Multi-Objective Genetic Algorithm to find the Most Relevant Slides in MRI for Parkinson's Disease

Alberto García, Olga Valenzuela, Ignacio Rojas

Abstract -The aim of this paper is to study the potential use of an intelligent/automatic classification system for Parkinson's Disease (PD), using magnetic resonance images (MRI). For feature extraction of MRI Discrete Wavelet Transform has been used, followed by minimum Redundancy Maximum Relevance critérium (mRMR) for feature selection and Principal Component Analysis (PCA) for feature reduction. We then applied Support Vector Machine (SVM) for classification and Genetic Algorithms (GA) for optimization. To discover which slices and regions of the MRI are the most relevant in the brain for identification of Parkinson's Disease region, optimization was carried out using a multi-objetive genetic algorithm. The slices obtained in this optimization process were consistent with those recommended by medical experts. The methodology presented outperformed most of the research available in the bibliography, achieving accuracies of 95% in classification of subjects. 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 regions of the brain associated with each disease.

Keywords-Parkinson Disease (PD), Magnetic Resonance Image (MRI), Discrete Wavelet Transform (DWT), minimum Redundancy Maximum Relevance (mRMR), Support Vector Machine (SVM), Genetic Algorithms (GA).

I. INTRODUCTION

Nowadays, brain disease and dementia research is one of the most important fields of the investigation. According to the Parkinson's Foundation, 46.8 million people worldwide are living with some form of dementia. Previous estimates for the year 2020 were that there would be 42.3 million people globally with dementia. Already that number has been surpassed by more than 4 million people, two years earlier than predicted. The same report predicted the numbers of people with dementia will double every 20 years. Dementia is the condition characterized by the deterioration of brain functions. This deterioration or loss of faculties leads to impairment in daily life and can even lead to loss of consciousness. It is estimated that there are more than 10 million people worldwide now living with Parkinson's Disease (PD from now on) [1]. This fact makes the research of the disease very important, since an early diagnosis could slow the

development of the disease. Currently there is no cure for Parkinson's, but there are drugs which inhibit its development, giving patients a minimum quality of life for an extended period. The main diagnosis methods consist of using clinical criteria, which are based on medical evaluation and covers a range of tests performed on the patients in order to address their stage. Current diagnoses rely on the presence of motor signs meaning that the patient is invariably diagnosed with PD at an advanced stage. Moreover, any neuroprotective therapy initiated at such a late stage may have fewer substantial effects on limiting the disease's progression. Thus, it is critical to find valid and objective biomarkers to distinguish early PD patients from the healthy population. Over the past two decades, various objective measures have been adopted for the differential diagnosis of PD, including a range of olfactory, electrophysiological and neurophysiological tests [2]. However, the most developed area in providing an objective assessment is neuroimaging.

There are relevant paper in the bibliography that construct automatic classifier for detecting Parkinson Disease [3-9].
O. Cigdem et al. [3] present a methodology using voxel-based morphometry (VBM) technique, to compare morphological alterations of PDs versus healthy controls (HCs) in gray matter (GM) and white matter (WM). The effects of using different covariates (i.e. total intracranial volume (TIV), age, sex and combination of them) as well as two different hypotheses, t-contrast and f-contrast, on classification of PD from HCs were analyzed in a population of 40 PDs and 40 HCs obtained from the PPMI dataset. The maximum accuracies of distinguishing between PDs and HCs are obtained when TIV is used as a covariate and f-contrast is used for model building: 73.75%, 72.50%, and 93.7% for GM, WM, and the combination of them, respectively.

In P. Peran et al. [4] a methodology based on multimodal MRI approach to identify brain differences was presented that could discriminate between matched groups of patients with multiple system atrophy (MSA, 29 patients), Parkinson's disease (26 patients), and healthy controls (26 patients). Two different MSA variants, namely, MSA-P, with predominant parkinsonism, and MSA-C, with more prominent cerebellar symptoms were analyzed. Using whole-brain voxel-based MRI, authors combined gray-matter density, T2* relaxation rates, and diffusion tensor imaging scalars to compare and discriminate PD, MSA-P, MSA-C, and healthy controls. Using Logistic regression

Alberto García, Ignacio Rojas and Olga Valenzuela
University of Granada
CITIC-UGR, Granada (Spain)

analysis, authors showed that around 95% of discrimination between PD and MSA patients can be obtained.

An interesting approach was presented by S. Talai et al [5] to classify patients with Parkinsons disease (PD) and patients with progressive supranuclear palsy (PSP), an atypical variant of PD. Apparent diffusion coefficient (ADC) parameter maps were calculated based on diffusion-tensor MRI datasets. Mean ADC values were determined in 82 brain regions using an atlas-based approach. The extracted mean ADC values for each patient were then used as features for classification using a linear kernel SVM classifier. To increase the classification accuracy, a feature selection was performed, which resulted in the top 17 attributes to be used as the final input features. A leave-one-out cross validation based on 56 PD and 21 PSP subjects revealed that the presented method is capable of differentiating PD and PSP patients with an accuracy of 94.8%.

In L.Salamanca et al. [6] a novel approach based on the structural characterization of brain areas affected in PD by diffusion magnetic resonance imaging (dMRI) is analyzed. The authors proposed a novel diagnosis pipeline based on the Fisher vector algorithm. This technique allows for a precise encoding into a high-level descriptor of standard diffusion measures like the fractional anisotropy and the mean diffusivity, obtained from the regions of interest (ROIs) typically involved in PD. In a test cohort of 50 PD patients and 50 controls, the implemented methodology obtained good results (77% of accuracy) using a logistic linear regressor for classification of each ROI independently, which are subsequently combined into a single classification decision.

The novel methodology presented in this paper involves MRI and hybridization of intelligent algorithms in order to first: obtain the most relevant slides of the MRI images (to the best of our knowledge, the papers presented in the bibliography require the information of an expert in order to select the most relevant slides); and second: constructing an accuracy classifier with the most relevant features. We also intend to test different feature extraction and feature selection methods. Most of the research is oriented to slices and two-dimensionality transforms (2D Wavelets). A significant goal of this paper is to explore the main regions affected in PD and find the most relevant slices. To this end, we decided to use NGSA-II, an established and proven Multi-Objective Genetic Algorithm. The results of a SVM classifier are finally presented.

II. MATERIALS
A. MRI acquisition

The data used in this contribution are obtained from the PPMI dataset (www.ppmi-info.org/data)[1]. The protocol included T1-weighted MRI images based on a scanner by Siemens with acquisition plane = sagittal, acquisition type = 3D, coil = Body, flip angle = 9.0°, matrix X/Y/Z = 240.0/256/176 pixel, mfg model = TrioTim, pixel

spacing X/Y = 1.0/1.0 mm, pulse sequence = GR/IR, slice thickness = 1 mm, and TE/TI/TR = 2.98/900/2300 ms. The PPMI is a landmark observational clinical study to comprehensively evaluate cohorts of significant interest using advanced imaging, biologic sampling and clinical and behavioral assessments to identify biomarkers of PD progression. PPMI is taking place at clinical sites in the United States, Europe, Israel, and Australia. Data and samples acquired from study participants will enable the development of a comprehensive Parkinson's database and biorepository, which is currently available to the scientific community to conduct field-changing research.

B. Image Preprocessing

Spatial normalization is required when dealing with MRI images of different subjects since it is needed to bring all of them to a common space.
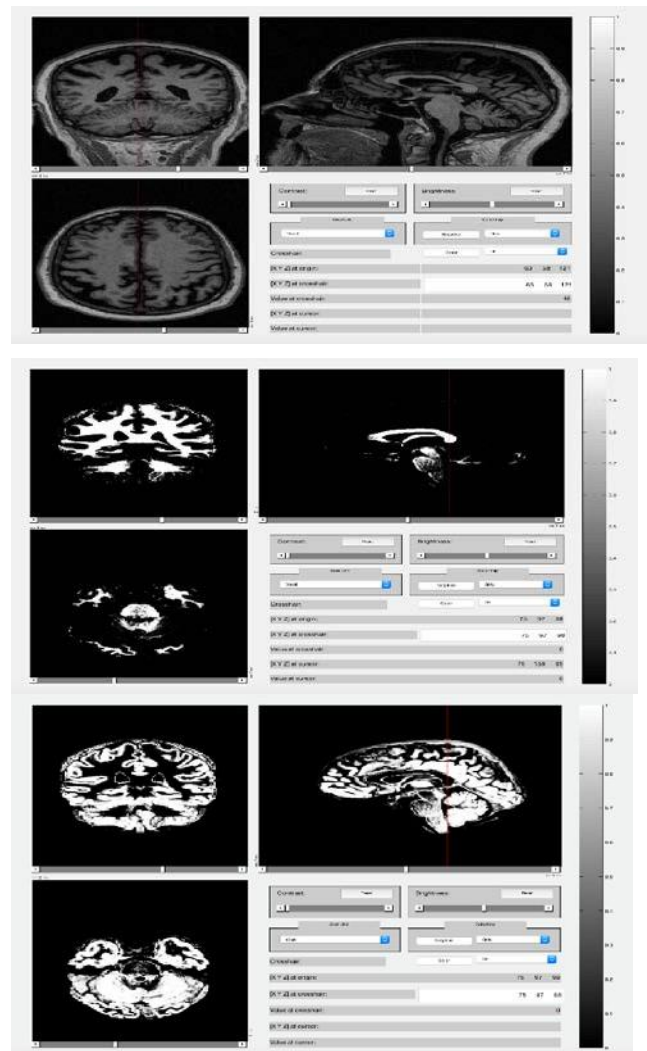

Figure 1.Whole matter image (top), gray matter image (middle), white matter image (bottom).

To run the normalization of the images we used SPM12 toolbox for Matlab [10]. Statistical Parametric Mapping (SPM) refers to the construction and assessment of spatially extended statistical processes used to test hypotheses about functional imaging data. These ideas have been instantiated in software that is called SPM. The SPM software package has been designed for the analysis of brain imaging data sequences. The sequences can be a series of images from different cohorts, or time-series from the same subject. Every subject in the experiments around this project, was spatially normalized to a bounding box of 157x189x136 voxels in the X, Y, and Z directions respectively and voxels (remember that one voxel is a three dimensional pixel) reshaped so that each one measures 1mm x 1mm x 1mm. Brain images have also been segmented so that whole matter (W images), gray matter (C1 images) and white matter (C2 images) are available in different files. The effect of this segmentation is shown in Figure 1 using "Tools for NIfTI and ANALYZE image" [11].

C. Subjects

We downloaded more than 250 brain images in DICOM format. DICOM is an acronym for Digital Imaging and Communications in Medicine. DICOM is both a communications protocol and a file format, which means it can store medical information, such as ultrasound and MRI images, along with a patient's information, all in one file. The format ensures that all the data stays together, as well providing the ability to transfer said information between devices that support the DICOM format. After deleting corrupt files and duplicated images, our data base was formed by 214 images (67 Control Patients and 147 Parkinson's Disease Patients). The size of the data base is 10 GB (almost 30 GB after normalization and segmentation).

III.   NOVEL METHODOLOGY FOR SLICE OPTIMIZATION

The patients from the PPMI database are normalized to MNI space in order to find the best Coronal slices for classification, with the goal to define a set of Y slices that provided the maximum accuracy. In order to do this, images were preprocessed and 2D wavelet coefficients were extracted to form a feature matrix. Using a feature selection algorithm called mRMR we extracted the best features from the matrix, then the dimension of the feature vectors was reduced using PCA of 95 % and the resulting dataset was finally used to train an SVM to perform classification. In order to find the best combinations of Coronal slices we used Genetic Algorithms, and a set of different solutions were extracted from the Pareto front and its histogram. When performing classification of the obtained slices, the method achieved an accuracy of 95%

A. Feature Extraction

Wavelet transform (WT) represents a very well-known thecnique: a windowing technique with variable size. Thus,

it preserves both time and frequency information of the signal. The development of signal analysis is shown in Figure 2. When applied to 2-D images, 1D-DWT is applied separately to each dimensions, resulting in four sub-bands (Low-Low, Low-High, High-High, and High-Low or LL, LH, HH, HL). In this two dimensional case, LL is regarded as the approximation coefficients of the image, while LH, HH, and HL remain as the detailed coefficients. To compute the next level of decomposition, the same process is applied to the approximation coefficients. As the level of decomposition is increased, further sub-bands are computed and more compact yet coarser approximation coefficients are obtained.
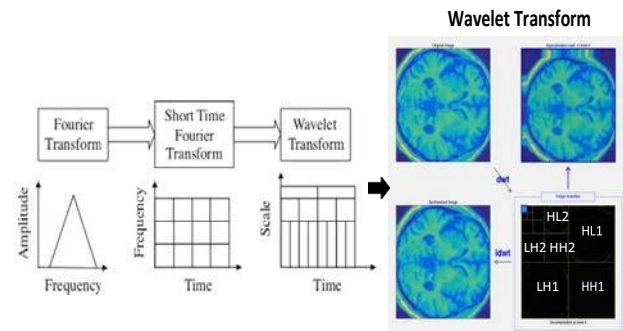

Figure 2. Block diagram of two dimension wavelet decomposition with application to images

There are different type of wavelets that can be used for image processing in MRI analysis. In the presents contribution, an empirical study (by evaluating the accuracy of the system obtained) has been performed using seven types of wavelets (Figure 3). The Db-4 and Biorthogonal 3.3wavelet functions up to level 2 have the best trade-off between efficiency and computational cost. They extract the largest amount of information because they are more precise in smoothly changes, so in order to get the 2D Wavelet coefficients , both, the Daubechies function of order 4 up to level 2 for each slice or the biorthogonal 3.3 are the best candidate (in this contribution, the biorthogonal 3.3 is used).
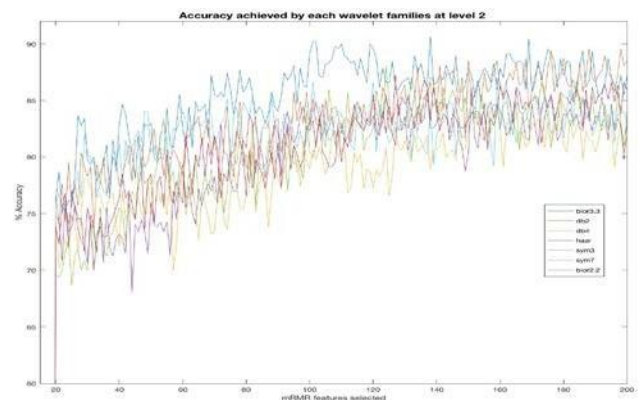

Figure 3 Accuracy achieved by each wavelet family at level 2 for different number of mRMR features selected

B. Feature Selection: mRMR and PCA

First, the minimum Redundancy Maximum Relevance (mRMR) by Peng et al is briefly described [12]. For discrete variables, mutual information I of two variables x and y measures the level of similarity between them and is defined according to their joint probabilistic function p(x, y) and their respective marginal probabilities p(x) and p(y):

$$II(xx, yy) = \sum_{ii,jj} pp(x_{ii}, yy_{jj}) \log \frac{pp(x, yy)}{pp(xx)pp(y)}$$

Minimum redundancy intends to select the features such that they are mutually maximally dissimilar, making the feature set more representative. This minimum redundancy condition is:

$$\min WW_{II}, \quad WW_{II} = \frac{1}{|SS|^2} \sum_{ii,jj \in SS} (ii, jj)$$

Where |S| denotes the number of features in S, the feature subset we are seeking from the feature set, and i, j are two different features. Maximum relevance condition is met when total relevance of all features are maximized. Relevance measures the level of discriminant power of features according to their difference between target classes. Mutual information is again used between target classes h = {$h_1, h_2, \dots, h_k$} and features i. In this way, I(h, i) quantifies the relevance of feature i for the classification task. Thus, maximum relevance means to maximize the total relevance of all features in S:

$$\max VV_{II}, \quad VV_{II} = \frac{1}{|SS|^2} \sum_{ii \in SS} ii) h,$$

The mRMR feature set is obtained optimizing both conditions $\min W_I$ and $\max V_I$. This optimization requires combining them into a single criterion function, two of the simplest combination criteria being:

$$MID = \max(W_I - WW_I); \quad MMIIMM = \max \left(\frac{W_I}{WW_I} \right)$$

Where MID stands for Mutual Information Difference criterion and MIQ stands for Mutual Information Quotient criterion. Either of the criteria being used, features are selected in an incremental way, hence sorting them in an ordered fashion according to their mRMR rank achieved. It is noteworthy to remember that in order to reduce noise and improve the performance of the classifier later on, it is useful (and recommended) to discretized the data into categorical data with zero mean value and unit variance [13].

Finally, PCA is applied in order to reduce the final number of coefficient that will be the input to the SVM classifier. PCA is an efficient tool to reduce the dimension of a data set consisting of a large number of interrelated variables while retaining the most significant variations. It is achieved by transforming the data set to a new set of ordered variables according to their degree of variance or importance.

C. NSGA-II

A important goal of this paper is to know which coronal slices achieved the best classification accuracy. In order to find the best set to perform the classification, a Multi-Objective Genetic Algorithm called NSGA-II is used [14]. Genetic Algorithms (GA)is an optimization tool which is based on Darwin's survival of the fittest principle, which states that an initial population of individuals evolves through natural selection in such a way that the fittest individuals have a higher chance of survival. The GA maintains a population of competing feature transformation matrices. To evaluate each matrix in this population, the input patterns are multiplied by the matrix, producing a set of transformed patterns which are then sent to a classifier. The classifier typically divides the patterns into a training set, used to train the classifier, and a testing set, used to evaluate classification accuracy. The accuracy obtained is then returned to the GA as a measure of the quality of the transformation matrix used to obtain the set of transformed patterns. Using this information, the GA searches a transformation that minimizes the dimensionality of the transformed patterns and maximizing classification accuracy. Each feature is encoded into a vector called a chromosome. One element of the vector represents a gene. Each bit in the binary vector is associated with a feature. If the i-th bit of this vector is equal to one then the i-th feature is allowed to participate in classification. All of the chromosomes make up of a population and are estimated according to the fitness function. In this paper, a population of individuals consisting of bit strings of 161 bits is used, each bit coding whether a specific slice should be selected or not. Only slices from Y=20 to Y=180 were tested. The two fitness functions that the algorithm has to optimize (minimize) were:

- f(1) = 100 − accuracy(%)
- f(2) = number of slices

A population of 100 individuals were used and ran the optimization for 200 generations. Each individual of each generation was stored in a data file along with the accuracy achieved, so that we could perform a post-optimization analysis.

D. Classification

For classification purposes, a Support Vector Machine (SVM) is used [15]. In order to train and test an SVM model, it is first trained with a training set for which features and target labels are known. Hyperplane parameters are fit so that the model achieves its best accuracy separating the different data points, and then, a testing set for which labels are known is fed into the SVM model and classified, getting a vector of labels as output. This output labels are compared to the actual target labels of the input feature vectors and the SVM model accuracy is obtained.

IV. RESULTS

The output of the Multi-Objetive Genetic Algorithm s a Pareto front. A Pareto front is the most usual representation of the result of GA and this graph shows the dominant individuals.
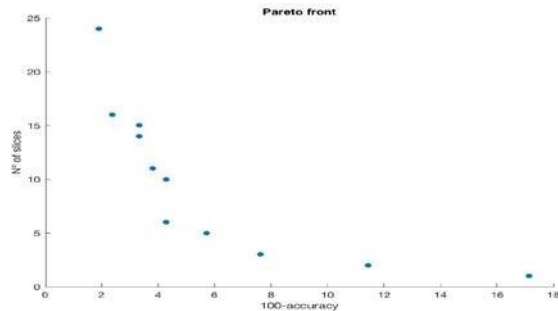

Figure 4 A Pareto front diagram. It shows the trade-off between accuracy and number of slices used

An individual is called dominant if none of the minimization functions can be improved without worsening the others; this means an individual belongs to the Pareto front if for a given accuracy, no other individual has the same accuracy.

The Pareto front obtained in the proposed methodology is presented in Figure 4. We obtained the histograms of slices for Pareto solution, shown in Figure 5. Observation of the histograms gives an idea of the set of best coronal slices to classify subjects. Since the genes of the individuals that are outperformed by others tend to disappear with further generations, the dominant individuals, which have dominant genes, exhibit similar genome, so the frequency of each slice in the Pareto front is somewhat an approximation of the relevance of that slice when extracting features for classification.
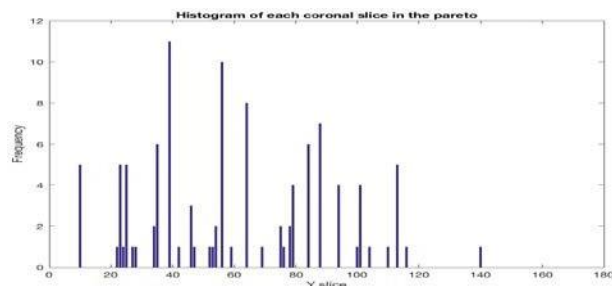

Figure 5 Histogram of pareto. There are some dominant slices

In order to show results on a MRI image, we selected the slices from the best Pareto solution. This solution has 24 slices with the highest accuracy of classification (sorted by relevance: 29, 42, 43, 44, 47, 54, 58, 61, 65, 66, 71, 73, 75, 78, 83, 88, 98, 103, 107, 113, 120, 129, 132 and 135). Furthermore, we selected slices from the Pareto front that appeared on at least 3 individuals, which occurred in 13 slices (58, 75, 83, 54, 103, 29, 42, 44, 132, 98, 113, 120, 65). These results (best slides) are shown in Figures 6 and 7 respectively.
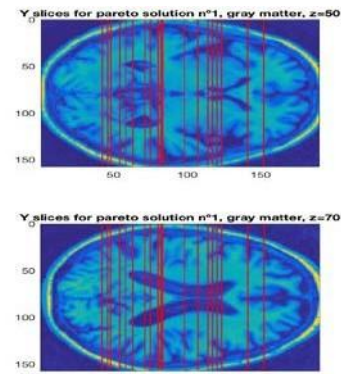

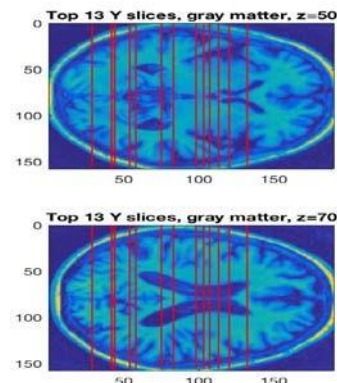Figure 6 Best coronal slices of the best pareto solution

Figure 7 Top 13 coronal slices (most frequent) of the pareto front

The accuracy of the classification system (SVM) is analyzed for different solutions of the Pareto front. Firstly, we used the best Pareto solution to extract features using the 24 slices with the highest accuracy which formed this solution. Secondly, we used the top 13 slices, according to their frequency of occurrence in the Pareto front, which appeared on at least 3 individuals' data. In this way we defined SET1 and SET2 respectively. For feature extraction we used the 2D wavelet coefficients of gray matter, for each Coronal slice using the transform bior3.3 up to level 1. For feature selection we selected the top 100 mRMR features for each slice. For feature selection we used PCA and kept 95% of the variance in the data. For classification, we used SVM with RBF kernel. The cost parameter C was determined from $\{2^{-5}, 2^{-3}, 2^{-1}, 2^1, 2^3, 2^5, 2^7, 2^9, 2^{11}, 2^{13}, 2^{15}\}$ and the gamma parameter γ from $\{2^{-15}, 2^{-13}, 2^{-11}, 2^{-9}, 2^{-7}, 2^{-5}, 2^{-3}, 2^{-1}, 2^1, 2^3\}$ using gridsearch and a 10-fold CV. Subjects were randomly selected to belong to the training set (80% of whole set) or testing set (20% of whole set) ten times independently and the results were averaged.. The results of the classification process is presented (confusion matrix) in Figure 8, for SET1.
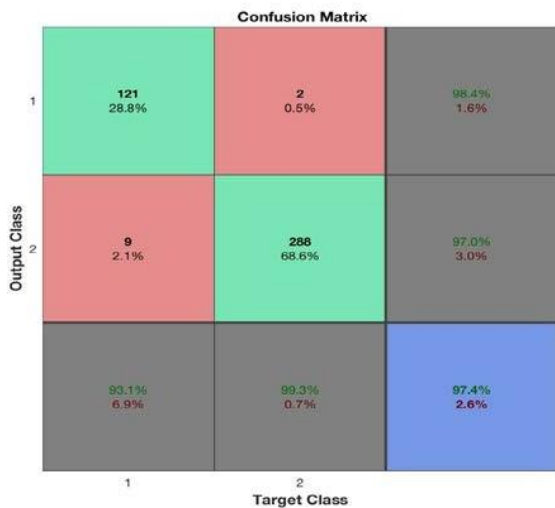
16

Figure 8 Confusion matrix of SET1, formed by slices of the best pareto solution

V.   CONCLUSION

The depth of information that MRI provides about the soft tissue anatomy has dramatically improved the quality of brain pathology diagnosis and treatment. However, the amount of data is far too much for manual interpretation and hence there is a clear need for automated analysis tools. Consequently, in this study we combined MRI imaging with algorithm-based extraction of features to allow automatic classification of subjects into PD or Control groups at the highest accuracy possible. The main object of this paper was to design and develop an advanced framework to automatically classify extracted MRI data, providing best possible accuracy in assessment of brain function. The classifications of Alzheimer's, the most common dementia, or Parkinson's, are often based around evaluation of the gait, and more general motor skills of the individual. But this Parkinson's study, using MRI images and classifying patients with information extracted from these images, is novel. Using the main techniques seen in the bibliography in this type of dementia study, our classifier was able to classify between PD and Control subjects with accuracy of up to 95%.

Not only is it important to obtain an accurate classification result, it is also relevant to obtain information about which areas of the brain (in our case, the slides) are most relevant to the classification. In this sense, and to the best of our knowledge, this is a very novel paper, presenting a methodology for selecting the best slides for Parkinson's Disease

ACKNOWLEDGMENTS

REFERENCES

[1] http://www.parkinson.org/understanding-parkinsons/what-is-parkinsons (2018).

[2] Kashmere, Jodi; Camicioli, Richard; Martin, Wayne (2002) Parkinsonian syndromes and differential diagnosis. Curr Opin Neurol 15: 461 466.

[3] Cigdem, Ozkan; Beheshti, Iman; Demirel, Hasan, Effects of different covariates and contrasts on classification of Parkinson's disease using structural MRI. Computers in Biology and Medicine Vol. 99, pp.173-181, 2018.

[4] Peran, Patrice; Barbagallo, Gaetano; Nemmi, Federico; et ál. MRI supervised and unsupervised classification of Parkinson's disease and multiple system atrophy. Movement Disorders, Vol.33, N.4, pp.600-608, 2018

[5] Talai, Sahand; Boelmans, Kai; Sedlacik, Jan; et ál. Automatic classification of patients with idiopathic Parkinson's disease and progressive supranuclear palsy using diffusion MRI datasets, Medical Imaging 2017: Computer-Aided Diagnosis. Proceedings of SPIE, Vol.10134, 2017.

[6] Salamanca, Luis; Vlassis, Nikos; Diederich, Nico; et ál. Improved Parkinson's Disease Classification from Diffusion MRI Data by Fisher Vector Descriptors. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, 2015

[7] Gu, Quanquan; Zhang, Huan; Xuan, Min; et ál. Automatic Classification on Multi-Modal MRI Data for Diagnosis of the Postural Instability and Gait Difficulty Subtype of Parkinson's Disease. Journal Of Parkinsons Disease. Vol.6, N.3, pp.545-556, 2016

[8] Pazhanirajan, S.; Dhanalakshmi, P. MRI Classification of Parkinson's Disease Using SVM and Texture Features, 2nd International Conference on Computer and Communication Technologies Ubicación: CMR Tech Campus, Hyderabad, INDIA Fecha: JUL 24-26, 2015

[9] Zhang, Li; Liu, Chang; Zhang, Xiujun; et ál.. Classification of Parkinson's disease and Essential Tremor Based on Structural MRI. 7th International Conference on Cloud Computing and Big Data (CCBD) Macau, Peoples R China NOV 16-18, 2016.

[10] John Ashburner, Gareth Barnes, Chun-chuan Chen, Jean Daunizeau, Rosalyn Moran, Rik Henson, Volkmar Glauche, and Christophe Phillips. SPM8 Manual The FIL Methods Group (and honorary members). Functional Imaging Labora- tory, pages 475–1, 2013. ISSN 09621083. doi: 10.1111/j.1365-294X.2006.02813.x.

[11] Jimmy Shen "Tools for NIfTI and ANALYZE image" http://www.mathworks.com/matlabcentral/fileexchange/8797

[12] H C Peng, F H Long, and C Ding. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8):1226–1238, 2005.

[13] C Ding and H Peng. Minimum redundancy feature selection from microarray gene expression data. Bioinformatics and Computational Biology, 3(2):185–206, 2005. ISSN 0219-7200. doi: 10.1142/S0219720005001004

[14] K. Deb ; A. Pratap ; S. Agarwal ; T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, Vol.6, Issue: 2, 2002.

[15] Mavroforakis, ME; Theodoridis, S A geometric approach to support vector machine (SVM) classification. IEEE Transactions on Neural Networks Vol.17, 2006.