Computer Aided Diagnosis for Parkinson's disease detection: State-of-the-Art

Olfa Laouini[1,2], Safouane Hamdi[1,3]

[1]University of Tunis El Manar, Faculty of Medicine of Tunis, Biophysics Department
[2]University of Tunis, Productivity Research Center, National Superior Engineering School of Tunis
[3]University of Sfax, National School of Electronics and Telecommunications of Sfax, Department of Telecommunications,

Abstract—Parkinson's disease (PD) is a chronic neurodegenerative disorder occurs at the central nervous system. PD involves the malfunction or death of certain vital nerve cells in the brain called neurons. Factor which reduce obviously the quality of life of patients. Still there is no particular test to confirm the existence of PD. In fact, typical clinical procedures leads to errors, excessive costs and provide insufficient adequate information. Therefore, diagnosing PD may need deep experience and highly skilled specialists. Improvement of diagnosis and assessment in early stages of disease can gets resolved by the integration of high-performance computers. Advancements in artificial intelligence has led to the emergence of expert systems for medical applications. Computer Aided Diagnosis (CAD) can be used to improve accuracy, sensitivity and specificity of diagnosis as well as making the detection of PD more time efficient. CAD is a concept established by taking into account the opinion of physicians as well as the data yielded by computers in an equal way. Using CAD, the performance achieved by computers does not have to be comparable to that made by physicians, but they both need to be complementary. In fact, a large number of CAD systems have been employed for assisting physicians in the early detection of a wide range of tumors and lesions. This paper describes the state of the art of CAD systems used in the detection of cerebral neurodegenerative disorders, such as, PD. Besides, it discusses the latest proposed solutions for each step of processing, their performance and limits.

Index Terms—Parkinson's disease (PD), Computer Aided Diagnosis (CAD), Support Vector Machine (SVM), machine learning.

1 Introduction

Parkinson's disease is a neural condition, in which, part of the brain becomes progressively damaged through certain number of years. Normally, in the human brain, there are a large number of dopamine-producing cells, called neurons. These neurons concentrate in a particular area of the midbrain, called substantia nigra. Dopamine is a chemical which relays messages between the substantia nigra and other parts of the brain. Such kind of nervous communications is necessary to control the movement of human body. The fact which keeps smooth and coordinated muscles. When approximately 60% to 80% of the dopamine-producing cells are damaged, the motor symptoms of PD begin to appear [1] PD affects mainly the quality of life of its patients. In fact, problems with imbalance, tremor, postural instability and rigidity are all the major symptoms of PD.

In neurodegenerative disorders, such as, PD, an early stage diagnosis can result in significant lifesaving. For neurologists who diagnose PD, it is almost indispensable to make decisions before going through neuroimaging. However,

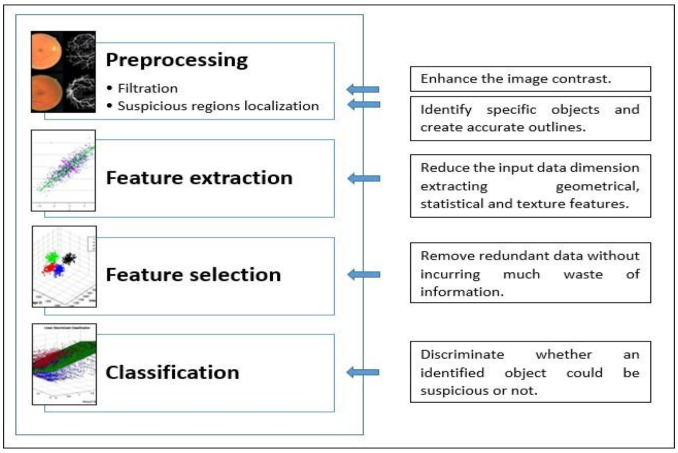

Figure 1. Flowchart of typical CAD system

clinical diagnosis of PD leads to serious errors as well as high medical costs. By contrast, it provides insufficient meaningful data about the target areas.

Since no specific reliable test to detect PD is available yet, physicians have to analyze and evaluate a great quantity of information in a limited time. Based on that, CAD technology contributes to assist them focus on the expected tumors as well as mark the suspect structures and tissues. Therefore, these systems can give accurate results in the diagnosis activities. In fact, by using CAD, the reduction of nerve cells in the brain can be identified in early stages [2]. Thus, distinguishing PD patients from healthy subjects has become pretty easier.

With the rapid advances in computing and digital imaging techniques, CAD concept is emerging as an advanced interdisciplinary technology. It combines fundamental elements of different areas, such as, digital image processing, artificial intelligence, machine learning, mathematical modeling and statistics. CAD systems are applied for a large set of imaging modalities, including, X-rays, Ultrasonography (US), Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and several nuclear tomographic techniques. Moreover, they are also used for all body parts, such as, the brain, chest, bones, lungs, blood-vessels, abdomen and limbs.

International Journal of Scientific & Engineering Research Volume 9, Issue 10, October-2018
ISSN 2229-5518

2 Historical Review of the Development of CAD

Early studies on quantitative analysis of medical images by computers [3]–[8] were reported in the 1950s. Because computers and machines were better than human beings at performing certain tasks, it was almost assumed at that time that computers could fully replace physicians in detecting abnormalities. Thus, since the 1960s, many different approaches to automated computer diagnosis have been attempted as aids in decision-making. Generally, these early attempts did not go as expected, because computers were not powerful enough.

In the 1980s, another approach emerged which assumed that computer output could be utilized by physicians, but not replace them. This concept is currently known as Computer Aided Diagnosis (CAD). It has been spread widely and quickly. However, in the early phase of research and development of CAD technology, some computer scientists criticized the new systems by saying that it simply would not work. In fact, that has been proved to be completely wrong. The reason for this strong criticism at that time might have been related to unsuccessful attempts in previous research efforts toward the development of automated computer diagnosis.

The year 1998 was one of the most important years in the history of CAD. It marked the transition of CAD concept from the research phase to industrial practice. However, in 2006, CAD technology succeeded in obtaining the Food and Drug Administration (FDA) approval [9].

Recently, CAD has become one of the major research subjects in medical imaging and diagnostic radiology. Many different types of CAD models have been developed for detection and characterization of various lesions in various medical imaging techniques.

3 Overview of Latest Developments in CAD

3.1 Systems Used in the Diagnosis of PD

Diagnosis of PD at early stages, by using CAD systems, is the main intension of all techniques and methods described below. In this section, different algorithms based on machine learning covering all CAD approches were mentioned. These processing algorithms provide more accuracy, sensitivity and specificity over the detection of PD.

3.2 Preprocessing

Image preprocessing is the technique that can be used to enhance image data prior to computational processing. Preprocessing an image involves: removing noise, normalizing image intensity, removing reflections and masking significant portions.

Normalization, filtering and masking are three common preprocessing techniques. Spatial normalization It is a kind of image registration. However, human brain may differ in size and shape. It can be seen in several views, such as, sagittal, coronal and hemispheric. The main objective of spatial normalization is that the location of the brain scan

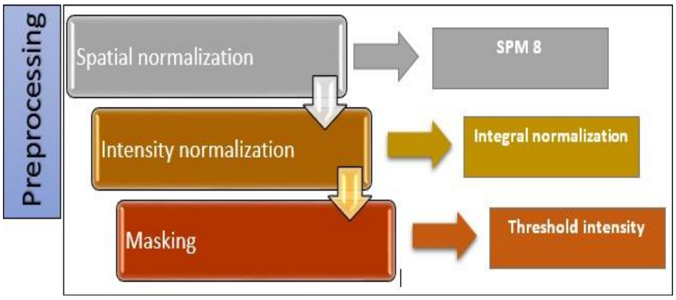

Figure 2. Preprocessing methodologies proposed by Martínez-Murcia et al

of a defined subject can be correlated to the same location as the brain scan of another subject. Intensity normalization is aimed to reduce the differences in intensity values between the voxels due to various reasons related to the acquisition defects. The resulting values can be used as input feature to identify PD patients.

Martínez-Murcia et al [10] have presented a specific normalization technique for SPECT database preprocessing. It is based on using spatial and intensity normalization. Firstly, All the images were spatially normalized using SPM8 software. Thereafter, In order to determine the difference in activities between the area of specific activity (related to dopamine transporters) and the area of non-specific activity (blood vessels), some kind of intensity normalization were required. The intensity normalization applied in this work is based on the obtaining of a fundamental parameter from the image I_p and the estimation of its activity as follows:

$$t^j = \frac{t}{I_p} \qquad (1)$$

Where t denotes the spatial normalized image, and t the^j intensity normalized image. The value I_p is determined using integral normalization. This method approximates the expression $I_p = t$ as the sum of all the intensity values of the image. Finally, to distinguish regions of interest from non-specific areas, a binary mask was applied. This kind of masks selects all voxels that are higher than a specific intensity threshold I_{th}. So that only selected voxels are considered for further processing. The threshold was established as the average of all intensity values in the whole image. It is computed as follows:

$$I_{th} = \frac{1}{2}\left(\max\left(I_{mean}\right) - \min\left(I_{mean}\right)\right) + \min\left(I_{mean}\right) \qquad (2)$$

Where I_{mean} indicates the mean of all images in the database. The proposed methodologies were tested on two databases: 208 DaTSCAN images from the "Virgen de la Victoria" Hospital (VV) in Málaga, Spain and 289 DaTSCAN images from the Parkinson Progression Markers Initiative (PPMI). High accuracy rates were achieved, 94.7% for VV database and 91.3% for PPMI database.

Brahim et al [11] have made comparison between different intensity normalization perform. Two techniques were proposed, Gaussian Mixture Model (GMM) image filtering and Mean-Squared Error (MSE) optimization. GMM is aimed

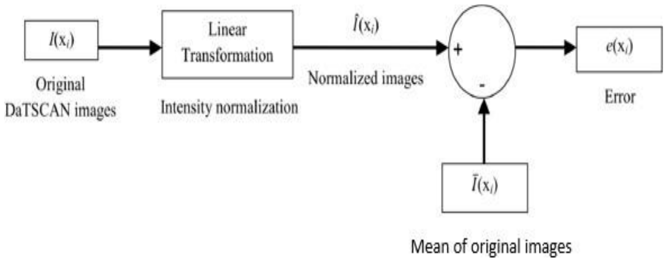

Figure 3. Flowchart of linear intensity normalization using MSE optimisation approach proposed by Brahim et al

to modulate the intensity in any position of the image, according to this equation:

$$I_{Gauss}(x) = I \times p(x) \quad (3)$$

Where I is the average intensity of all voxels and $p(x)$ is the probability distribution for a spatial coordinate x. The GMM image filtering method used in this work is based on a parametric model. In which, the probability density function of the gray levels of an image is modeled by a mixture of k Gaussians:

$$p(x) = \sum_{i=1}^{k} \varphi_i \, F(\mu_i \theta_i) \quad (4)$$

Where $\theta_{i=1\ldots k}$ denotes the parameter of distribution associated with component i. μ_i is the mean of component i. $F(\mu_i, \theta_i)$ is the density of the i-th Gaussian. $\varphi_{i=1\ldots k}$ is the mixture weights ($\varphi_{1\ldots k} = 1$).

In MSE optimization, an estimate $\hat{I}(x_i)$ was found. It minimizes the cost function ξ:

$$\xi = \frac{1}{N_{ns}} \sum_{i=1}^{N_{ns}} \left| \hat{I}(x_i) - I(x_i) \right|^2 \quad (5)$$

Where N_{ns} is the number of voxels in the non-specific region. $I(x_i)$, $\bar{I}(x_i)$ and $\hat{I}(x_i)$ denote the intensity values of the original, mean and normalized images in non-specific region. Normalizing the intensity using GMM image filtering and MSE optimization raised the classification rates to almost 92.91% of accuracy, 94.64% of sensitivity and 92.65% of specificity.

Kaya et al [12] have showed that discretization of continuous valued features increases the performance of classification. Discretizing a dataset involves the transformation of continuous valued features into discrete valued features. Firstly, continuous valued feature of 32 PD subjects' voice measures were sorted. Then, the candidate cut points are determined for this continuous valued feature. The fitness values of the candidate cut points are computed. Thereafter, the values of the continuous valued feature are splitted according to candidate cut point which has the best fitness value. These steps are used recursively until a stopping criterion is verified. The classification rates raised up to 95%.

Ma et al [13] have introduced a new hybrid method named SCFW-KELM. It combines Subtractive Clustering Features Weighting (SCFW) and Kernel-based Extreme Learning Machine (KELM). In the proposed method, SCFW is used as a data preprocessing technique. It aims at decreasing the variance in features of the PD dataset in order to strengthen the discrimination between normal and affected subjects. Experimental results have shown advantageous performs in term of classification: 99,49% of accuracy, 100% of sensitivity and 99,39% of specifity.

4 Feature Extraction

The feature extraction can be used to reduce the input data dimension and minimize the training time taken by the classifier. From the region of interest, we can extract several features, such as, geometrical, statistical and texture moments. Statistical moments include: mean value, median value and standard deviation. Geometrical moments include: shape, size and surface.

Ramírez et al [14] have proposed a new extraction technique based on Partial Least Squares (PLS). The purpose of this work is to reduce the large dimensionality of the input data by downscaling the SPECT images and extracting score features using PLS. PLS is used to find a linear regression model by projecting the predicted and observable variables into a new relevant space. The experimental results proved that PLS is effective for extracting discriminating information from ambiguous data. Yielding peaks of accuracy, sensitivity and specificity of 96.9%, 100%, 92.7%, respectively.

Brahim et al [11] have made comparison between two feature extraction methods: Voxels-as-Features (VAF) and Principal Component Analysis (PCA). VAF uses all voxels of the image as a feature vector. This vector will be used as an input to the classifier. PCA is used to obtain the most relevant information from row data. Besides, it presents this information in a lower dimensionality space. However, PCA was found to be more accurate than VAF. In fact, PCA achieved high accuracy rates for PD diagnosis with peak of 92.91% against 90.55%, achieved using VAF.

Martínez-Murcia et al [15] have used FastICA to extract pertinent data. FastICA it is an adaptive algorithm based on gradient descent. It can be problematic when algorithms are used on an environment in which adaptation is not necessary. The single unit FastICA algorithm has the following form:

$$w(k) = E\{x g(w(k-1)^T x)\} - E\{g'(w(k-1)^T x)\} w(k-1) \quad (6)$$

Where the loadings vector w is normalized to unit norm in each iteration and the function $g(x)$ is a derivative of the contrast function. The expected values are estimated in practice by using the mean of a significant number of samples of the input data. The achieved rates of accuracy were up to 94.7%.

5 Feature Selection

The transformation of data space into a feature space is known as feature selection. This technique is very effective when the data contains many features that are either redundant or irrelevant. Feature selection is used to remove these insignificant data without incurring much waste of information. Selection task should be distinguished from feature

International Journal of Scientific & Engineering Research Volume 9, Issue 10, October-2018
ISSN 2229-5518

extraction. While feature extraction creates new features from functions of the original features, feature selection returns a subset of these features.

Salas-Gonzalez et al have presented two different feature selection techniques. The first technique [16] is based on the selection of voxels of interest using Welch's t-Test, as following:

$$W_t = \frac{\bar{I}_{NOR} - \bar{I}_{PAT}}{\sqrt{\frac{I^{\sigma}_{NOR}}{N_{NOR}} + \frac{I^{\sigma}_{PAT}}{N_{PAT}}}} \qquad (7)$$

Where N_{NOR} and N_{PAT} are the numbers of normal subjects and affected subjects. $\bar{I}_{NOR}$ and $\bar{I}_{PAT}$ are the mean images of normal and affected subjects. $I^{\sigma}{}_{NOR}$ and I^{σ}_{PAT} are the root-mean-square deviation of normal subjects and affected subjects images. W_t denotes the resulting image with the value given by Welch's t-test calculated in each voxel. Those voxels which present a Welch's t-test value greater than a given threshold will be selected for the classification task. Those voxels have been used as a feature vector for two different classifiers: SVM with linear kernel and classification trees. Accuracy rates went up to 98.7%. The second technique [17] have shown the performance of Mann-Whitney-Wilcoxon U-test in comparison to Welch's t-test in selecting voxels of interest. The absolute value of the U-statistic of a two-sample unpaired Wilcoxon test was used to measure the voxel separability between classes. First, all the observations are arranged into a single ranked series. Besides, the ranks for the observations which came from samples 1 and 2 were added up. The statistic U is then given by:

$$U_i = R_i - \frac{n_i\,(n_i + 1)}{2} \quad , i = \{1,2\} \qquad (8)$$

The distance U given by the U-test is the smaller value of U_1 and U_2, $U = min\ \{U_1, U_2\}$. Those voxels which provide the greatest U-values are selected and used as input features. The proposed methodology entails an accuracy rate of greater than 90%.

Most PD patients (approximately 90%) are suffering from speech rigidity, such as, dysphonia and dysarthria. On this basis, Revett et al [18] have performed a new automated system based on vocal analysis as feature selection for PD subjects suffering from dysphonia. In this work, a relatively small dataset containing sustained vowel phonation data was examined. The feature set is analysed using the rough sets paradigm, which maps feature vectors associated with objects onto decision classes. The results from applying rough sets is a set of rules that map features via rules into a decision support system. The proposed methodologies increased the accuracy rate to almost 98%.

Bennasar et al [19] have developed two nonlinear feature selection methods, named Join Mutual Information Maximization (JMIM) and Normalized Join Mutual Information Maximization (NJMIM). Both these methods used mutual information and the maximum of the minimum criterion, which alleviates the problem of overestimation of the feature significance. The achieved results demonstrated that JMIM outperformed significantly the classification results.

1. Represents the sample standard deviation of the differences between predicted values and observed values.

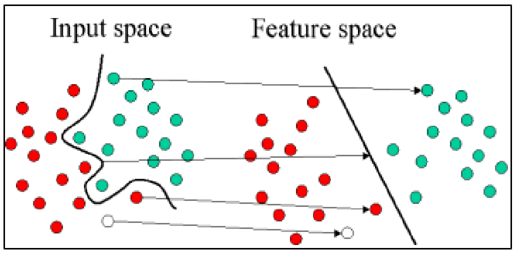

Figure 4. SVM divides two different classes, converting a non-linear separable function into another linear separable function as well as keeping the largest possible margin between them.

6 CLASSIFICATION

Classification is the last process in a CAD system. It is used to discriminate whether a subject is affected by a defined condition or not. The training and classification can be done using several techniques.

Support Vector Machine (SVM) is a set of related supervised learning methods. SVM with linear kernels defines decision hypersurfaces or hyperplanes in a multidimensional feature space, that is:

$$g(x) = W^T X + \omega_0 = 0 \qquad (9)$$

Where W is known as the weight vector and ω_0 as thethreshold. The weight vector W is orthogonal to the decision hyperplane. The optimization task consists of finding the unknown parameters ω_i, $i = \{1, ..., n\}$ by defining the decision hyperplane. X_i, $i = \{1, ..., n\}$ are the feature vectors of the training set X.

SVM classifiers have been widely used in the classification issues due to their high accuracy and ability to deal with high dimensional data [20].

Salas-Gonzalez et al [21] have presented a new combined technique to increase the accuracy of cerebral disorders diagnosis. The presented methodology is based on the combination of SVM with linear kernels and classification trees. Firstly, the brain image is divided into several parts or components (each one denoted by C). These components will be used later as feature vectors for the SVM classification. The proposed methodology computes a systematic scan of the whole brain image. For the analysis, M is the number of patients, N is the number of components whom the brain is divided into, and C_{ij} is the component i of the patient j where $i = \{1,2, ..., N\}$, $j = \{1,2, ..., M\}$. The whole brain image I of the patient j is:

$$I_j = C_{1j} \cup C_{2j} \cup \ldots \cup C_{2j} \qquad (10)$$

Each image I_j has an associate label y_j, which is chosen to be "0" for normal subjects and "1" for affected subjects. Each component is used as a feature vector input, therefore an SVM classifier with linear kernel is trained and tested on each component C_{ij}. After processing the whole database, the outcome is a set of individual SVM results which will be used to construct a tree classifier. The training sample consists of data $\{C_{ij}, y_j\}$ This classification tree will al- low to choose which components are appropriate to use.

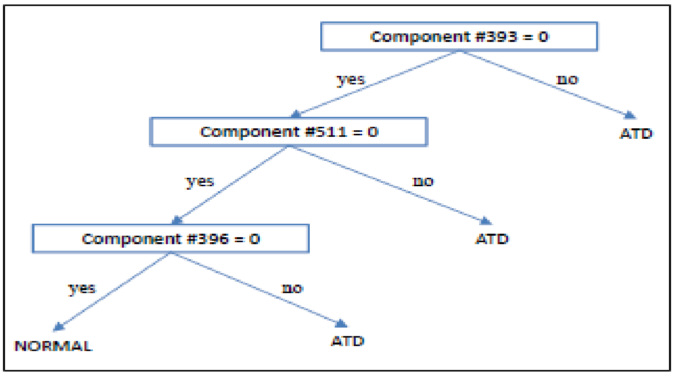

Figure 5. Classification tree proposed by Salas-Gonzalez et al

in the classification task. The classification trees increases considerably the classification accuracy. Improving from 84 % of accuracy using SVM to 90 % combining SVM and classification trees.

Artificial Neural Network (ANN) is an interconnected group of nodes. In machine learning and cognitive sciences, ANNs are a family of models inspired by biological neural networks. They are often similar to the vast network of neurons in the brain, where each circular node represents an artificial neuron. Whereas, each arrow represents a connection from the output of one neuron to the input of another. ANNs are mainly used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown. Due to its high performance in term of classification and recognition of suspect patterns, ANN has been widely used in developping of new CAD solutions. [22].

Mehmet Can [23] has designed an eleven-parallel networks system which is combined with a majority voting system. The proposed system aims to raise true recognition rates in an imbalanced dataset. The studied dataset is based on a voice recording performed at the university of Oxford by M.A.Little [24]. It is demonstrated that a true positive rate of almost 90% of positive class was achieved. Hadjahmadi et al [25] have made a comparison to determine the performance of several types of classifiers. Bayesian networks, Regression, Classification and Regression Trees (CART), SVM and ANN are all included to propose a decision support system for PD diagnosis.. By comparing all these classifiers, CART was found to be the most efficient with an accuracy rate up to 93.75%. ANN and SVM were the next powerful classifiers with approximately 89% and 87% of accuracy.

kaya et al [12] have demonstrated that combining the discretization of continuous valued features with traditional classifiers such as ANN and SVM, can increase the diagnosis results. Abraham et al [26] have pointed out the effect of combining discretization as a preprocessing tool with Naive Bayes classifiers. Accuracy rate of approximately 97 was achieved. Demsar et al [27] have created a predictive model for cerebral disorders diagnosis. It is based on appling discretization of continous features bound together with two kinds of classifiers: classification tree and Naive Bayes. Over 82% of accuracy was performed for classification tree and 80% for Naive Bayes.

Neuro fuzzy technology can be very efficient for classification [28]. Fuzzy systems can handle uncertainties associated with information or data in the knowledge bases. They have been widely applied in solving different real world problems. Fuzzy systems use data and knowledge specific to chaotic dynamics of the process and increases the performance of the system. Various neural and fuzzy structures have been proposed to solve diverse problems [29]–[34]. Rahib H. Abiyev et al [35] have implemented a new computerized model which integrates fuzzy systems and neural networks. It exploits the performance of both ANN and fuzzy logic.

Cluster analysis is an iterative process. It involves processing data and modeling parameters until getting desired properties. Sajid Ullah Khan [36] has applied three clustering techniques, including: k-nearest neighbors (KNN), Random Forest and Ada-Boost. KNN classifies an instance based upon K number of closest neighbors. Random Forest works by generating several decision trees and combining their result for classification. Ada-Boost uses many weak learning algorithms to produce the results of strong learning algorithms.

7 Conclusion

The brain contributes to the balance of the whole body. PD is a chronic disorder of central nervous system. Dopamine is a chemical that relays messages between the midbrain and other parts of the brain. PD involves the reduction of dopamine in nerve terminals. The fact which produces rigid movement, slight tremor and postural instability. Although PD is in continuous progress and the number of people suffering from the disease is expected to rise, no particular test to detect the disorder is avialable. Usally, functional cerebral diagnosis yields a large volume of information. These data are subjected to advanced visual analysis made by experienced clinicians. The procedures involve often a predefined classification or the analysis of regions of interest. Generally, these procedures prone to several errors. Lately, some methods based on the machine learning paradigm and neural networks have been applied to analyse cerebral scans. These methods have led to the emergence of high-performance CAD systems for the detection of PD. These systems have been applied to various imaging techniques, in order to extract complex high-dimensional features. These features will be subsequently used to train an automatic classifier. Extracting features and selecting features will considerably reduce the number of wrong decisions. Classification is a set of algorithms that are used to discriminate between normal subjects and PD patients.To improve the overall performance of a CAD model, development in each processing step may be required.

References

[1] M. M. Goldenberg, “Medical management of parkinsons disease,” *Pharmacy and Therapeutics*, vol. 33, no. 10, pp. 590–594, October 2008.

[2] A. Valli and G. Wiselin Jiji, “Parkinsons disease diagnosis using image processing techniques a survey,” *International Journal on Computational Sciences & Applications (IJCSA)*, vol. 4, no. 6, December 2014.

[3] G. Lodwick, C. Haun, and W. Smith, "Computer diagnosis of primary bone tumor," *Radiology*, vol. 80, pp. 273–275, 1963.

[4] P. Myers, C. Nice, and H. Becker, "Automated computer analysis of radiographic images," *Radiology*, vol. 83, pp. 1029–1033, 1964.

[5] F. Winsbarg, M. Elkin, and J. May, "Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis," *Radiology*, vol. 89, pp. 211–215, 1967.

[6] R. Kruger, J. Towns, and D. Hall, "Automated radiographic diagnosis via feature extraction and classification of cardiac size and shape descriptors," *IEEE Transactions on Biomedical Engineering*, no. 3, pp. 174–186, May 1972.

[7] R. Kruger, W. Thompson, and A. Turner, "Computer diagnosis of pneumoconiosis. IEEE Transactions on Systems, Man, and Cybernetics," no. 1, pp. 44–47, January 1974.

[8] J. Toriwaki, Y. Suenaga, and T. Negoro, "Pattern recognition of chest x-ray images," *Computer Graphics and Image Processing*, vol. 2, pp. 252–271, 1973.

[9] F. Hiroshi and U. Yoshikazu, "Computer-aided diagnosis: The emerging of three CAD systems induced by Japanese health care needs," vol. 92, pp. 238–248, April 2008.

[10] F. J. Martínez-Murcia, J. M. Górriz, J. Ramírez, I. A. Illán, and A. Ortiz, "Automatic detection of Parkinson using significance measures and component analysis in DaTSCAN imaging," *Neurocomputing*, vol. 126, pp. 58–70, February 2014.

[11] A. Brahim, J. Ramirez, J. M. Gorriz, L. Khedher, and D. Salas-Gonzalez, "Comparison between Different Intensity Normalization Methods in 123I-Ioflupane Imaging for the Automatic Detection of Parkinsonism," *PLOS ONE*, June 2015.

[12] K. Ersin, F. Oguz, B. Ismail, and A. Ahmet, "Effect of discretization method on the diagnosis of Parkinsons disease," *January*, 2011.

[13] M. Chao, O. Jihong, C. Hui-ling, and Z. Xue-Hua, "An efficient diagnosis system for parkinsons disease using kernel-based extreme learning machine with subtractive clustering features weighting approach," November 2014.

[14] J. Ramirez, J. Gorriz, and F. Segovia, "Computer aided diagnosis system for the Alzheimers disease based on partial least squares and random forest SPECT image classification," *Neuroscience Letters*, vol. 472, no. 2, pp. 99–103, February 2010.

[15] F. J. Martínez-Murcia, J. M. Górriz, J. Ramírez, I. A. Illán, and A. Ortiz, "Automatic detection of Parkinson using significance measures and component analysis in DaTSCAN imaging," *Neurocomputing*, vol. 126, pp. 58–70, February 2014.

[16] D. Salas-Gonzalez, J. Górriz, J. Ramírez, M. López, I. Álvarez, F. Segovia, and C. G. Puntonet, "Selecting Regions of Interest for the Diagnosis of Alzheimers Disease in Brain SPECT Images Using Welchs t-Test," *Bio-Inspired Systems: Computational and Ambient Intelligence*, vol. 5517, pp. 965–972, 2009.

[17] D. Salas-Gonzalez, J. Górriz, J. Ramírez, F. Segovia, R. Chaves, M. López, I. A. Illán, and P. Padilla, "Selecting Regions of Interest in SPECT Images Using Wilcoxon Test for the Diagnosis of Alzheimers Disease," *Hybrid Artificial Intelligence Systems*, vol. 6076, pp. 446–451, 2010.

[18] R. Kenneth, g. Florin, and s. abdel-badeeh Mohamed, "Feature selection in Parkinson's disease : A rough sets approach," *Proceedings of the international multiconference on computer science and information technology*, pp. 425–428, October 2009.

[19] B. Mohamed, h. yulia, and s. Rossitza, "Feature selection using joint mutual information maximization," 2015.

[20] A. Ben-Hur, C. S. Ong, S. Sonnenburg, and B. Schölkopf, "Support Vector Machines and Kernels for Computational Biology," *PLoS Computional Biology*, vol. 4, no. 10, October 2008.

[21] D. Salas-Gonzalez, J. M. Gorriz, and J. Ramirez, "Computer Aided Diagnosis of Alzheimers Disease Using Support Vector Machines and Classification Trees," *Advances in Neuro-Information Processing*, vol. 5507, pp. 418–425, 2009.

[22] F. ASTROM and R. KOKER, "A parallel neural network approach to prediction of PD," *Expert systems with applications*, vol. 38, pp. 12 470–12 474, 2011.

[23] M. can, "Neural networks to diagnose the Parkinson's disease," 2013.

[24] J. McDonough, K. Kumatani, T. Gehrig, E. Stoimenov, U. Mayer, S. Schacht, M. Wolfel, and D. Klalow, "Machine learning for multimedia interaction," *Springer-Verlag Berlin Heidenberg*, vol. 4892/2008, 2007.

[25] A. Hadjahmadi and T. J. Askari, "A decision support system for PD Diagnosis using classification and regression tree," *the journal of mathematics and computer science*, vol. 4, no. 2, pp. 257–263, 2012.

[26] R. Abraham, S. J, and S. Iyengar, "A comparative analysis of discretization methods for medical data mining with NaiveBayesion classifier," *the 9th international conference on Information technology*, 2006.

[27] J. Demsar, B. Zupan, M. N.Aoki, J. Wall, G. T.H., and J. Beck, "Feature mining and predictive model construction from trauma patient data," *International journal of medical informatics*, 2011.

[28] J. Blahuta and T. souk up, "Ultrasound medical image recognition with artificial intelligence for parkinson's disease," *MIPRO, 2012 Proceedings of the 35th International Convention*, pp. 958–962, May 2012.

[29] R. H. Abiyev and O. Kaynak, "Fuzzy wavelet neural networks for identification and control of dynamic plants: a novel structure and a comparative study," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 8, pp. 3133–3140, 2008.

[30] R. H. Abiyev, "Fuzzy wavelet neural network based on fuzzy clustering and gradient techniques for time series prediction," *Neural Computing & Applications*, vol. 20, no. 2, pp. 249–259, 2011.

[31] R. Abiyev, O. Kaynak, T. Alshanableh, and F. Mamedov, "A type-2 neuro-fuzzy system based on clustering and gradient techniques applied to system identification and channel equalization," *Applied Soft Computing*, vol. 11, no. 1, 2011.

[32] R. H. Abiyev, "Credit rating using type-2 fuzzy neural networks," *Mathematical Problems in Engineering*, vol. 2014, p. 8, 2014.

[33] R. Abiyev, R. Aliev, O. Kaynak, I. Turksen, and K. Bonfig, "Fusion of computational intelligence techniques and their practical applications," *Computational Intelligence and Neuroscience*, vol. 2015, p. 3, 2015.

[34] H. Do and J. Chen, "A neuro-fuzzy approach in the classification of students academic performance," *Computational Intelligence and Neuroscience*, vol. 2013, p. 7, 2013.

[35] H. Abiyev and S. Abizade, "Diagnosing Parkinson's Diseases Using Fuzzy Neural System," *Computational and Mathematical Methods in Medicine*, vol. 2016, p. 9, December 2015.

[36] S. U. Khan, "Classification of PD using data mining techniques," July 2015.