Design of intelligent systems for Parkinson's Disease classification using MRI imaging

ABSTRACT

The aim of this project is to study the potential for MRI evaluation of Parkinson's Disease, perfoming experiments to test the effectiveness of different algorithms (SVM, stacking, Deep Learning) and methods in the disease's identification. More specifically, the objective of this project is to develop an automated software that could help to classify patients of Parkinson's disease and aid in the discovery of new regions affected by PD or, for that matter, any neurological disease.

Nowadays, brain disease and dementia research is one of the most important fields of investigation. According to the Parkinson's Foundation, 46.8 million people worldwide are living with some form of dementia. Dementia is the condition characterized by the deterioration of brain functions. This loss of faculties leads to impairment in daily life and can even lead to loss of consciousness. Particularly, it is estimated that there are more than 10 million people worldwide now living with Parkinson's Disease (PD). In this project we explain carefully the different stages of PD. Over the past two decades, various objective measures have been adopted for the differential diagnosis of PD, including electrophysiological and neurophysiological tests. However, the most developed area in providing an objective assessment is neuroimaging. This method involves Magnetic Resonance Imaging (MRI) and computational algorithms. We focus our project in this kind of techniques to build a robust Computer-Aided Diagnosis system.

We divide the brain 3D images in slices (2D images) to study the different brain's regions. To discover which slices of the MRI are the most relevant in the brain for identification of Parkison's Disease region, we intend to test differente feature extraction and feature selection methods. Most of the research is oriented to slices and two-dimensionality transforms. When working with MRI images, which have a number of slices of 157x189x136 in planes X,Y,Z respectively, we have 4035528 voxels with 255 possible values. We would thus find more than one billion different feature vectors. We need to reduce the amount of resources required to describe such a large dataset. Usually, a very efficient strategy to extract features from images is to convert the images to a different domain. Fourier's Transform would bring us a representation of the frequencies making up the image, thus highlighting some information which could be otherwise hidden and inaccesible. The main problem stemming from the utilization of Fourier's Transform is that some information is lost, specifically the time-domain information. It would be a better approach to use a transformation that didn't lose information. The most common in MRI studies is Wavelet transform (2D), which does not have the same problem of loss of time information as Fourier. Therefore, we select discrete wavelet transform as our tool to extract features from MRI images. After feature extraction, we have a matrix of vectors of features which represents our inicial data set. These vectors have thousands of features, each describing a specific characteristic of the elements which they represent. As a result of this huge amount of information, redundancy might be a tough issue. In order to address it, we choose Principal Component Analysis as a powerful statitical tool to reduce the dimensionality and remove redundancy at the same time that the most relevant information is kept. In particular, we set 95% of the amount of variance as the miminimum percentage that the number of components given by PCA needs to explain. All this techniques conform the data preprocessing phase. Prior to any manipulation or information extraction, we need to normalize the whole data set to represent the same reality. Every MRI image should represent the same space, which means that each region of each image has to describe the same brain region. The ventricles should be placed in the same coordinates for each image of each patient. Given that brains vary very much from subject to subject, this is a hard goal to achieve because different genetic inheritance and life experiences to each brain making having its own peculiarities. So, we have to take into consideration two main objectives to be accomplished. The first one is normalize the gray color levels of each image so that

all of them belong to the same scale. The second one is reshape the image to the same bounding box so that all brains have the same measures. To realice the normalization we used Statisticall Parametric Mapping (SMP 12), which is an algorithm that analyses each voxel using a standard statistical test and then reassembles parameters into a spatially normalized image. Taking into account that we have two kind of subjects, patients who suffer Parkinson's Disease and healthy patients, the explained procedures were combined in different ways to reach the hypothesized results and to perform four main experiments:

- The first experiment was to determine which plane of the MRI image is more important to classify subjects. In MRI images, there are three planes of acquisition: X (Axial), Y (Coronal) and Z (Sagittal). To reduce computational time and wasted slices, we selected a plane to extract slices. We have found that Coronal plane is the best one to study when it comes to diagnose Parkinson's Disease.

- After we de decided which plane of X,Y,Z (Y, as we have just said) we were going to use, we wanted to know the accuracy of each slice in order to select wich slices of the brain could be most suitable to extract features from and to base a classification on. This will be our first approach to find the best slices in diagnose.

- We compared the use of white matter or gray matter instead of the whole matter to improve the perfomance of the classification. However, the results were not conclusive, so we decided to carry on using the whole matter as our source of information in order to find the best slices.

- A first optimization of relevant slices, Coronal slices (Y plane), based on Ensemble Learner methods, in particular Stacking, was carried out. First, after dividing the data set in training set and test set, we develop a first phase in which 189 SVM models are trained (one per each Coronal slice). 10-fold cross validation was used to avoid overfitting. Besides, a grid search were made to tune SVM hyperparameters (C and γ). As a result, by the end of this phase, we had a matrix in which each row was a patient (from the test set) and each column refers to a slice, i.e., we had a num_patients x 189 matrix. Each element of the matrix is a prediction corresponding to the patient in the row as a result of the study in each slice. For example, we could have a row like this:

  $$\text{Patient } 1 \rightarrow (0,0,0,1,0,\ldots,0), \text{ where 0 means the patient is ill and 1, healthy.}$$

  We know the real state of the patient (the system doesn't know it at all). Taking this into account, we carried out the second phase, where we try to find which slices are the most important when it comes to diagnose the real state (which is known for us) of the patient. To assure the interpretability of the results, we used a linear model (Logistic Regression) to study which columns (slices) are most relevant in the linear combination to generate the label. We chose the absolute value of the coefficients as an indicator of the importance. This procedure and the results are deeply explained in Section 4.3. We studied the 10 best slices to take into consideration for diagnosing Parkison's Disease. The precision values obtained are greater than 72%. In addition, the recall values obtained are algo greater than 80%.

- In the second optimization, which had the same structure, we established a new way of facing neurodegenerative illneses' diagnose. We use Deep Learning, in particular Convolutional Neural Network, as the model of the first phase, i.e., the model to learn the physical features of Parkinson's Disease in a Magnetic Resonance Image. In this case, no data preproccesing was carried out in order to take advantage of the power of deep learning. We designed the topology of the net using convolutional layers, dropout layers and pooling

layers. We fixed the number of epochs to 100 and we used early stopping to reduce computational time when the neural network doesn't seem to be learning new features. After different simulations, the final topology of the net is: A 2D convolutional layer (320 parameters), a pooling layer, a 2D convolutional layer (4624 parameters), a pooling layer, a 2D convolutional layer (2320 parameters), a pooling layer and dropout layers (dense to finish). As in the first optimization, we studied the 10 best slices (images were specified in Section 4.3 and 4.4 to visualize the slices) to find out Parkinson's Disease. However, results were not as promising as in the first approach, so we have found that deep learning is not suitable for problems with too little amount of data.

Furthermore, we were very interested in developing each concept and tool used mathematically, in order to guarantee the rigor needed in this challenging issue. The precision values obtained are greater than 72%. In addition, the recall values obtained are algo greater than 80%. This suggests that the proposed workflow and its application could help in the investigation of Parkinson's disease and aid in the research of other neurodegenerative diseases, improving their diagnostic accuracy and identifying the most relevant zones and regions of the brain associated with each disease.

KEY WORDS

Parkinson's Disease (PD), Feature Extraction, Discrete Wavelet Transform (DWT), Feature Selection, Principal Component Analysis (PCA), Classification, Support Vector Machine (SVM), Logistic Regression, Stacking, Ensemble Learner, Deep Learning, Convolutional Neural Networks (CNN), Pooling, Dropout, Convolution.

Diseño de sistemas inteligentes para la clasificación automática de la enfermedad del Parkinson utilizando imágenes MRI

RESUMEN

El objetivo de este proyecto es estudiar la potencialidad de las imágenes de resonancia magnética (MRI) en la evaluación de la enfermedad del Parkinson, llevando a cabo experimentos y probando la efectividad de distintos algoritmos (SVM, stacking, Deep Learning) y métodos para el diagnóstico de la enfermedad.

Para la extracción de características de las imágenes utilizamos la Transformada Discreta Wavelet 2D, seguido de Análisis de Componentes Principales (PCA) para la selección y reducción de características. Después, aplicamos Máquinas de Soporte Vectorial (SVM), Regresión logística y Redes Neuronales Convolucionales para la clasificación y optimización.

Para determinar qué cortes de la imagen MRI son los más relevantes en el cerebro para identificación del Parkinson, una optimización basada en Ensemble Learners, más concretamente Stacking, ha sido llevada a cabo. Además, introducimos las Redes Neuronales Convolucionales como una nueva alternativa en la batalla que supone el diagnóstico precoz de este tipo de demencia. Llevamos a cabo una clasificación entre pacientes enfermos de Parkinson y pacientes sanos (grupo de Control), a fin de valorar nuestros resultados.

Por otra parte, nuestro objetivo ha sido desarrollar cuidadosamente el trasfondo matemático subyacente a las herramientas y conceptos usados con el objetivo de garantizar el rigor tan necesario ante un problema de tal calibre. Obtenemos unos valores de *precision* mayores al 72%. De la misma manera, obtenemos un *recall* mayor del 80%. Dichos resultados sugieren que la metodología de trabajo aplicada en este Trabajo Final de Grado pudieran ayudar en la investigación y diagnóstico del Parkinson, así como otras enfermedades neurodegenerativas identificando las zonas más relevantes del cerebro asociadas a la enfermedad.

PALABRAS CLAVE

Enfermedad del Parkison (PD), Extracción de características, Transformada Discreta Wavelet (DWT), Selección de Características, Análisis de Componentes Principales (PCA), Clasificación, Máquinas de Soporte Vectorial (SVM), Regresión Logística, Stacking, Ensemble Learner, Deep Learning, Redes Neuronales Convolucionales (CNN), Pooling, Dropout, Convolución.