ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/318134013

A 3D Convolutional Neural Network Approach for the Diagnosis of Parkinson's Disease

Conference Paper · May 2017

DOI: 10.1007/978-3-319-59740-9_32

CITATION
1

READS
119

8 authors, including:

Francisco Jesús Martínez-Murcia
University of Granada
67 PUBLICATIONS **315** CITATIONS
SEE PROFILE

Andrés Ortiz
University of Malaga
113 PUBLICATIONS **742** CITATIONS
SEE PROFILE

Juan M Gorriz
University of Granada
377 PUBLICATIONS **3,799** CITATIONS
SEE PROFILE

Ignacio Alvarez Illan
University of Granada
120 PUBLICATIONS **1,738** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

LAGRANGE - Multimodal and longitudinal biomarker analysis for diagnosis and prediction of Alzheimer's and Parkinson's View project

Intelligent Automated System for Detecting Diagnostically Challenging Breast Cancers View project

All content following this page was uploaded by Francisco Jesús Martínez-Murcia on 05 October 2017.

The user has requested enhancement of the downloaded file.

A 3D Convolutional Neural Network Approach for the Diagnosis of Parkinson's Disease

Francisco Jesús Martinez-Murcia[1⋆], Andres Ortiz[2], Juan Manuel Górriz[1], Javier Ramírez[1], Fermin Segovia[1], Diego Salas-Gonzalez[1], Diego Castillo-Barnes[1], and Ignacio A. Illán[3]

[1] Dept. of Signal Theory, Networking and Communications
18071 University of Granada, Spain
[2] Dpt. of Communications Engineering
29071 University of Málaga, Spain
[3] Dept. of Scientific Computing,
The Florida State University, Tallahassee, FL 32306-4120

Abstract. Parkinsonism is the second most common neurodegenerative disease, originated by a dopamine decrease in the striatum. Single Photon Emission Computed Tomography (SPECT) images acquired using the DaTSCAN drug are a widely extended tool in the diagnosis of Parkinson's Disease (PD), since they can measure the amount of dopamine transporters in the striatum. Many automatic systems have been developed to aid in the diagnosis of PD, using traditional feature extraction methods. In this paper, we propose a novel system based on three-dimensional Convolutional Neural Networks (CNNs), that aims to differenciate between PD-affected patients and unaffected subjects. The proposed system achieves up to a 95.5% accuracy and 96.2% sensitivity in the diagnosis of PD.

1 Introduction

Parkinsonism is the second most common neurodegenerative disease, surpassed only by the prevalence of Alzheimer's Disease [11]. The most common cause of Parkinsonism is Parkinson's Disease (PD), a disease originated due to the progressive loss of dopamine transporters (DaT) of the nigrostriatal pathway, which leads to a decrease in the dopamine content of the striatum [5, 6].

Currently nuclear imaging is being used consistently to assist the diagnosis of PD. The most common drug is ^{123}I-ioflupane (also known as DaTSCAN), a tracer that binds to the DaT in the striatum [3] emitting photons that can be detected using Single Photon Emission Computed Tomography (SPECT) equipment.

DaTSCAN images have been widely used in Computer Aided Diagnosis (CAD) systems, that aim at differentiating between PD affected subjects and normal controls (CTL) [2, 24, 18, 20, 14, 13, 15], or even differentiate between

⋆ Corresponding author: fjesusmartinez@ugr.es

other diseases that lead to parkinsonism, such as multiple system athropy (MSA) or progressive supranuclear palsy (PSP) [22, 21]. These methods use feature extraction techniques, such as moments [18], Independent Component Analysis (ICA) [14], Partial Least Squares (PLS) [20] or Texture analysis [13] to correctly classify PD affected subjects.

Convolutional Neural Networks (CNN) are a particular type of Artificial Neural Networks (ANN) which are becoming increasingly popular in image analysis [10, 15, 22], given their adaptability to many types of projects. They have been used in medical imaging previously [15, 22] with great success. Therefore, we propose that the use of a three-dimensional CNN based on Tensorflow [1] can be of great help to assist in the diagnosis of PD and provide useful information about the variability contained within.

In this work, we present an application of CNNs to the diagnosis of PD and SWEDD subjects. In Section 2 we present the methodology used to build the convolutional neural network and the evaluation methodology. In Section 3 we present and discuss the performance of the system under two different experiments, and finally, at Section 4, we draw some conclusions about this and propose future work.

2 Methodology

2.1 Volume selection

DaTSCAN binds mainly to dopamine transporters at the striatum. Therefore, most of the space contained within these images is not relevant for diagnosis. Without loss of generality, we can extract the volumes of interest containing the striatum and obviate the rest, thus obtaining an significant feature reduction.

The procedure used here consist of a simple thresholding that has been previously used in PD [13, 15], by which we binarize the average of all images in the dataset:

$$I_{BW} = I_{mean} > I_{th} \tag{1}$$

For which the intensity threshold I_{th} is computed as a percentage T of the maximum value of the average image. That is:

$$I_{th} = T \times I_{max} \tag{2}$$

Once we obtained I_{BW}, we use the minimum and maximum indices in all directions where any of the voxels of I_{mean} is greater than the threshold to define a box. Finally, the portion of any image that falls between these coordinates is taken as the input image I_{cut}. Figure 1 shows the final box using $T = 0.35$ superimposed on a sample subject.

2.2 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a specific type of Artificial Neural Network (ANN) that is becoming increasingly important in the Machine

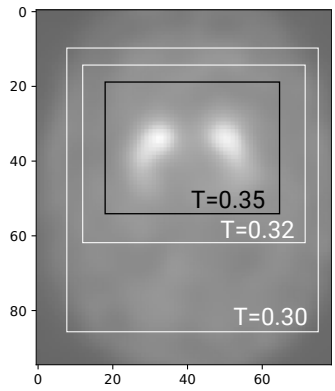

Fig. 1. Example of selected area for different threshold values.

Learning community [10, 4, 15, 22]. They are a bioinspired variant of Multilayer Perceptrons (MLPs), in which the response of a neuron is approximated by a convolution operation. Since 2012, when an ensemble of CNNs [10] achieved lowest error on the ImageNet classification benchmark [19], CNNs prevail over any other pattern recognition algorithm in the literature.

The architecture of CNNs usually comprises many types of layers, of which the most important are: convolution layers, max pooling layers and fully connected layers. Many combinations of these can be found throughout literature [10, 4, 1, 16, 15, 22]. All CNNs share these properties: a local connectivity of the hidden units, use of pooling to introduce translation invariance, and parameter sharing. A 1-layer typical approach would include convolution of the input image with a set of filters and applying its corresponding activation function, downsmapling of the resulting signal by max-pooling and an output MLP that transform the activation signal into probability of classes.

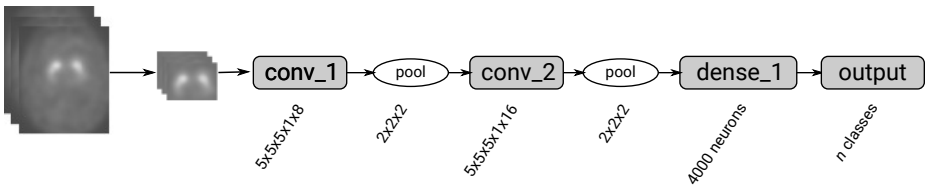

Fig. 2. Schema of our system.

Convolutional Layer CNNs owe their name to the convolutional layer. The operation at a convolutional layer can be defined as follows. Let $\mathbf{V}_{i-1}$ be a tensor the activation map of the $(i-1)$-th layer (in the case of $i = 1$, it is the

input volume). Let $\mathbf{W}_i$ be the set of K filters of the i-th layer, and $\mathbf{b}_i$ a bias term, a vector of length K. Then, the mathematical operation performed at the convolutional layer is:

$$\mathbf{V}_i = f_a\left(\mathbf{W}_i * \mathbf{V}_{i-1} + \mathbf{b}_i\right) \tag{3}$$

with $f_a(*)$ being the activation function, that we will detail later.

In the case of two-dimensional convolutions, $\mathbf{V}_{i-1}$ is a tensor of size $H \times W \times C$ (height, width and number of channels), and $\mathbf{W}_i$ a tensor of size $P \times Q \times R \times K$, containing K filters of size $P \times Q \times R$, with $R_i = K_{i-1}$ and $R_1 = 1$ for the first layer. With this definitions, the k-th convolution term for the k-th filter is:

$$\mathbf{W}_{ik} * \mathbf{V}_{i-1} = \sum_{u=0}^{P-1} \sum_{v=0}^{Q-1} \mathbf{W}_{ik}(P-u, Q-v)\mathbf{V}_{i-1}(x+u, y+v) \tag{4}$$

which, extended to a three-dimensional environment, with a $\mathbf{V}_{i-1}$ of size $H \times W \times D \times C$, and a $\mathbf{W}_i$ of size $P \times Q \times R \times S \times K$ would result in:

$$\mathbf{W}_{ik} * \mathbf{V}_{i-1} = \sum_{u=0}^{P-1} \sum_{v=0}^{Q-1} \sum_{w=0}^{R-1} \mathbf{W}_{ik}(P-u, Q-v, R-w)\mathbf{V}_{i-1}(x+u, y+v, z+w) \tag{5}$$

After the convolution, the activations for each of the K filters are stacked, and passed to the following layer, usually a max-pooling layer.

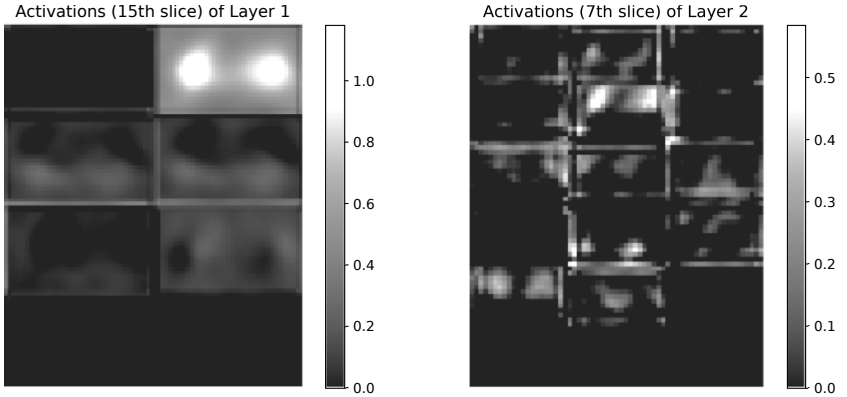

Fig. 3. Summary of the activations in the first and second layers, after feeding a normal control patient with a threshold $T = 0.35$.

The filter size $P \times Q \times R$ is an hyperparameter that is usually set a priori, often making $P = Q = R$ to create a cube, taking values of 3, 5 or even 7 in the literature. In this work we will use a value of $P = Q = R = 5$, with a structure of 2 layers with $K_1 = 8$ and $K_2 = 16$ filters respectivetly. In Figure 3 we display the activation of the filters at the first and second layers after inputting a normal control patient.

The other two hyperparameters that control the output size of the activation layers are stride and zero-padding. Stride controls the step at which the convolution is computed, or in other words, how much overlapping there is between convolutions. That defines the receptive field of a neuron in the convolution layer, which is the part of the image to which each convolution filter is connected to. Using a stride of 1, the convolution is performed at each voxel of the input. With a higher stride, there is less overlapping between receptive fields, and the output volumes will be smaller.

The zero-padding is a convinient technique that provides control of the output volume by padding the input volume with zeros. That way, after the convolution, the output volume can have the same size as the input volume, which in some cases is desirable. In this work we use a stride of 1 and zero-padding of 2 so that the output volume is the same size as the input volume.

Activation The activation function is common to all types of ANNs. It is applied to the output of all the operations performed at the network, and provides an activation of the operations. Many different types of activation functions exist, among them the traditional sigmoid, although in the context of CNN the Rectified Linear Unit (ReLU) has gained a lot of popularity. The function itself is a non-saturating activation function:

$$f(x) = \max(0, x) \tag{6}$$

The ReLU function makes the CNN training several times faster than other approaches, since the calculation of its derivative has a lower computational cost. And it does so without losing any generalization ability [10]. Throughout this work, we will use the ReLU function in all convolutional layers.

Max Pooling The max pooling layer performs a non-linear downsampling by keeping the maximum value over a $M \times M \times M$ block of the activation layer. This layer prevents the following layers from processing non-maximal values, reducing computational load. Furthermore, by reducing the input space and keeping the receptive field of the filters, we can achieve translational invariance.

Currently, there is a trend to discard the max-pooling layer [23], favouring other alternatives such as using smaller filters or increasing the stride of the convolutional layer. However, in this work we use max-pooling layers after both 1st and 2nd convolutional layers, all with a block size of $2 \times 2 \times 2$.

Dense Layers At the end of a series of convolutional and max pooling layers, there is always a fully connected layer, also known as dense layers, in which all neurons all connected to all outputs from the last max pooling step. The structure of this part usually mimic a multilayer perceptron (MLP) in which the input layer is the output of the last max pooling layer, with one or several hidden layers and an output layer with as many neurons as classes. If we consider the convolution and max pooling layers a sophisticated feature extraction system,

this can be considered the high-level reasoning part of the CNN. The activation here is computed by a matrix multiplication and a softmax activation function.

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \quad \text{for} j = 1, \dots K \tag{7}$$

Dropout A popular method for reducing overfitting in fully connected layers is Dropout [19]. It works by "turning off" some neurons at with a probability $1 - p$, and using only the reduced network. After this step, the "off" neurons are turned on again with their last weight matrix. This procedure is repeated in every training iteration. At testing time, all neurons are active, so their outputs are weighted by a factor of p, an approximation of using all possible 2^n networks. In this work we have used a dropout probability of 0.5.

2.3 Dataset

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. The images in this database were imaged 4 + 0.5 hours after the injection of between 111 and 185 MBq of DaTSCAN. Raw projection data are acquired into a 128×128 matrix stepping each 3 degrees for a total of 120 projection into two 20% symmetric photopeak windows centered on 159 KeV and 122 KeV with a total scan duration of approximately 30 - 45 minutes [8].

Afterwards, the images were registered to a custom template [17] using SPM-8 [7], resulting in $91 \times 109 \times 91$ images. A total of $N = 301$ DaTSCAN images from this database were used in the preparation of the article, 158 suffering from PD, 32 SWEDD and 111 normal controls.

Data Augmentation To reduce overfitting and increase the generalization capabilities of our neural network, we have performed a simple data augmentation procedure, by feeding the neural network with the training set and a mirrored (over the sagital plane) version of this set.

2.4 Evaluation

We have performed a stratified 10-fold cross-validation evaluation of our proposed neural network [9]. In each fold, a confusion matrix has been estimated, from which parameters such as accuracy, sensitivity and specificity can be estimated. However, when using more than 2 classes, sensitivity and specificity cannot be computed. Instead, the accuracy and full confusion matrix are given.

To evaluate the performance of our system, we tested our system under two different experiments:

– **Experiment 1**: We tested the performance of the system under the same conditions, but varying the intensity threshold T, as proposed in Section 2.1 (note that due to memory restrictions on the GPU used, the minimum T used was 0.32).
– **Experiment 2**: We tested the system including and excluding the SWEDD subjects from the dataset, to see if there was any difference at all.

3 Results and Discussion

3.1 Experiment 1

In this experiment we show how the performance of our system evolves when varying the threshold T, in a PD vs CTL scheme. Results show that our system achieves high performance in the detection of PD in the PPMI dataset, obtaining up to a 0.955 ± 0.044 of accuracy and 0.962 ± 0.051 sensitivity for a threshold $T = 0.34$. Due to memory restrictions, we could not test the system for $T < 0.32$. However, a global tendency can be seen, in which the performance decreases with the threshold.

In Figure 4 we compare the accuracy obtained with the histogram of the mean image (from which the threshold T and selection area is computed). We can see that most of the voxels are contained in for a threshold $T < 0.3$, and, is in this transition range (between 0.3 and 0.4) where more performance differences are obtained. Afterwards, the performance decreases slowly, at the same pace that the selection box size. This relation is easily seen when compared with the histogram, since the distribution of voxel intensities is concentrated in the lower intensities (background and internals of the brain) and there is a higher variability at the striatum.

3.2 Experiment 2

In this experiment we test the ability of our convolutional model to detect subjects labelled with SWEDD. SWEDD subjects display no evidence of dopaminergic deficit, whilst showing symptoms of Parkinsonism. Since there is no evidence of a decline in dopamine transporters, their images are extremely similar to controls, and current attempts have been unable to differenciate between them [12, 6, 15].

In this test, we use a selection threshold $T = 0.35$ for the images, and perform the same analysis as the previous one, obtaining an accuracy of 0.820 ± 0.068, and the following confusion matrix:

$$\begin{pmatrix} 94 & 27 & 4 \\ 13 & 1 & 2 \\ 4 & 4 & 152 \end{pmatrix} \tag{8}$$

Compared to accuracies obtained excluding SWEDD subjects (0.955 ± 0.044), we can see that the performance of our system degrades. Looking at the confusion

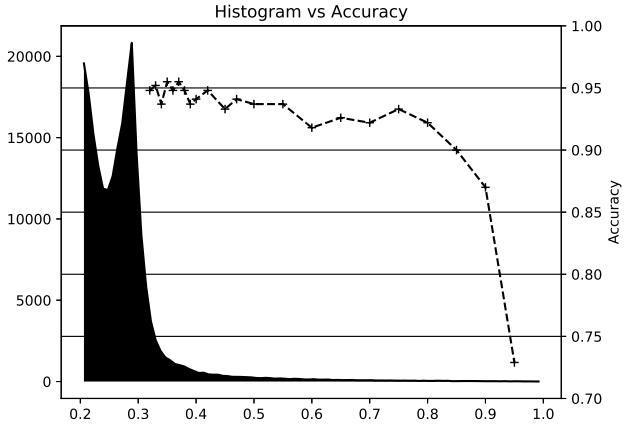

Fig. 4. Accuracy obtained at each threshold, compared with the histogram of the image.

matrix, it is easy to conclude that SWEDD subjects are extremely difficult to classify, since all but one were misclassified. The inclusion of the SWEDD class only meant a source of additional noise in the control class, and by extension, in our classification scheme.

4 Conclusions and Future Work

In this work we have demonstrated the ability of deep learning and convolutional neural networks in the diagnosis of Parkinson's Disease (PD). Particularly, we have applied a three-dimensional approach to convolutional layers that has never been tried before in DaTSCAN images. The high performance obtained in PD detection is an indication of our system's ability to detect patterns in DaTSCAN images. Furthermore, it demonstrates the utility of using 3D convolutional layers in the analysis of three-dimensional medical imaging. The CNN architecture proposed in this work was small, compared to large CNN such as LeNET or GoogLeNET, among others. That, together with the insignificant differences in dopamine transporters, may have had an impact on the smaller ability to differentiate SWEDD patients from controls. In the future, we plan to go deeper on this architecture in order to differentiate these a priori indistinguishable subjects, and obtain a higher performance.

Acknowledgements

This work was partly supported by the MINECO/ FEDER under the TEC2015-64718-R project and the Consejera de Economía, Innovación, Ciencia y Empleo (Junta de Andalucía, Spain) under the Excellence Project P11-TIC- 7103.

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), http://tensorflow.org/, software available from tensorflow.org
2. Benamer, T.S., Patterson, J., Grosset, D.G., Booij, J., de Bruin, K., van Royen, E., Speelman, J.D., Horstink, M.H., Sips, H.J., Dierckx, R.A., Versijpt, J., Decoo, D., Van Der Linden, C., Hadley, D.M., Doder, M., Lees, A.J., Costa, D.C., Gacinovic, S., Oertel, W.H., Pogarell, O., Hoeffken, H., Joseph, K., Tatsch, K., Schwarz, J., Ries, V.: Accurate differentiation of parkinsonism and essential tremor using visual assessment of [123I]-FP-CIT SPECT imaging: the [123I]-FP-CIT study group. Movement Disorders: Official Journal of the Movement Disorder Society 15(3), 503–510 (May 2000), PMID: 10830416
3. Booij, J., Habraken, J.B., Bergmans, P., Tissingh, G., Winogrodzka, A., Wolters, E.C., Janssen, A.G., Stoof, J.C., van Royen, E.A.: Imaging of Dopamine Transporters with Iodine-123-FP-CIT SPECT in Healthy Controls and Patients with Parkinson's Disease. Journal of Nuclear Medicine 39(11), 1879–1884 (1998)
4. Ciresan, D.C., Meier, U., Masci, J., Maria Gambardella, L., Schmidhuber, J.: Flexible, high performance convolutional neural networks for image classification. In: IJCAI Proceedings-International Joint Conference on Artificial Intelligence. vol. 22, p. 1237. Barcelona, Spain (2011)
5. Eckert, T., Edwards, C.: The application of network mapping in differential diagnosis of parkinsonian disorders. Clinical Neuroscience Research 6(6), 359–366 (2007), http://www.sciencedirect.com/science/article/pii/S1566277207000023, neural Networks in the Imaging of Neuropsychiatric Diseases
6. Erro, R., Schneider, S.A., Quinn, N.P., Bhatia, K.P.: What do patients with scans without evidence of dopaminergic deficit (swedd) have? new evidence and continuing controversies. Journal of Neurology, Neurosurgery & Psychiatry pp. jnnp–2014 (2015)
7. Friston, K., Ashburner, J., Kiebel, S., Nichols, T., Penny, W.: Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press (2007)
8. Initiative, T.P.P.M.: PPMI. Imaging Technical Operations Manual, 2 edn. (Jun 2010)
9. Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of International Joint Conference on AI. pp. 1137–1145 (1995), http://citeseer.ist.psu.edu/kohavi95study.html
10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
11. de Lau, L.M.L., Breteler, M.M.B.: Epidemiology of parkinson's disease. The Lancet. Neurology 5, 525–535 (Jun 2006)
12. Marshall, V.L., Reininger, C.B., Marquardt, M., Patterson, J., Hadley, D.M., Oertel, W.H., Benamer, H.T., Kemp, P., Burn, D., Tolosa, E., et al.: Parkinson's disease is overdiagnosed clinically at baseline in diagnostically uncertain cases: A 3-year european multicenter study with repeat [123i] fp-cit spect. Movement Disorders 24(4), 500–508 (2009)

13. Martinez-Murcia, F., Górriz, J., Ramírez, J., Moreno-Caballero, M., Gómez-Río, M., Initiative, P.P.M., et al.: Parametrization of textural patterns in 123i-ioflupane imaging for the automatic detection of parkinsonism. Medical physics 41(1), 012502 (2014)
14. Martínez-Murcia, F.J., Górriz, J.M., Ramírez, J., Illán, I., Ortiz, A.: Automatic detection of parkinsonism using significance measures and component analysis in datscan imaging. Neurocomputing 126, 58–70 (2014)
15. Ortiz, A., Martínez-Murcia, F.J., García-Tarifa, M.J., Lozano, F., Górriz, J.M., Ramírez, J.: Automated diagnosis of parkinsonian syndromes by deep sparse filtering-based features. In: Innovation in Medicine and Healthcare 2016, pp. 249–258. Springer (2016)
16. Payan, A., Montana, G.: Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks. arXiv preprint arXiv:1502.02506 (2015)
17. Salas-Gonzalez, D., Górriz, J.M., Ramírez, J., Illán, I.A., Padilla, P., Martínez-Murcia, F.J., Lang, E.W.: Building a fp-cit spect brain template using a posterization approach. Neuroinformatics 13(4), 391–402 (2015)
18. Salas-Gonzalez, D., Górriz, J.M., Ramírez, J., López, M., Illan, I.A., Segovia, F., Puntonet, C.G., Gómez-Río, M.: Analysis of SPECT brain images for the diagnosis of Alzheimer's disease using moments and support vector machines. Neuroscience Letters 461, 60–64 (Sep 2009)
19. Schmidhuber, J.: Deep learning in neural networks: An overview. Neural Networks 61, 85–117 (2015)
20. Segovia, F., Górriz, J.M., Ramírez, J., Álvarez, I., Jiménez-Hoyuela, J.M., Ortega, S.J.: Improved Parkinsonism diagnosis using a partial least squares based approach. Medical physics 39(7), 4395–4403 (2012)
21. Segovia, F., Gorriz, J., Ramírez, J., Salas-Gonzalez, D.: Multiclass classification of 18 f-dmfp-pet data to assist the diagnosis of parkinsonism. In: Pattern Recognition in Neuroimaging (PRNI), 2016 International Workshop on. pp. 1–4. IEEE (2016)
22. Segovia, F., García-Pérez, M., Górriz, J.M., Ramírez, J., Martínez-Murcia, F.J.: Assisting the diagnosis of neurodegenerative disorders using principal component analysis and tensorflow. In: International Conference on EUropean Transnational Education. pp. 43–52. Springer (2016)
23. Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.: Striving for simplicity: The all convolutional net. arXiv preprint arXiv:1412.6806 (2014)
24. Towey, D.J., Bain, P.G., Nijran, K.S.: Automatic classification of 123I-FP-CIT (DaTSCAN) SPECT images. Nuclear Medicine Communications 32(8), 699–707 (Aug 2011), http://www.ncbi.nlm.nih.gov/pubmed/21659911, PMID: 21659911

View publication stats