of dimension 91x109x91 voxels, with voxel size 2x2x2mm^3.

B. Generation of an atlas of the brain

Because the available DaTscan images were in the MNI space, an atlas of the brain was used to obtain a prior segmentation of relevant structures to detect Parkinson's disease such as the striatum. However, brain atlases are usually available on Magnetic Resonance (MR) images. To that end, AssemblyNet [9], [10], a Deep Learning-based brain parcellation algorithm accessible from the volBrain platform [11], was applied to the 2mm isotropic T1-weighted MNI template image of FSL [12]. It provided a prior segmentation of 132 brain structures, associated with 264 labels making the distinction between right and left brain hemispheres.

Among the segmented structures, the putamen, pallidum and caudate composing the striatum were available. The union of these regions was used to obtain a prior segmentation of the striatum in each brain hemisphere.

Moreover, the lentiform nucleus, composed of the putamen and pallidum, and the neostriatum, made of the caudate and putamen, could also be potentially relevant for the analysis. In this article, the augmented atlas refers to 135 brain structures: the 132 regions of the previously generated atlas, the striatum, neostriatum and lentiform nucleus.

Finally, a prior binary mask of the brain was generated, excluding regions associated with the background of the atlas. After filling the holes of the binary image of the brain, slice by slice, this mask was applied to each DaTscan image.

C. Creation of the graph of the striatum

The first step of the method requires the 3D segmentation of the striatum, appearing as a well delimited structure of bright intensity on DaTscan images, more especially for healthy subjects (see Fig. 1). Therefore the Max-Tree [13], a hierarchical structure highlighting regions with intensity values greater than their surrounding in an image, notably used for segmentation [14], is suitable to detect the striatum. Indeed, the Max-Tree is a graph, each of its nodes being associated with a connected component of the binary image obtained after thresholding the input image to one of its intensity values. In this work, the striatum is expected to be associated with two nodes of the Max-Tree, one in each brain hemisphere.

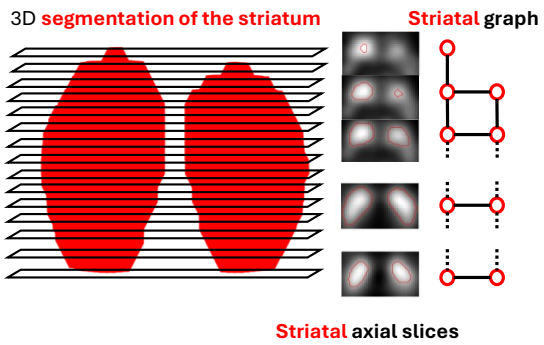

Fig. 2: Illustration of the creation of the graph of the striatum on a healthy subject.

The Max-Tree of each DaTscan image is first created, applying all possible thresholds to the associated requantified image (see section III-A). Then, the nodes having the highest overlap with the right and left prior striatum segmentations are selected, this overlap being measured with the Dice's coefficient.

Finally, the striatal graph is created from the axial slices of the 3D segmentation of the striatum, therefore composed of the best two nodes of the Max-Tree of the associated DaTscan image. More precisely, each node of the graph of the striatum represents an axial slice of the right or left segmented striatal region. Edges are then created between nodes associated with striatal regions on adjacent axial slices belonging to the same hemisphere. Two nodes associated with striatal regions on a same axial slice are also linked by an edge. Thus, the resulting graph, as illustrated in Fig. 2, is specific to each subject, and can be used for Parkinson's disease classification associating relevant features to each node.

D. Classification of striatal nodes using Random Forest

Once the striatum's tree has been created, its nodes are first classified individually using machine learning. To that end, an unique label is assigned to all nodes from a same subject, *i.e.* 1 for patients with Parkinson's disease, 0 otherwise. The label of each node is then predicted based on features falling into six categories, described above.

a) Clinical data: the chosen clinical features are the sex and age. Thus, the associated feature values are the same for all the nodes of a same subject.

b) Shape: geometric properties are measured for each node. They include area (in mm^2 and as the ratio with the area of the brain on the same slice), sphericity (computed as the overlap between the node and its best fitted circle) and the overlap with the node on the same hemisphere on the previous slice. Geometric information from the augmented atlas is also considered, computing the ratio with the area of each augmented atlas region on the same slice, and the overlap with each augmented atlas region, on the same slice or restricted to the right or left hemisphere.

c) Symmetry: these features cover both shape and intensity information. For each slice, the symmetric image is computed as the vertically flipped image. For each node, the symmetric node is computed as the vertically flipped binary image of the node on the opposite hemisphere on the same slice. Symmetry information is extracted computing the overlap between the node and its symmetric node, and the ratio between the mean node intensity on the original and symmetric image.

d) Localization: spatial information is computed on each node and its associated slice. The distance between the node's slice and a reference slice, namely the first or median striatal slice, is computed in mm. The augmented atlas is also used to determine whether a node belongs to a specific region, either on the whole slice or restricted to the right or left hemisphere.

e) Intensity: these features are based on the ratio between the mean intensity value of the node and the intensity value of a reference region. Thus, the mean and maximum intensity values of the brain, computed on the whole volume or restricted to the same slice as the node, are used as a reference. The mean intensity value of the regions outside the segmented striatum (on the whole volume or restricted to the same slice as the node) and the area surrounding the node on the same slice are also used. Augmented atlas-based features consider as a reference the mean intensity value of each augmented atlas region, on the whole volume or restricted to the same slice as the node.

f) Tree: these features embed properties computed on the whole graph, namely the whole striatal volume measured in mm^3 and the ratio between the number of nodes and slices.

Thus, the chosen features are unitless or computed in $mm(^2)$. Besides, the overlap between regions is measured using the Dice's coefficient. Finally, the 717 features assigned to each node feed a classifier. More precisely, Random Forest [15], an ensemble method relying on the prediction from several decision trees, is applied to assess whether the subject associated to each input node has Parkinson's disease. This algorithm offers robust results, controls over-fitting, and easily provides information on the importance of each feature to guide the decision. The average of the probabilities predicted by the Random Forest trees is then considered to classify the whole striatal graph in a Deep Learning-based approach.

E. Classification of striatal graphs using Deep Learning

After applying the Random Forest classifier on each node of the striatal graph, the whole graph is classified using Deep Learning. Supposing the Random Forest classifier is efficient, and to reduce the computation time during training, only important features according to this classifier are used to train the Deep Learning model. Practically, only features with a non-zero Gini importance are selected for the Deep Learning step.

Moreover, the average probability of a node of being associated with Parkinson's disease, as predicted by the Random Forest trees, is used as an additional feature. Because these predictions are made on nodes, without any consideration of their surrounding, filtering techniques are also applied on the graph of the average predicted probabilities. More especially, maximum, minimum (namely dilation and erosion [16]), mean and median graph filters are applied, considering only adjacent nodes as neighbourhood.

In short, for the Deep Learning step, the nodes of the striatal graph are associated with five features derived from the Random Forest node's prediction, and features defined in section II-D considered as important by the Random Forest classifier. Finally, a Graph Convolutional Neural Network (GCN) [17] is applied to the resulted striatal graph to assess whether the associated subject has Parkinson's disease. The chosen architecture, among other parameter settings, is described in the next section.

III. Experimental results

A. Experimental setting

The method was implemented in Python, using Scikit-Learn [18] for Random Forest and PyTorch Geometric [19] for Graph Deep Learning. The experimental setting is describe above.

a) Data distribution: to assess the robustness of the proposed method, 10-fold cross-validation was performed: the subjects were split into ten equal sized folds, each one following the dataset proportion of patients and healthy control subjects. Thus, all the nodes or graphs of a same subject were attributed to the same fold. Moreover, nested cross-validation [20] was also considered to optimize the hyperparameters of the involved models. However, to avoid any bias, the process was adapted to rely on different subjects during the training of the node and graph classification models. To that end, at each of the ten training and testing phases of the cross-validation, one fold was used for testing, the first four remaining folds for training the node classifier and the five remaining folds for training the graph classifier. Balanced accuracy was chosen for each cross-validated metric.

b) Graph creation: the Max-Tree was performed on the brain-masked DaTscan images linearly requantified to 64 gray levels. The 6-connectivity was used to label the binary images obtained at each threshold of the requantified image.

c) Node classification: the hyperparameters of the Random Forest model chosen to be optimized using Grid Search during a 4-fold cross-validation were the number of trees, the maximum depth of a tree, and the class weight strategy.

d) Graph classification: the chosen architecture was made of two alternations of GCN and Rectified Unit (ReLU) layers, followed by a GCN layer. Each GCN layer had 64 hidden channels. A global mean average pooling layer was then applied, followed by a dropout of 50%, and a linear layer. A sigmoid function was finally applied on the two outputs of the network, one for each class. Each model was trained with a batch size of 32 on 100 epochs at the most and using Adam with a learning rate of 0.01 to optimize the binary cross-entropy loss. Moreover, each training was stopped at the epoch leading to the highest average balanced accuracy when fitting the model to the validation set, each validation set being one of the five folds used for graph classification.

TABLE I: Performances of the Random Forest algorithm (mean over the 10 test folds)

Applied Graph Filter	Node classification		Consistent graphs		Consistent graph classification	
	TNR	TPR	HC	PD	TNR	TPR
None	0.812	0.907	61.9%	84.0%	0.861	0.972
Erosion	**0.861**	0.871	**79.7%**	86%	**0.893**	0.948
Dilation	0.757	**0.937**	64.3%	**90.6%**	0.830	**0.973**
Mean	0.821	0.908	71.8%	88.5%	0.861	0.961
Median	0.819	0.908	72.3%	87.7%	0.861	0.963

The Deep Learning process was divided in two steps. First, forward-Sequential Feature Selection (SFS) [21] was performed with 5-fold cross-validation. This greedy algorithm starts by selecting the best feature, according to the chosen cross-validated metric, training the model with only one feature. Then, the best new feature, training the model with two different features, including the first selected one, is added to the selection. Thus, forward-SFS adds features one-by-one until reaching the required number of features. In this article, to reduce computation time, the algorithm ends when the cross-validated metric stops increasing.

Once the features were selected, 5 classifiers were used in an ensemble approach to obtain a robust graph classification. For each classifier, four folds were used for training, and one fold - different for each classifier - for choosing the best epoch. The 5 trained models were then fitted to the graphs in the test fold. One common strategy to make a labeling decision is majority voting: the class predominantly predicted by the classifiers is assigned to the associated test sample. Other strategies rely on the soft rather than the hard labels. For example, the sum [22] or mean over the classifiers of each predicted class output is computed for each test sample, hence associated with the class maximizing that score. In this article, additional strategies are explored, selecting for each test graph the classifier giving the minimal, maximal or median Parkinson's class output when fitted, and assigning to the related test sample the class predicted by the selected classifier.

B. Results

As shown in I, the performances of the Random Forest (RF) algorithm were better when the nodes to classify were associated with Parkinson's Disease (PD), covering 77.9% of the 25942 nodes on the whole dataset. Moreover, results related to real cases of PD, respectively Healthy Control (HC), were improved after applying a dilation, respectively an erosion, on the graph of RF predicted probabilities. As shown by the True Positive Rate (TPR) and True Negative Rate (TNR) values, the RF model generally provided a proper node prediction. Besides, the node manual labelling relied on the supposed uniformity of the node labels belonging to a same striatal graph. This consistency was observed in many RF predictions, with in average, up to 90.6% of PD graphs in the test folds having an unique predicted node label, against 79.7% for HC graphs. The classification performances on these consistent graphs turned out to be yet very encouraging. These

TABLE II: Properties of features preferentially chosen by the Sequential Feature Selection (mean over the 10 Deep Learning trainings)

Feature type	Filter/norm	SFS rank	Selection rate	SFS-1 CV BA	SFS CV BA	RF rank
RF	Dilation	1.0	20.0%	0.882	0.971	-
RF	Mean	1.0	10.0%	0.906	0.983	-
RF	Median	1.0	10.0%	0.905	0.968	-
RF	Erosion	1.0	20.0%	0.873	0.956	-
RF	-	1.0	40.0%	0.884	0.946	-
Intensity	Lateral ventricule	2.0	20.0%	0.538	0.971	58.0
Distance	Median slice	2.25	40.0%	0.5	0.965	83.5
Age	-	2.4	50.0%	0.5	0.943	94.0
Distance	Top slice	3.0	10.0%	0.5	0.967	151
Intensity	Lateral Orbital gyrus	3.0	10.0%	0.603	0.976	70.0
Area	Caudate	3.0	10.0%	0.5	0.935	94.0
Intensity	Medial frontal cortex	3.0	10.0%	0.593	0.968	133
Area	Putamen	3.17	60.0%	0.541	0.962	107
Intensity	Anterior orbital gyrus	3.5	20.0%	0.514	0.961	144

RF predicted probabilities, with or without filtering, turned out to be the most relevant features for the Deep Learning graph classification, as reported in II. Thus, they were always selected first by the SFS method, as they were associated with the highest cross-validation (CV) metric, namely balanced accuracy (BA), when training the model with only one feature (SFS-1). However, adding features related to nodes or subject significantly increased the training CV metric. Among these additional features, the ratio between the areas of the node and the putamen segmented in the atlas, the age and the distance from the median slice were chosen most frequently. Finally, SFS showed good training performances while keeping only in average 7 features among the 212 features highlighted by the RF algorithm. Moreover, SFS offered a complementary selection, as it did not keep the most important features according to the RF algorithm. More particularly, the whole striatal volume, ranked in average as the most important feature by the RF method, was never selected in the Deep Learning step.

Once the features were selected using SFS, the choice of an ensemble approach between the 5 graph classifiers was decisive, as reported in III. While majority voting gave the highest accuracy (ACC) averaged over the test sets, the highest BA was obtained applying minimum voting, and so, associating subjects with healthy controls in absence of consensus between the classifiers. Although leading to potential under-diagnosis, this strategy turned out to give the most reliable Parkinson's disease prediction, as measured by the positive predictive values (PPV). Finally, the performances were noticeably higher in case of full consensus between the classifiers, occurring in average for 91% of the test subjects.

TABLE III: Performances of the Deep Learning graph classifiers according to the ensemble strategy (mean over the 10 test runs)

Decision rule	BA	TPR	TNR	ACC	PPV
Majority voting	0.908	0.946	0.871	**0.931**	0.968
Mean voting	0.905	0.944	0.866	0.929	0.966
Minimum voting	**0.911**	0.891	**0.93**	0.899	**0.981**
Maximum voting	0.865	**0.964**	0.767	0.925	0.944
Median voting	0.884	0.957	0.811	0.928	0.953
Consensus	0.94	0.962	0.918	0.954	0.981

IV. Conclusion

In this article, the proposed method benefited from the complementarity of features computed at different scales. The Random Forest algorithm classified nodes associated with axial slices of the striatum, embedding both local and subject information, such as clinical data and anatomic pattern. It provided relevant features, further used in the Deep learning classification of the striatal graph in an ensemble approach. Besides, the choice of a voting strategy between the involved classifiers turned out to be crucial, and should be further explored. Future improvements will also include the segmentation of the striatum, actually based on an atlas, that will be done using Deep Learning on the Max-Tree of DaTscan images. Finally, this approach will be compared with a Deep learning classification directly made on the Max-Tree of these DaTscan images.

Acknowledgment

Data used in the preparation of this article was obtained on 2024-10-27 from the Parkinson's Progression Markers Initiative (PPMI) database (https://www.ppmi-info.org/access-data-specimens/download-data) RRID:SCR_006431. For up-to-date information on the study, visit http://www.ppmi-info.org/. PPMI - a public-private partnership - is funded by the Michael J. Fox Foundation for Parkinson's Research, and funding partners; including 4D Pharma, Abbvie, AcureX, Allergan, Amathus Therapeutics, Aligning Science Across Parkinson's, AskBio, Avid Radiopharmaceuticals, BIAL, BioArctic, Biogen, Biohaven, BioLegend, BlueRock Therapeutics, Bristol-Myers Squibb, Calico Labs, Capsida iotherapeutics, Celgene, Cerevel Therapeutics, Coave Therapeutics, DaCapo Brainscience, Denali, Edmond J. Safra Foundation, Eli Lilly, Gain Therapeutics, GE HealthCare, Genentech, GSK, Golub Capital, Handl Therapeutics, Insitro, Jazz Pharmaceuticals, Johnson & Johnson InnovativeMedicine, Lundbeck, Merck, Meso cale Discovery, Mission Therapeutics, Neurocrine Biosciences, Neuron23, Neuropore, Pfizer, Piramal, Prevail Therapeutics, Roche, Sanofi, Servier, Sun Pharma Ad- vanced Research Company, Takeda, Teva, UCB, Vanqua Bio, Verily, Voyager Therapeutics, the Weston Family Foundation and Yumanity Therapeutics.

References

[1] W. Poewe, K. Seppi, C.M. Tanner, G.M. Halliday, P. Brundin, J. Volkmann, A.E. Schrag and A.E. Lang (2017). Parkinson disease. Nature reviews Disease primers, 3(1), 1-21.

[2] I. Gayed, I, U. Joseph, M. Fanous, D. Wan, M. Schiess, W. Ondo, and K.S Won (2015). The impact of DaTscan in the diagnosis of Parkinson disease. Clinical nuclear medicine, 40(5), 390-393.

[3] P. R. Magesh, R. D. Myloth and R. J. Tom (2020). An explainable machine learning model for early detection of Parkinson's disease using LIME on DaTSCAN imagery. Computers in Biology and Medicine, 126, 104041.

[4] H. D. Tagare, C. DeLorenzo, S. Chelikani, L. Saperstein and R. K. Fulbright, (2017). Voxel-based logistic analysis of PPMI control and Parkinson's disease DaTscans. Neuroimage, 152, 299-311.

[5] A. Kurmi, S. Biswas, S. Sen, A. Sinitca, D. Kaplun, and R. Sarkar (2022). An ensemble of CNN models for Parkinson's disease detection using DaTscan images. Diagnostics, 12(5), 1173.

[6] A. Ortiz, J. Munilla, M. Martínez-Ibañez, J. M. Górriz, J. Ramírez and D. Salas-Gonzalez (2019). Parkinson's disease detection using isosurfaces-based features and convolutional neural networks. Frontiers in neuroinformatics, 13, 48.

[7] M. Thakur, H. Kuresan, S. Dhanalakshmi, K. W. Lai and X. Wu (2022). Soft attention based DenseNet model for Parkinson's disease classification using SPECT images. Frontiers in Aging Neuroscience, 14, 908143.

[8] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard and L. Jackel (1989). Handwritten digit recognition with a back-propagation network. Advances in neural information processing systems, 2.

[9] P. Coupé, B. Mansencal, M. Clément, R. Giraud, B. Denis de Senneville, V.-T Ta, V. Lepetit, J. V. Manjon. AssemblyNet: A large ensemble of CNNs for 3D Whole Brain MRI Segmentation. NeuroImage, 219, 117026, 2020.

[10] De Senneville, B.D., Manjon, J.V. and Coupé, P., 2020. RegQCNET: Deep quality control for image-to-template brain MRI affine registration. Physics in Medicine and Biology, 65(22), p.225022

[11] J V. Manjon and . Coupe. volBrain: an online MRI brain volumetry system. Frontiers in Neuroinformatics, 10, 30, 2016.

[12] M. Jenkinson, C.F. Beckmann, T.E. Behrens, M.W. Woolrich, S.M. Smith. FSL. NeuroImage, 62:782-90, 2012

[13] P. Salembier, A. Oliveras, and L. Garrido (1998). Antiextensive connected operators for image and sequence processing. IEEE Transactions on Image Processing, 7(4), 555-570.

[14] E., Grossiord, H. Talbot, N., Passat, M., Meignan, and L., Najman (2017, April). Automated 3D lymphoma lesion segmentation from PET/CT characteristics. In 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017) (pp. 174-178). IEEE.

[15] L. Breiman, "Random Forests", Machine Learning, 45(1), 5-32, 2001.

[16] J. Cousty, L. Najman, F. Dias and J. Serra (2013). Morphological filtering on graphs. Computer Vision and Image Understanding, 117(4), 370-385.

[17] T.N. Kipf and M. Welling (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

[18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, B., O. Grisel, M. Blondel et al (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

[19] M. Fey and J.E. Lenssen (2019). Fast graph representation learning with PyTorch Geometric. ICLR (RLGM Workshop)

[20] J. Samper-González, N. Burgos, S. Bottani, S. Fontanella, P. Lu, A. Marcoux, A. Routier, J. Guillon, M. Bacci, J. Wen, A. Bertrand, H. Bertin, M-O Habert, S. Durrleman, T. Evgeniou and O. Colliot, Alzheimer's Disease Neuroimaging Initiative. (2018). Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data. NeuroImage, 183, 504-521.

[21] F. J. Ferri, P. Pudil, M. Hatef and J. Kittler (1994). Comparative study of techniques for large-scale feature selection. In Machine intelligence and pattern recognition (Vol. 16, pp. 403-413). North-Holland.

[22] D. Lu, K. Popuri, G.W. Ding, R. Balachandar and M.F. Beg (2018). Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. Scientific reports, 8(1), 5697.