Turkish Journal of Computer and Mathematics Education Vol.12 No.10 (2021), 5257-5276

Research Article

A novel supervised machine learning algorithm to detect Parkinson's disease on its early stages

Lavanya Madhuri Bollipo[1], Dr.Kadambari. KV[2]

[1]Research Scholar, Department of Computer Science and Engineering, National Institute of Technology Warangal, India.
[2]Assistant Professor, Department of Computer Science and Engineering, National Institute of Technology Warangal, India.
[1]lavanyabollipo@student.nitw.ac.in

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract: Early and accurate Parkinson's disease (PD) diagnosis are usually complex as clinical symptoms often onset only when there is extensive loss of dopaminergic neurons in substantia-nigra and symptoms are atypical at an early stagesof the disease. Recent brain imaging modality such as single photon emission computed tomography (SPECT) with ^{123}I- Ioflupane (DaTSCAN) have shown to be a better diagnostic tool for PD even in its initial stages. Presently machine learning algorithms have become trendier and play important role to automate PD diagnosis and predict its progression. In machine learning community, support vector regression (SVR) has recently received much attention due to its ability to negotiate between fitting accuracy and model complexity in training prediction models. This workpresents an optimized SVR with weights associated to each of the sample data to automate PD diagnosis and predict its progression at primary stages. The proposed algorithm (W-SVR) is trained with motor and cognitive symptom scores in addition to striatal binding ratio (SBR) values calculated from the ^{123}I-Ioflupane SPECT scans (taken from the Parkinson's progression markers initiative (PPMI) database) for early PD prognosis accurately. In model building, different kernels are used to check the accuracy and goodness of fit. We observed promising results obtained by W-SVR in comparison with classic Support vector regression.

Keywords*:* Parkinson's disease, Support Vector Regression, Supervised machine learning, Prediction and lassification.

1. Introduction

Parkinson's disease (PD) is a brain disorder that eventually affects motor and cognitive behavior. The main neuro-pathological characteristic of PD is gradual loss of dopaminergic (DA) neurons in the substantia-nigra and basal ganglia, which includes the caudateand putamen nucleus [1]. As a result,there is a decrease of dopamine content in the striatum, and a corresponding dissipation of dopamine transporters (DAT). These DATs are responsible for controlling functions like movement, cognition, mood and reward [2]. Thus, the loss will lead to deterioration in nervous system which gives riseto: motor disturbances that include slowness of movement, resting tremor, muscular rigidity and impaired co-ordination [3]; and cognitive illness like depression, olfactory, and sleep disturbances [4]. As the disease continues to progress, one can observe significant change in motor and cognitive behavior worsening by the day.

Studies verify that there is currently no way to repair neurons once they have been destroyed and clinical symptoms in PD arise only when there is more than 60% loss of dopaminergic (DA) neurons [5]. However, by the time of symptoms of PD are detectable clinically, critical nigrostriatal dopaminergic neurons would have already been damaged. Conversely, the disease continues to progress with accumulation of significant motor and cognitive disability, worsening quality of life, reduced productivity, nursing home placement, and increased mortality [6]. There is still no standard cure for PD because the cause for death of DA cell is still mysterious. The main challenge of clinical diagnosis ofPD is to properly identify the PD subjects at an early stage when symptoms are atypical. However, early signs and symptoms of this disease may go unnoticed as they can overlap with other disease's symptoms. Thus, timely and clear-cut diagnosis of PD is important for initiating neuro-protective therapies. These treatments can assist the PD subjects to recover and retain their quality of life without further deterioration.

To detect PD further in its early stages, researchers have resorted to neuroimaging techniques. Applications of brain imaging techniques in the process of diagnosis of neurological disorders have increased the accuracy rate for predicting the disease at an early stage. The introduction of brain imaging modalities such as Single photon emission computed tomography (SPECT) with ^{123}I-Ioflupane (DaTSCAN), a pre-synaptic radio- pharmaceutical of the dopaminergic transporters (DAT) showing a

substantial uptake decrease in basal ganglia of PD subjects. These DaTSCANs have revealed that there is a significant depletion of DAT in PD subjects even in their early stages [7–10]. Therefore, DaTSCANs are suggested to be a suitable method for the clinician to increase the diagnostic accuracy for predicting the disease even in its initial stages [11, 12]. Conventionally, the DaTSCAN images obtained from suspected PD individuals are exposed to a visual analysis performed by clinical experts. A predefined ratinggiven according to Tolosa et al. (2007) [13] or the analysis of regions of interest (ROIs) attributed by Lozano et al.(2010) [14] typically involves in this process. This procedure is independent and can be susceptible to error, sinceit relies on overall changes in dopamine concentration throughout the ROI.

Recently, machine learning paradigms are used to automate the prediction of neurological disease progression and assess the stage of pathology, yieldingto the construction of computer-aided-diagnosis (CAD) systems. These systems are applied to semi quantitative parameters to train an automatic classifier like support vector machines (SVM) indistinguishing PD subjects. Hence, some researchers have applied machine learning techniques to estimate clinical scores from brain SPECT image [15–20] and found reliable correlation between estimated clinical scores and different PD progression stages. Therefore, appropriate and targeted treatment can then be carried out to treatPD effectively.

SVM have become popular in machine learning community. An important property of SVMs is, the determination of model parameters is a convex optimization problem so the solution is always global optimum and has emerged as an important learning technique for solving classification problems in various fields with excellent performance [24-28]. With the introduction of ε-insensitive loss function (ε : error deviation), SVM has been extended to solve the regressionproblems called support vector regression (SVR) [25]. SVR has recently received much attention due to its competitive performance compared to other regression techniques such as logistic regression, and Neural Networks (NN). In general, SVR constructs decision functions in high dimensional space for linear regression while the training data is mapped to a higher dimension in kernel Hilbert space.ε-SVR is the first popular SVR strategy [26]. ε-SVR aims to find a function whose deviation is not more thanε, thus forming the ε-tube, to fit all training data. In order to find the best fitting surface, ε-SVR tries to maximize the minimum margin containing data points in the ε-tube as much as possible.

In PD diagnostic system, the data is mostly imbalanced. In such scenario, a classic implementation of SVR is inefficient as they may provide inaccurate results. This paper proposes a variant of SVR, optimized with weights associated to each of the sample dataset, which results in anew approach called as weighted support vector regression (W-SVR). This methodgives more accuracy than the classical SVR, and the resulting support vectors are sparser and much more robust with respect to changes in the regularization hyper-parameter, while retaining a comparable accuracy.A comparison of proposed algorithm (W-SVR) with standard SVR using kernels like linear, polynomial, sigmoid, radial basis function (RBF) and logistic over PD dataset indicates that W-SVR fits better and shows comparable accuracy for diagnosing Parkinson's data in less computational time. this work also gives the predictive model for PD dataset using multivariate logistic regression (MLR) to show the class probabilities. Thus, the proposed model can be used as a better tool for early detection of Parkinson's disease.

Main points of the present work:

- Input data contains 634 subjects with 12 features obtained from PPMI database (http://www.ppmi-info.org/data) [29]. As PPMI is a multi-center international study involves subjects from different geographical locations, it addsdiversity in the database that makes the model robust.
- We use 12 baseline features to train our algorithm. Due to the evident of substantial decrease of DATin striatal regions of PD subjects [15-20], we choose four striatal binding ratio (SBR) values of left and right caudate, putamen nucleus as features. The age and gender also influence the classification and prediction accuracy, we use these two features.
- Motor impairment and cognitive are clinical symptoms of PD [16, 17], we considered MDS-UPDRS, MoCA, Tremor dominant (TD), postural instability/gait difficulty (PIGD) scores as features. In addition to these,we also use Handedness, Family history as features to get better results.
- We use weighted SVR and MLR models to train the algorithm over PD dataset. A performance comparison of the proposed models with the classic SVR over a Parkinson's data is given. Our experimental results shows that proposed method providesbetter classification, predictive results than SVR and can be used as better diagnostic tool in medical system.

The rest of the paper is organized as follows. Section 2 contains related research Section 3 contains the description and analysis carried out on input data, Section 4 contains problem formulation and mathematical modeling, statistical analysis of features, building of classification and prediction/prognostic model and

presents proposed W-SVR forecasting method. Experimental results are discussed and compared with existing models in Section 5. Finally conclusion is given in Section 6.

2. Related work

A closely related works of PD diagnosis are: Palumboet al. (2014) [18] investigated the diagnostic performance of ^{123}I-FP-CIT brain SPECT with semi-quantitative data by Basal Ganglia V2 software. The authors trained SVM by giving different set of descriptors suchas SBR values and patient's age for classification of PD and also evaluated the influence of age on disease onset. However, these researchers have used few features which may lead to loss of generality. Prashanthet al. (2014) [19,20] have appliedSVMs and multivariate logistic regression technique for classification and prediction of PD by using four striatal binding ratio (SBR) values (left and right caudate, left and right putamen) that were computed from DaTSCAN SPECT images as features to automate PD diagnosis and also predicted risk probability using multivariate logistic regression. Augimeri et al. (2016) [21] proposed a fully automated method for DaTSCAN analysis that generates quantitative measures based on striatal intensity, shape, symmetry and reached 100% classification accuracy with SVM. They also demonstrated the existence of a linear relationship and an exponential trend between pooled structural and functional striatal characteristics and the UPDRS motor score. Lei et al. (2017) [22] used multi-modal neuroimaging data for joint PD detection and clinical score prediction to design unique objective function and to capture discriminative features in training SVR. Oliveira et al. (2018) [23] assessed the potential of a set of features related to uptake ratios on the striatum, the estimated volume and length of the striatal region with normal uptake extracted from ^{123}I-FP-CIT SPECT brain imagesto diagnose Parkinson's disease, they obtained accuracyof 97.9% using SVM.

Although the authors have achieved high accuracy for classification but these procedural approaches require an effective algorithms for feature reductionand such techniques may also lead to lose of information which effects the decision making process. Consequently, these researchers have used standard support vector theory for automating classification of PD by considering large dataset. When standard SVR technique is used to deal with imbalanced medical data, it may decrease the generalization ability. This work proposes an optimized SVR technique with weights associated to each of the sample dataset, which results in a new approach called as weighted support vector regression (W-SVR). This methodgives better regression curve and good fit to data, and the resulting support vectors are sparser and much more robust with respect to changes in the regularization hyper-parameter, while retaining a comparable accuracy. A comparison of proposed algorithm (W-SVR) with standard SVR using kernels like linear, polynomial, sigmoid, radial basis function (RBF) and logistic over PD dataset indicates that W-SVR fits better and shows comparable accuracy for diagnosing Parkinson's data in less computational time. Thus, the proposed model can be used as a better tool for early detection of Parkinson's disease.

3. Materials

Work flow of the proposed prognostic model using Parkinson's data is shown in Figure 1. Statistical significance test was performed on all the features before going ahead with classification and prediction of PD.

3.1. Study participants

Data used in this work was taken from the Parkinson's progression markers initiative (PPMI) database (http:// www.ppmi-info.org/data). PPMI is a longitudinal study where subjects are evaluated longitudinally, i.e., evaluations occur at screening in 3 month intervals during the first year of participation and repeated after 6 months.

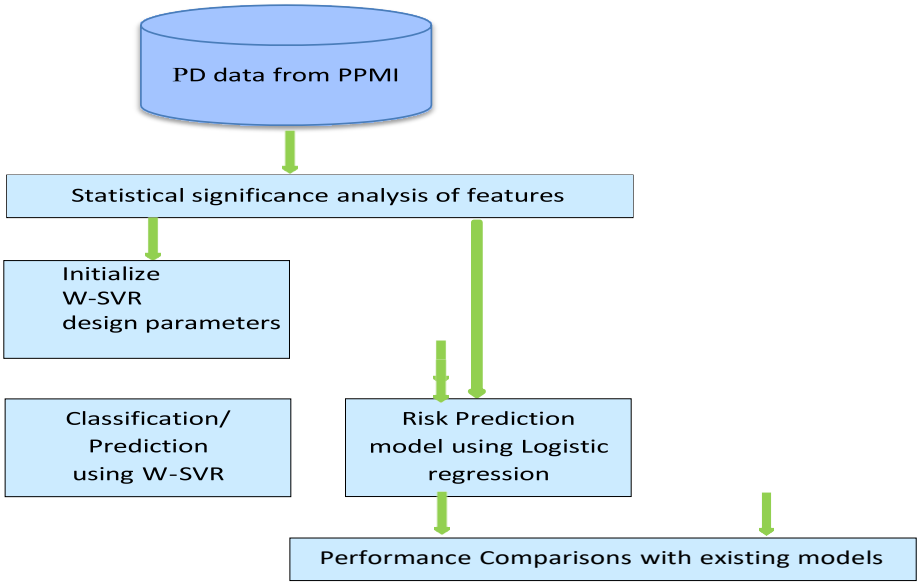

Figure 1: Scheme of the procedural work-flow

Data was downloaded on 24th Jan 2021. Total number of samples considered in the present work is *n* =634 subjects out of which 213 healthy and 421 were early PD subjects. Table 1 gives the details of the dataset.

Table 1: Database of Healthy and early PD population: Number of samples of Female and Male subjects in both the population (Gender), family history of PD (F.history) and handedness of subjects (Handedness).

Case(n=634)	Gender		F.history		Handedness	
	Female	Male	Yes	No	Left	Right
Healthy	75	138	103	203	40	173
Early PD	146	275	318		46	375

3.2. Feature selection

For the proposed work, 12 features were used: age, gender, handedness, family history, Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) total Pre-Dose, Montreal Cognitive Assessment (MoCA), tremor dominant (TD) score, postural instability/gait difficulty (PIGD) score, striatal binding ratio(SBR) values of the four striatal regions (left and right caudate, left and right putamen) that were computed fromDaTSCAN SPECT images [29]. Table 2 describes 9 predictors with respective mean and standard deviations. In the early PD group, all subjects were in their initial stage of the disease. All subjects were in Hoehn and Yahr (HY) stage 1 and 2 with mean ± standard deviation of the HY as 1.57 ± 0.48.

Table 2: Database of Healthy and early PD population: mean ± standard deviation(s) of Age, MDS-UPDRS, MoCA, TD score, PIGD score, SBR values for left caudate(Lt.Cad), right caudate(Rt.Cad), left putamen(Lt.Put) and right putamen(Rt.Put).

Case(n : 34)	*Age*	*MDS-UPDRS*	*MoCA*	*TD GD*	*PI*	*Lt.Cad.*	*Rt. Cad.*	*Lt. Put.*	*Rt. Put.*
Healthy	60.6±11.2	4.6±4.4	28.0 ±1.3	0±0	0±0	2.96 ± 0.6	2.93 ±0.5	2.12 ±0.5	2.12 ±0.5
Early PD	61.5±9.6	31.9±13.0	27.1 ±2.3	0.47 ±0.4	0.25 ±0.2	1.98 ±0.5	1.98 ±0.5	0.80 ±0.3	0.84 ±0.3

3.3. Statistical analysis of features

Linear regression analysis is used for obtaining statistically significant features (predictors). All statistical analysis was carried out using SPSS 25 software (SPSS Inc., Chicago, IL). $\rho < 0.05$ was considered to be the threshold value. When dealing with a large number of features in multivariate analysis, it is important to find those featuresthat are independent and overlap. i.e., groups of thee may be dependent in regression. Figure 2 shows the Glyph plot that depicts the multi-variability of randomly chosen 50 samplesfrom PD dataset. It is detected from Figure 2 that the multi-variability is high among the samples and hence leading to difficulty in classification of the problem.

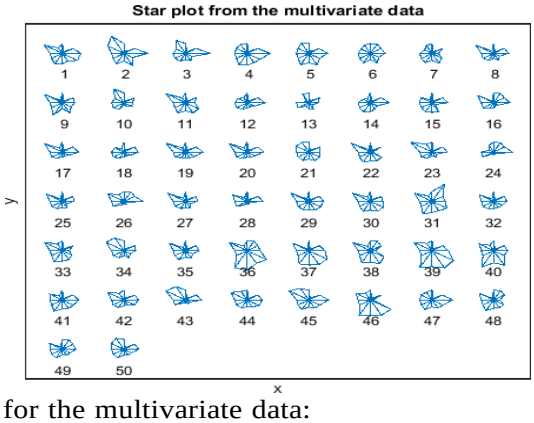

Figure 2: Glyph plot for the multivariate data:

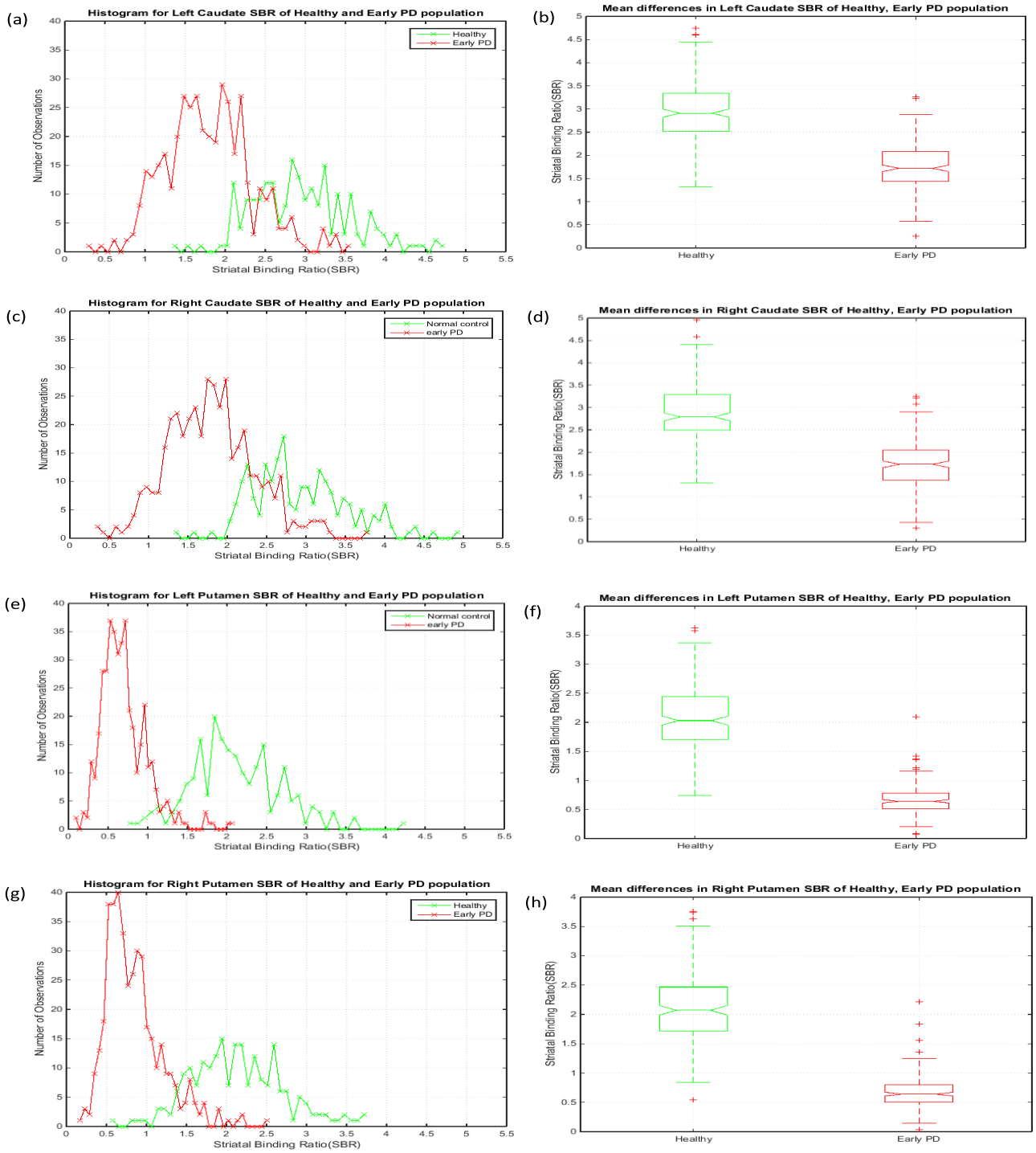

Figure 3: Histogram and notched box plots of the striatal binding ratio (SBR) values in Healthy and PD subjects. Histogram plots show the amount of overlap of SBR values and box plots show the mean differences of SBRs in Healthy and PD subjects. Figure (a,b,c,d) show distribution of left, right caudate SBRs and corresponding mean difference; Figure (e,f,g,h) show distribution of left, right putamen SBRs and corresponding mean difference. In each notched box plot, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, the whiskers extend to the most extreme data points that are not considered outliers, and outliers are plotted individually.

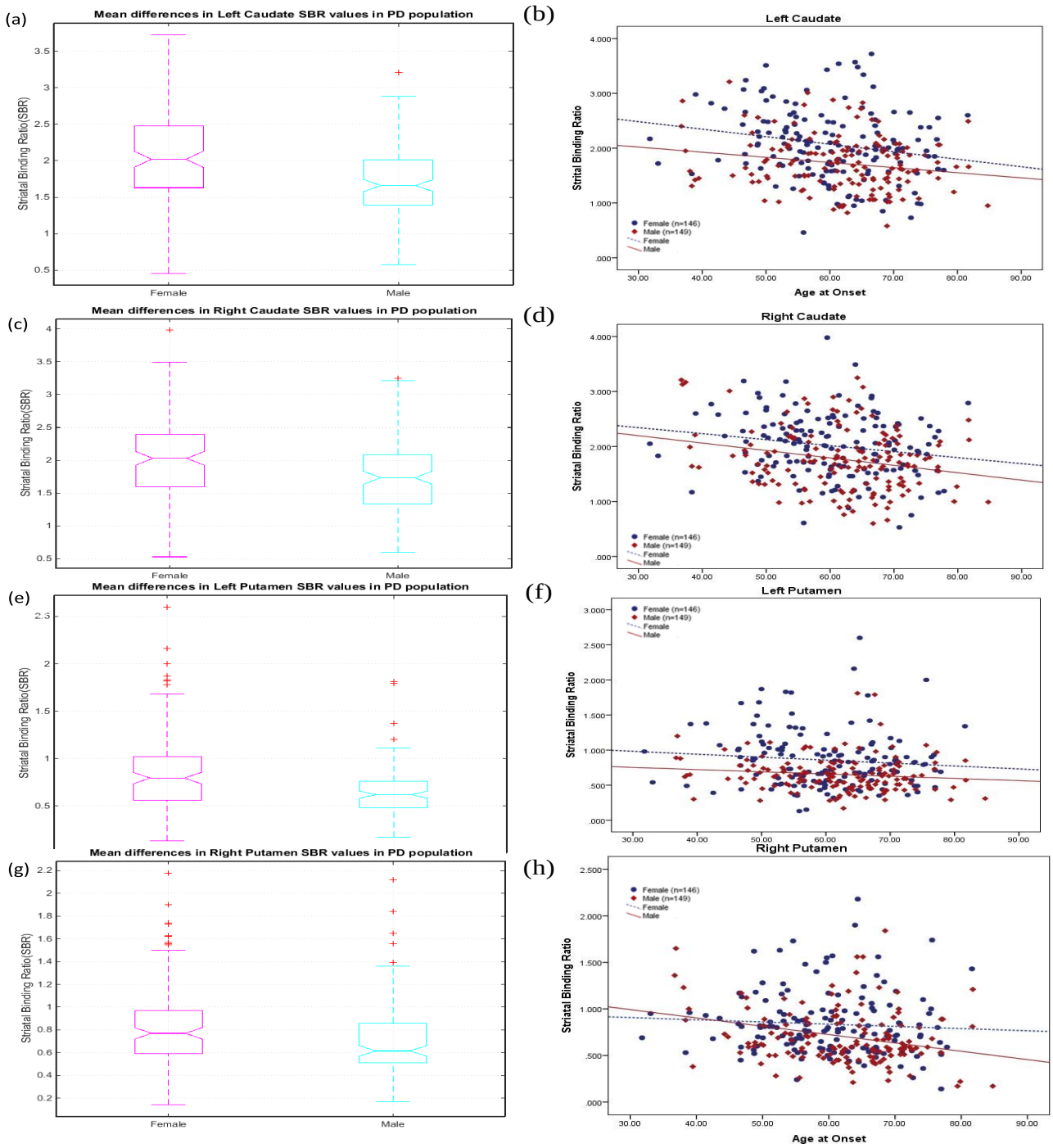

Figure 4: Notched box plots and regression graphs of the striatal binding ratio (SBR) values of female and male PD population. Regression plots show the change in SBR values with respect to age. Figure (a,b) shows mean difference and regression lines of left caudate SBRs; Figure (c,d) shows mean difference and regression lines of right caudate SBRs; Figure (e,f) shows mean difference and regression lines of left putamen SBRs; Figure (g,h) shows mean difference and regression lines of right putamen SBRs.

The histograms and notched box plots are plotted for each SBR feature to visualize its distribution for healthy and early PD population with its gender category. Figure 3 shows severe reductions in the dopamine concentration of striatum in PD patients, with greater reduction in putamen than the caudate compared to healthy subjects. The notched box plots fig. 3: (b,d,f,h) show that the notches for early PD and healthy subjects are fairly separated indicating the significance of these features. The histogram plots fig. 3: (a,c,e,g) show that amount of overlap of distribution between the healthy and early PD population. Overlapping is comparatively higher for the caudate SBRs (both left and right) when compared to the putamenal SBRs (both left and right). In fig. 3: (e,f), the amount of overlapping between Healthy and PD subjects is less for the left putamenal SBR values, indicating that this predictor has high discriminant power in classification problem. Figure 4 depicts mean and regression lines of SBR values of Female and Male PD population. It is shown that female subjects have lower SBR values than male population in all striatal regions with respect to age indicating the significance of these predictors.

From above figures it is witnessed that, multi-variability among samples are more (Figure 2) and the amount of overlap for SBR features are relatively high between healthy and early PD subjects (Figure 3, Figure 4). The amount of multi-variability and the overlap of the distributions of SBR values determine the difficulty of the classification or prediction problem, i.e., classifying or predicting early PD from healthy subjects. Higher the overlap/multi-variability, the more difficult it is the classification. Hence, we resort to machine learning tools to make this complex problem simpler. The research has shown that SVMs is widely used in automating PD diagnosis [18, 19, 21–23, 28]. Though SVMs give better accuracy than when compared to its counterparts, but when standard SVM technique is used to deal with regression estimation problems especially in medical data where samples are unbalanced, the performance may decrease. Therefore, there is a need for an effective classification/predictive technique which can improve the accuracy and reduce the computation time. Thus, in the present work, we propose a modified form of SVR which overcomes the above shortfalls.

4. Methodology

4.1. Support vector regression theory

Let $T = \{f(x_1, y_1), (x_2, y_2), , , , (x_n, y_n)\}$ be a training set of n samples, where $x_i \in \mathrm{R}^m$ are the input values and $y_i \in \mathrm{R}$ are the corresponding target values. Support vector regression, which evolved from the support vector classification by introduction of the ε-insensitive loss function, is a data-driven machine learning methodology for regression tasks. For linear regression, the objective function is

$$f(x) = w \cdot \varphi(x) + b \tag{1}$$

where $w \in \mathrm{R}^m$ and φ is the mapping function induced by a kernel K, i.e., $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$, which projects the data to a higher dimensional space. The function $f(x)$ should loosely fit the training data and be as flat as possible to avoid over-fitting problem by minimizing the $\|w\|$. To cope with these infeasible constraints, two slack variables ξ_i and ξ_i^* are introduced to measure the amount of difference between the estimated value and the target value. That is, ξ_i approximates the number of misclassified samples. This convex optimization problem is feasible based on the assumption that a function exists which approximates every data pair (x_i, y_i) with acceptable ε accuracy. Then, the objective function $f(x)$ is represented by the following constrained minimization problem:

$$\min_{w,\xi,\xi^*} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \tag{2}$$

s.t
$$y_i - w \cdot \phi(x_i) - b \leq \varepsilon + \xi_i$$
$$w \cdot \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^*$$
$$\xi_i, \xi_i^* \geq 0, i = 1,2, \cdots, n$$

C is a constant known as the penalty factor that denotes the trade-off between error and margin. i.e. the optimization criterion penalizes data points whose y-values differ from $f(x)$ by more than ε. After applying the Lagrange multiplier, the minimization problem can be handled as the dual optimization problem as

$$\max_{\alpha,\alpha^*} -\frac{1}{2}(\alpha-\alpha^*)^T K(x_i,x_j)(\alpha-\alpha^*) - \varepsilon\sum_{i=1}^{n}(\alpha_i+\alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_i-\alpha_i^*)$$
$$\text{s.t. } \sum_{i=1}^{n}(\alpha_i-\alpha_i^*) = 0, 0 \le \alpha_i, \alpha_i^* \le C, i = 1,2,\dots,n. \quad (3)$$

α_i and α_i^* are Lagrange multipliers and the samples with positive and non-zero α_i and α_i^* are called the support vectors (SVs). By exploiting Karush-Kuhn-Tucker (KKT) conditions [33] which determine necessary and sufficient conditions for a global optimum, the product between dual variables and constraints has to vanish at the optimal solution and the parameter b can be computed as

$b = y_i - w \cdot \varphi(x) - \varepsilon$ for $0 \le \alpha_i \le C$

$b = y_i - w \cdot \varphi(x) + \varepsilon$ for $0 \le \alpha_i^* \le C$

Then the function $f(x)$ can be rewritten in support vector expansion shown in Equation (1) and the final regression function becomes:

$$f(x) = \sum_{i=1}^{n}(\alpha_i-\alpha_i^*)K(x_i,x) + b \quad (4)$$

Thus from Equation (4), one can observe that with the help of kernels the complexity of a function is independentof the dimensionality of the input space, and depends only on the number of support vectors.

4.2. Building of the Proposed W-SVR model to classify and predict early PD from healthy controls

Proposed algorithm uses SVR which uses weights associated with each of the training sample. This approach allows to learn individual samples by retaining Karush-Kuhn-Tucker (KKT) conditions on all previously seen samples. A Leave-one-out cross validation is implemented in which a single sample of PD data is used for testing and rest are used for training. This process is repeated forevery sample in the data. The dataset can be represented as $x_i \in R^m; i = 1, \dots, n$, where $n = 634$, is the number of samples and m is the number of features which is equal to 12 in both the classes (Healthy and early PD) and the binary class label $y \in R; y_i \in (EarlyPD = -1, Healthy = +1)$. Total number of samples included in the analysis is shown in Table 3.

Table 3: Case processing summary of model

		n	%
Selected cases	Included in analysis	60	94.6
	Missing cases	34	5.4
	Total	63	100.0

4.3. Formulation of proposed W - SVR

W-SVR is a variant of SVR in which each training instance possesses its own weight C_i, the weight for i^{th} training instance. the proposed algorithm reduces to ordinary SVR as a special case when $C_i = C$ for $i = 1, \dots, n$. The primal optimization problem for the W-SVR is

$$\min_{w,\xi,\xi^*} \frac{1}{2} \| w \|_2^2 + C_i \sum_{i=1}^{n} (\xi_i + \xi_i^*)$$

$$\text{s.t.} \quad \begin{aligned} & y_i - w \cdot \phi(x_i) - b \leq \varepsilon + \xi_i \\ & w \cdot \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, i = 1,2, \cdots, n \end{aligned} \tag{5}$$

Corresponding dual problem is:

$$\max_{\alpha_i} \quad -\frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j K(x_i, x_j) - \varepsilon \sum_{i=1}^{n} |\alpha_i| + \sum_{i=1}^{n} y_i \alpha_i$$

$$\text{s.t. } \sum_{i=1}^{n} \alpha_i = 0, \quad -C_i \leq \alpha_i \leq C_i, \quad i = 1, \cdots, n. \tag{6}$$

The final regression function for W-SVR is given by

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i) + b \tag{7}$$

The KKT conditions for the dual problem is given as

$$\begin{aligned} & |y_i - f(x_i)| \leq \varepsilon, \quad \text{if} \quad \alpha_i = 0 \\ & |y_i - f(x_i)| = \varepsilon, \quad \text{if} \quad 0 < |\alpha_i| < Ci, \\ & |y_i - f(x_i)| \geq \varepsilon, \quad \text{if} \quad |\alpha_i| = Ci, \\ & \sum_{i=1}^{n} \alpha_i = 0 \end{aligned}$$

The margin function $h(x_i)$ for x_i is:

$$h(x_i) = y_i - f(x_i) \tag{8}$$

4.4. Prediction/Prognostic model for early PD using multivariate logistic regression

Multivariate binomial logistic regression technique is also used to develop prediction/prognostic models in order to estimate the probability of risk in PD [31,32]. We also applied Logistic regression method to our data to compare and analyze performance in diagnosis process. Logistic regression predicts the logit of outcome (early PD or healthy)

from a set of predictors. The predicted probabilities (occurrence of class label PD) obtained from logit can then be revalidated with the actual outcome to determine if high probabilities are truly associated with higher risk of PDand low probabilities with lower risk of PD. In this model, 5 predictors (UPDRS, Lt.Cad, Rt.Cad, Lt.Put, Rt.Put)are used to fit a logit transformation as they are most characteristic features (as per Table 4). The probability of PD for each sample (n_i) is given by:

$$\text{logit}\,(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta_1 x_{1_i} + \cdots + \beta_k x_{k_i}$$

For each subject n_i, the π_i (likelihood of class label to be PD) is given by:

$$\pi_i = P(y_i = 1 \mid X_i)\beta, \alpha)$$

where α is the intercept (constant), $\beta = \{\beta_1, \ldots, \beta_k\}$ are the regression coefficients for the predictors and $X_i : X_i = [x_{1i}, \ldots, x_{ki}]$ is a sample with a set of k features. The regression coefficients are obtained using maximum likelihood approximation, and then solving the logit, probability of PD for each sample is obtained. The risk predictor is given by

$$\pi_i = \frac{1}{1 + \exp - (\alpha + \beta \cdot X_i)}$$

This PD risk estimation might be useful to categorize subjects into different risk categories.

5. Results and discussion

The experimental results of the proposed W-SVR for the prediction of PD progression are explained in this section. The number of samples for both the cases (Healthy, PD) and their features used to build the model is shown in table 1 and table 2. The dataset used for implementation are normalized for balancing the influence of each feature. After normalization, we take statistical analysis with the contribution to 95% for feature extraction. All statistical analysis of PD dataset is shown in table 4 and was carried out using SPSS 25 software (IBM SPSS Inc., Chicago, IL). The threshold of significance was de- fined as $\rho < 0.05$. During the construction of the regression model, we divide the dataset into training set and testing set. For implementation, we set optimal hyper parameter values for PD dataset as $\varepsilon = 2e^{-5}$, $C = 10$ and T_e(tolerance-error) $= 1e^{-6}$. We evaluate confusion matrix values and performance measures for the linear, polynomial order 4, sigmoid, RBF and Logistic kernels. All runs were performed on a computer with 3.4 GHz Intel i7 2600 CPU and 6 GB RAM using MATLAB 2017a.

5.1. Statistical analysis of input features

In machine learning, properly optimized feature extraction is the key to effective model construction. So to prove our most efficient regression model, we assessed the statistical significance of all 12 features through Linear regression.

Table 4: Statistical testing of each feature through Linear regression. β is the value of regression coefficient for the predictor in the model, *SE* is its standard error, *F* is F-statistic value, *df* is the degree of freedom, *ρ-value* is the significance of regression coefficient, R^2 is the measure of coefficient of determination that tells how close the data to the fitted regression line. *MSE* is the Mean square error. Note: the table does not show the constants obtained in the model.

Predictor	β	SE_β	F	df	ρ-value	R^2	MSE
Age	0.0453	0.47	1.273	1	0.259	0.045	0.223
Gender	0.0053	0.47	0.018	1	0.895	0.000	0.224
Handed	0.1080	0.47	7.504	1	0.006	0.012	0.221
F.History	0.2449	0.45	40.010	1	0.000	0.60	0.210
UPDRS	0.7536	0.30	802.673	1	0.000	0.567	0.094
MoCA	-0.2162	0.46	30.867	1	0.000	0.047	0.213

TD	0.5670	0.39	299.784	1	0.000	0.322	0.152
PIGD	0.4629	0.41	171.740	1	0.008	0.214	0.176
Lt.Cad	-0.613	0.37	360.489	1	0.000	0.376	0.137
Rt.Cad	-0.599	0.37	334.805	1	0.000	0.359	0.141
Lt.Put	-0.821	0.26	1233.945	1	0.000	0.674	0.072
Rt.Put	-0.808	0.27	1122.235	1	0.000	0.652	0.077

Table 4 shows the result of regression analysis of feature set: Age, Gender, Handedness(Handed), Family History(F.History), MDS-UPDRS(UPDRS), MoCA, TD, PIGD, left caudate(Lt.Cad), Right caudate(Rt.Cad), Left putamen(Lt.Put), Right putamen (Rt.Put) SBR. It is observed that all of the ten features except Age and Gender are statistically significant with $\rho < 0.05$. The table also shows the R^2 and Mean Square error(MSE). High R^2 and Low Residual values are obtained for putamenal SBR (both right and left) when compared to the caudate SBR. This indicates that putamenal SBR had higher discriminative power than caudate SBR in distinguishing early PD from healthy controls. This is also evident from the histogram plots which shows lower overlaps for putamenal SBRs (Figure 3: (e,g)) than for caudate SBRs (Figure 3: (a,c)).

5.2. W-SVR algorithm for classification and prediction to distinguish early PD from Healthy controls

the proposed Support vector regression algorithm optimized with weights associated to each sample is used to automate the prediction and diagnosis of early PD from healthy controls. W-SVR is used with different kernels such as linear, Polynomial, sigmoid, radial basis (RBF) and logistic to show the margin variations in classifying data and compared with standard SVR. W-SVR can be used as both classification and regression method by maintaining the main feature thatcharacterize the algorithm i.e. maximal margin. In regression, a margin of tolerance ε is set in approximation tothe support vectors. The main idea behind this work is (i) to minimize classification error and (ii) individualizingthe hyper plane which maximizes the margin. These two objectives are achieved in W-SVR with better distance between classes of support vectors and reduced number of error vectors.

Figure 5:(a,b,c,d) Figure 6:(a,b,c,d) Figure 7:(a,b) depicts contour plots showing distribution lines of regression, number of support vectors (SV) and errorvectors(EV) for proposed W-SVR and standard SVR algorithms with Linear, Polynomial of order 4, sigmoid, RBF, Logistic kernel functions respectively. It can be seen that proposed model has achieved better margin distance between two classes of data for different kernels and the number of error vectors are drastically reduced than the standard SVR. This shows that the proposed modelis performing better than standard SVR in classifying the data. Proposed model with RBF kernel has depicted largemargin distribution with low error when compared to other kernels (Figure 6 (c,d)). However sigmoid kernel gives large margin distribution, the number of error vectors are more as shown in Figure 6: (a,b) and it is also evidentin Figure 6:(f) as its MSE is high.

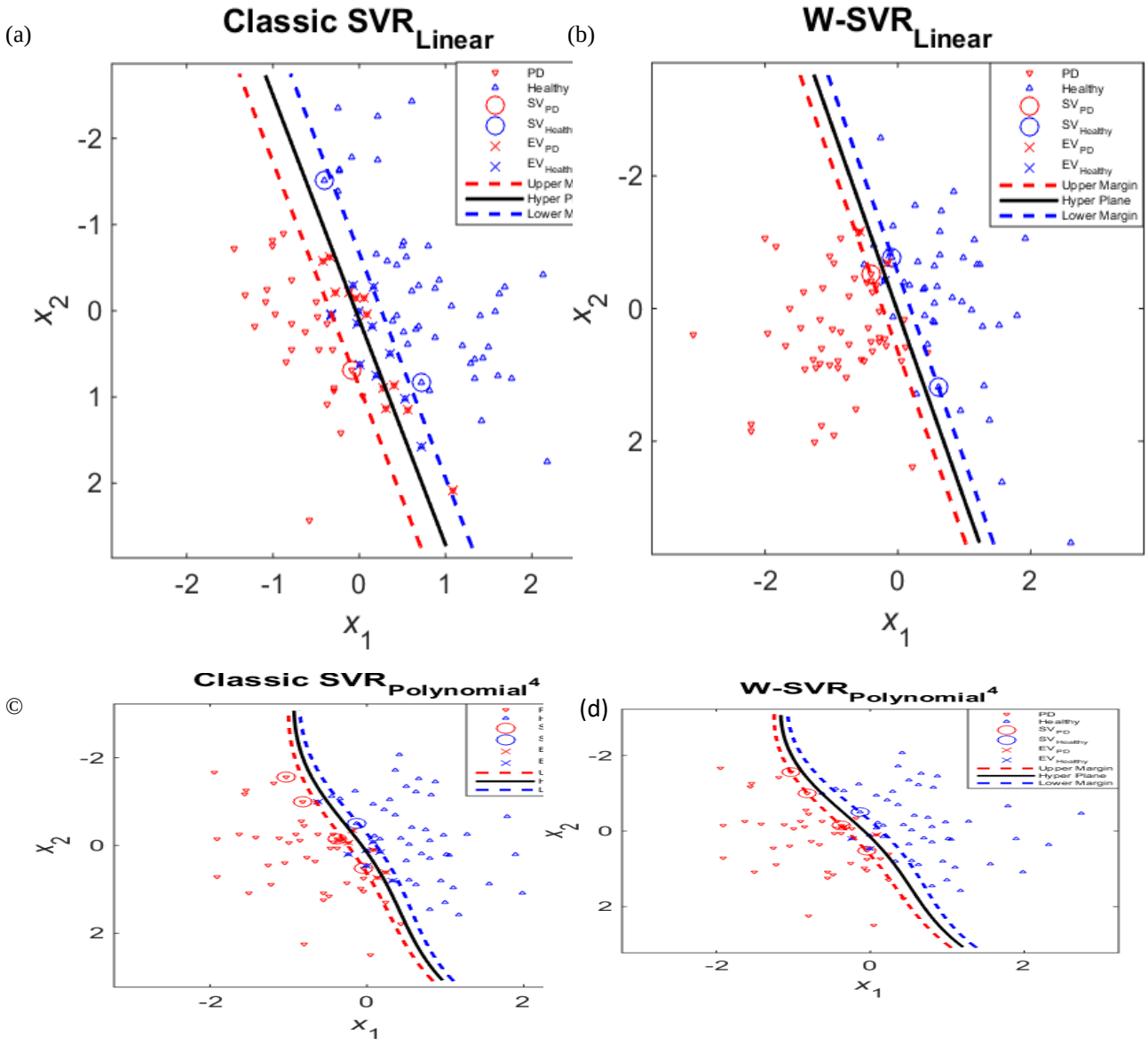

Figure 5: Contour plots showing the regression distribution lines of two classes of PD data using proposed W-SVR and standard SVR algorithms with Linear and polynomial of order 4 kernels. Figure (a,b,c,d,) depicts the margin separating two classes of data (Healthy and PD) where SV=support vectors and EV=error vectors. It can be seen that W-SVR gives the large margin distribution with less EVs when compared to classic SVR.

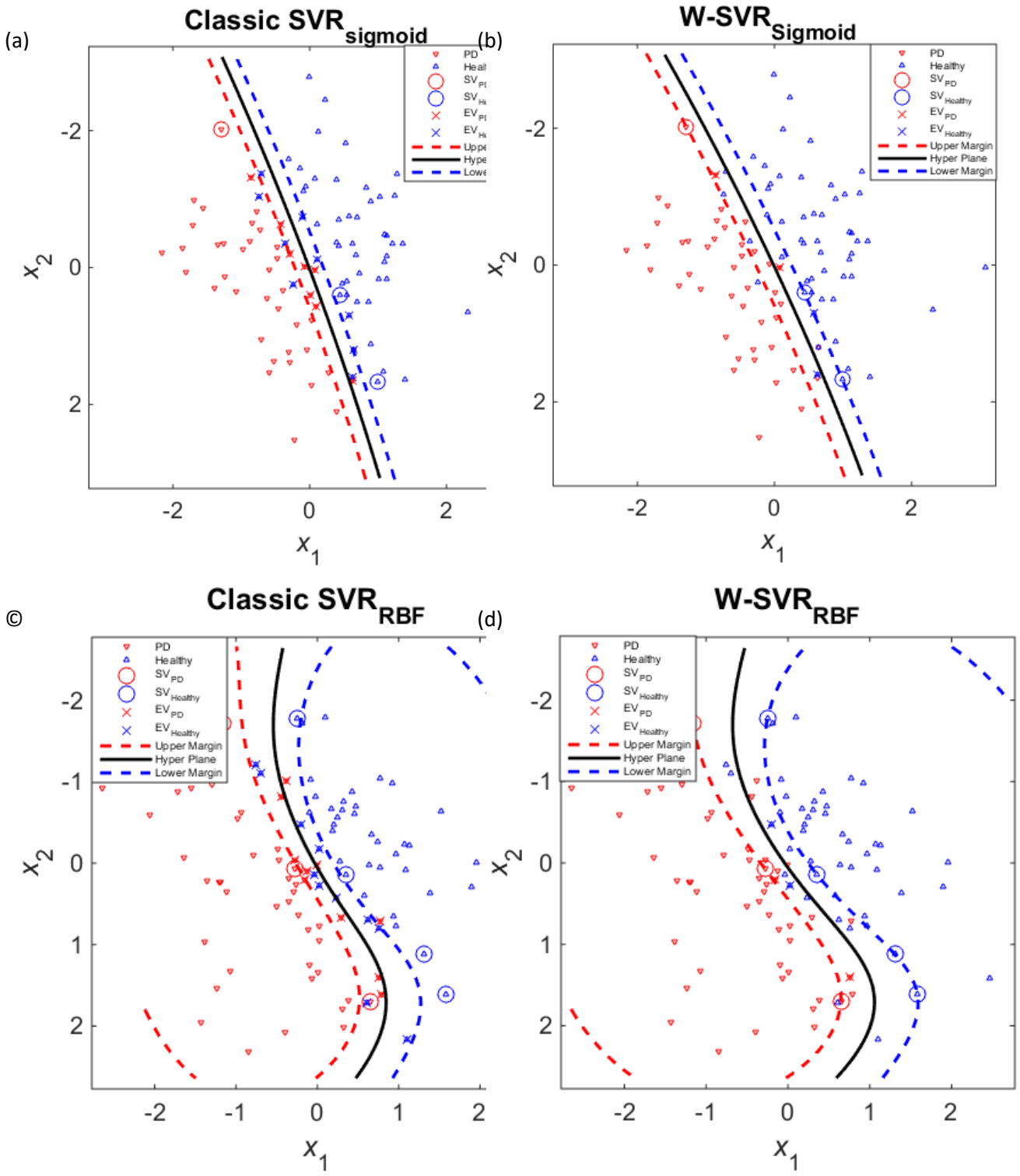

Figure 6: Contour plots showing the regression distribution lines of two classes of PD data using proposed W-SVR and standard SVR algorithms with Sigmoid and RBF kernels. Figure (a,b,c,d,) depicts the margin separating two classes of data (Healthy and PD) where SV=support vectors and EV=error vectors. It can be seen that W-SVR gives the large margin distribution with less EVs when compared to classic SVR.

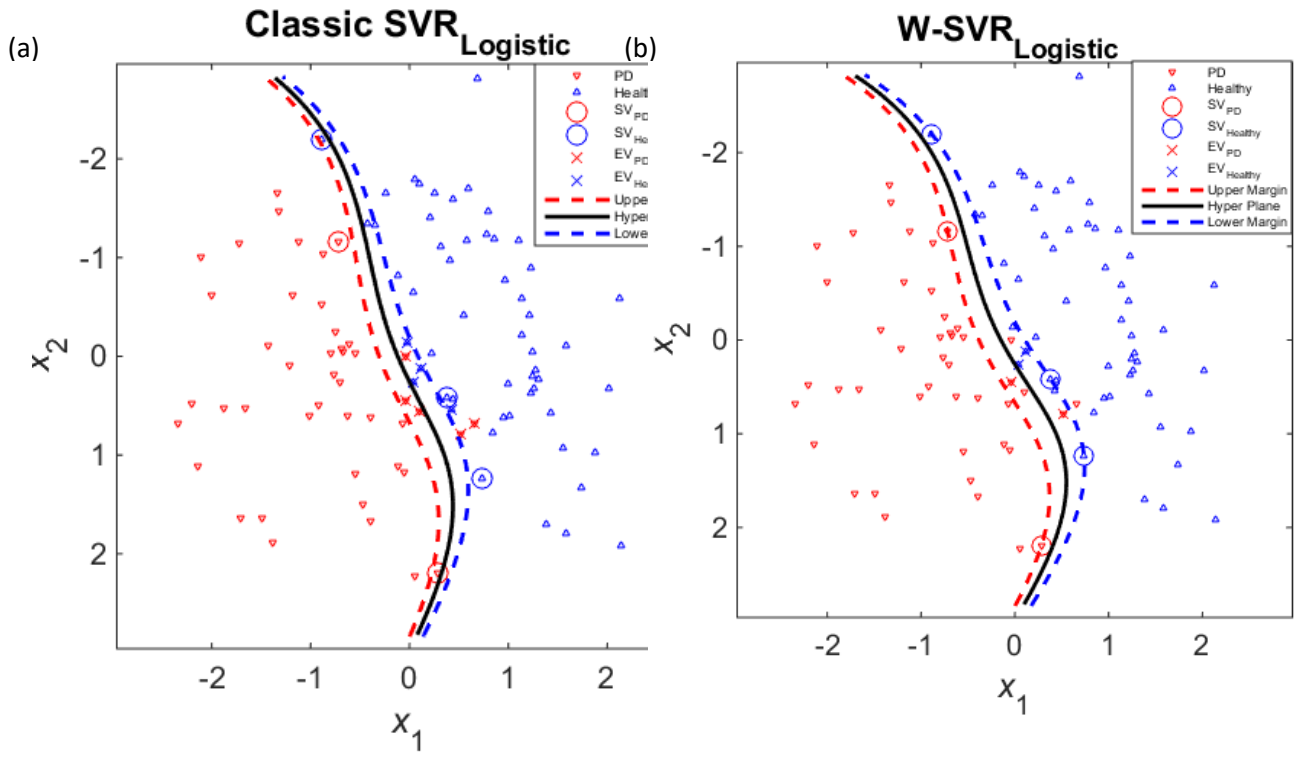

Figure 7: Contour plots showing the regression distribution lines of two classes of PD data using proposed W-SVR and standard SVR algorithms with Logistic kernel function. Figure (a,b) depicts the margin separating two classes of data (Healthy and PD) where SV=support vectors and EV=error vectors. It can be seen that W-SVR gives the large margin distribution with less EVs when compared to classic SVR.

5.3. Performance Measure comparison

To measure the forecasting accuracy, some widely used scale-dependent and scale-independent statistical indicators were examined as follows: mean square error (*MSE*) [30], the coefficient of determination (R^2) [30]. *MSE* is defined as the average of squares of the errors given in Equation (21). Smaller the value of *MSE*, better the forecasting performance of the model. R^2 ranging from 0 to 1 is a measure that allows one to determine the certainty of predictions from actual value given in Equation (22).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{X}_i - X_i\right)^2$$

$$R^2 = \frac{n\sum_{i=1}^{n}\hat{X}_i X_i - \sum_{i=1}^{n}\hat{X}_i \sum_{i=1}^{n} X_i}{\sqrt{\left(n\sum_{i=1}^{n}\hat{X}_i^2 - \left(\sum_{i=1}^{n}\hat{X}_i\right)^2\right)\left(n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2\right)}}$$

$\hat{X}_i$ is the vector denoting values of n number of predictions and X_i is a vector representing n number of true values. The Figure 8 describes a comparison of MSE values for proposed and existing algorithms. Figure 8 depicts the boxplotof MSE values for Linear, Polynomial order 4, Sigmoid, RBF and Logistic kernels. Boxplot show the minimum, median and maximum values of MSE. Results indicate that proposed W-SVR achieved the lower values ofMSE with each kernel functions when compared with standard SVR, which means that W-SVR can provide better fittingquality. Table 5 summarizes the results

of performance measures. It is observed that W-SVR performs better and can provide better fitting qualitythan SVR by achieving small MSE and large R^2 values and also consumes less time than standard SVR in each of the kernel used.

.

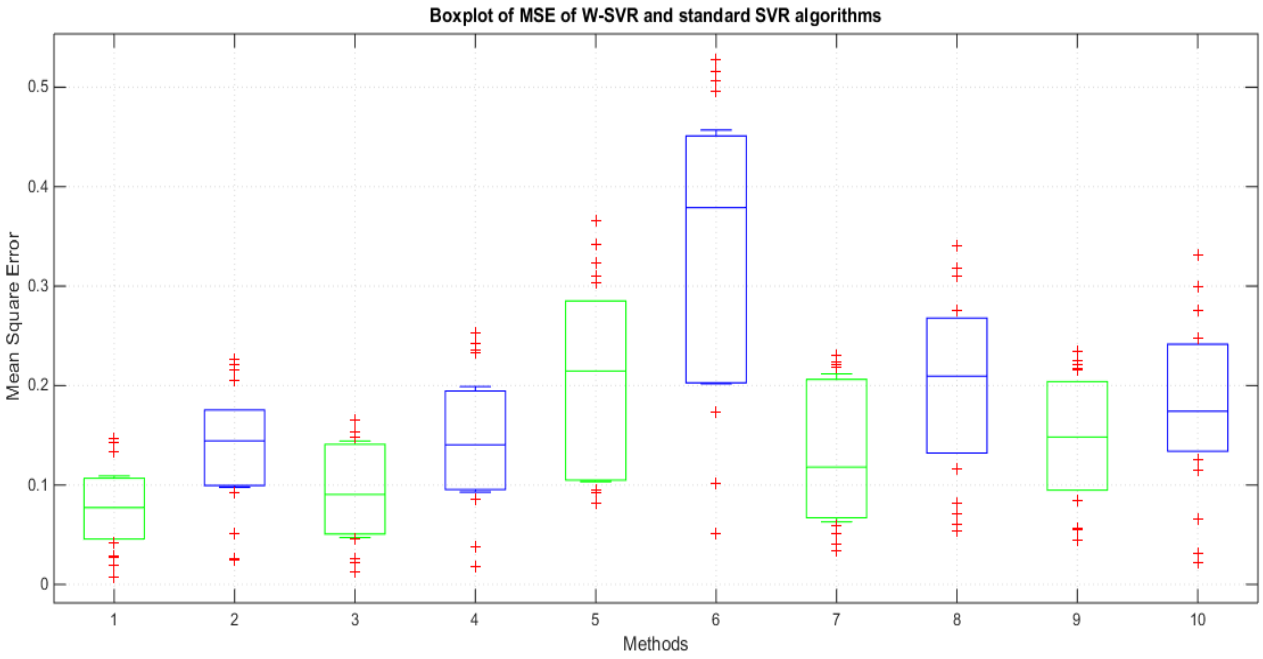

Figure 8: The comparison of Mean Square Error (MSE) of W-SVR and SVR algorithms with different kernels. Figure depicts the boxplot of the MSE of of proposed W-SVR and existing SVR algorithms. In figure, values of x axis from left to right are: 1.W-SVR_{Linear}, 2.SVR_{Linear}, 3.W-$SVR_{4^{th}orderPolynomial}$, 4.$SVR_{4^{th}orderPolynomial}$, 5.W-$VR_{Sigmoid}$, 6.$SVR_{Sigmoid}$ 7.W-SVR_{RBF}, 8.SVR_{RBF}, 9.W-$SVR_{Logistic}$, 10.$SVR_{Logistic}$.

Table 5: Confusion matrix and performance measures for the W-SVR and classic SVR with different kernels

Kernal	Model	Fscore(%)	CPU-time(sec)	MSE	R^2
Linear	W-SVR	97.71	1.60	0.078	0.899
	SVR	92.30	1.98	0.155	0.702
Polynomial^4	W-SVR	97.90	4.10	0.092	0.859
	SVR	91.54	5.40	0.143	0.729
Sigmoid	W-SVR	98.00	3.50	0.223	0.594
	SVR	91.20	3.82	0.372	0.435
RBF	W-SVR	97.74	2.26	0.131	0.758
	SVR	89.95	2.67	0.202	0.641
Logistic	W-SVR	97.81	3.20	0.156	0.695
	SVR	90.95	3.70	0.176	0.662

Table 6 gives the classification accuracy of proposed model with RBF kernel. The model achieved 96.73% accuracyin classifying the PD data which is more than the accuracy of standard SVR (see Table 5).

Table 6: Classification table using W-SVR_{RBF}

Observed group	Predicted group		
	Healthy	PD	% Accuracy
Healthy	188	7	96.40
Early PD	12	393	97.03
Overall %			96.73

5.4. Prediction/Prognostic model for early PD using multivariate logistic regression

Logistic regression outputs the predicted probability of occurrence of the PD class. To validate these probabilities, the degree to which predicted probabilities agree with actual outcome is shown through a classification table Table 7.The overall classification accuracy was as high as 93.5% indicating that the model with 5 predictors performs wellin predicting the subject outcome. This PD risk estimation might be useful to categorize subjects into different risk categories.

Table 7: Classification table using Logistic regression

Observed group	Predicted group		
	Healthy	PD	% Accuracy
Healthy	174	21	89.2
Early PD	18	387	95.5
Overall %			93.5

It is observed that the classification accuracy is not as high as that we obtained using proposed classifier withRBF kernel. This is because logistic regression models are discriminative models for classification that produceslinear decision boundaries, and are not flexible as the non-linear models.

From above results, we observe that all the kernels functions used with proposed model are performed well compared to standard SVR giving the better accuracy of 96.73% with RBF kernel and also observe that high R^2 and low MSE values. This shows that the proposed algorithm works fine even with large data by taking less computation time. In comparison to the related researches [18, 19, 21–23, 28], all the studies have used either quantitative or semi-quantitative parameters obtained from neuroimaging data and applied classification techniquessuch as SVMs. In our work, large PD dataset with considerably high number of features are used to train SVR incrementally and classification margin is optimized with modified Frank-Wolfe method. Thus, this work automates the early detection of PD from healthy subjects with better accuracy.

6. Conclusion and Future work

Initial and accurate diagnosis of PD is very important to apply early management approaches. Automated PD diagnostic techniques mostly rely on standard SVMs which suffer with slow convergence rate. In this paper, a variant of standard SVR algorithm is proposed, named as W-SVR to classify early PD subjects and predict disease progression with accelerated convergence rate than standard SVR. The model is built with Linear, 4^{th} order Polynomial, Sigmoid, RBF, Logistic Kernel to obtain better classification and predict the disease progression. It is shown that the W-SVR algorithm comparatively achieved better performance in less amount of time thanwhen compared to standard SVR for all kernel functions. From our work, we state that classification and prediction problems are solved in less computational time even when used with large dataset. In future work, W-SVR can be applied to solve several real-world problems, including financial data prediction, weather forecasting and anomaly detection.

Author's contribution

Lavanya Madhuri Bollipo built the conceptual design of the research, performed data analysis, experimental results and wrote the first draft and also successive revision of the manuscript. **Kadambari. KV** involved in correcting the manuscript draft and obtained important intellectual content.

Conflicts of Interest

None

Acknowledgment

PPMI - a public-private partnership (www.ppmi-info.org) - is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including list of all of the PPMI funding partners found at www.ppmi-info.org/fundingpartners.

References

1. DB Calne, J William Langston, WR Wayne Martin, A Jon Stoessl, Thomas J Ruth, Michael J Adam, Brian D Pate, and Michael Schulzer. Positron emission tomography after mptp: observations relating to the cause of parkinson's disease. *Nature*, 317(6034):246, 1985.

2. J Booij, G Tissingh, GJ Boer, JD Speelman, JC Stoof, AG Janssen, E Ch Wolters, and EA Van Royen. [123i] fp-cit spect shows a pronounced decline of striatal dopamine transporter labelling in early and advanced parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 62(2):133–140, 1997.
3. Garrett E Alexander. Biology of parkinson's disease: pathogenesis and pathophysiology of a multisystem neurodegenerative disorder. *Dialogues in clinical neuroscience*, 6(3):259, 2004.
4. Daniel Weintraub, Paul J Moberg, John E Duda, Ira R Katz, and Matthew B Stern. Effect of psychiatric and other nonmotor symptoms on disability in parkinson's disease. *Journal of the American Geriatrics Society*, 52(5):784–788, 2004.
5. Dino Muslimović, Bart Post, Johannes D Speelman, and Ben Schmand. Cognitive profile of patients with newly diagnosed parkinson disease. *Neurology*, 65(8):1239–1245, 2005.
6. Huajun Jin, Arthi Kanthasamy, Anamitra Ghosh, Vellareddy Anantharam, Balaraman Kalyanaraman, and Anumantha G Kanthasamy. Mitochondria-targeted antioxidants for treatment of parkinson's disease: pre- clinical and clinical outcomes. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1842(8): 1282–1294, 2014.
7. Jacques Darcourt, Jan Booij, Klaus Tatsch, Andrea Varrone, Thierry Vander Borght, Ö zlem L Kapucu, Kjell Nãgren, Flavio Nobili, Zuzana Walker, and Koen Van Laere. Eanm procedure guidelines for brain neuro- transmission spect using 123 i-labelled dopamine transporter ligands, version 2. *European journal of nuclear medicine and molecular imaging*, 37(2):443–450, 2010.
8. Barbara Palumbo, Mario Luca Fravolini, Susanna Nuvoli, Angela Spanu, Kai Stephan Paulus, Orazio Schillaci, and Giuseppe Madeddu. Comparison of two neural network classifiers in the differential diagnosis of essential tremor and parkinsons disease by 123 i-fp-cit brain spect. *European journal of nuclear medicine and molecular imaging*, 37(11):2146–2153, 2010.
9. Schillaci, A Chiaravalloti, M Pierantozzi, B Di Pietro, G Koch, C Bruni, P Stanzione, and A Stefani. Different patterns of nigrostriatal degeneration in tremor type versus the akinetic-rigid and mixed types of parkinson's disease at the early stages: molecular imaging with 123i-fp-cit spect. *International journal of molecular medicine*, 28(5):881–886, 2011.
10. David SW Djang, Marcel JR Janssen, Nicolaas Bohnen, Jan Booij, Theodore A Henderson, Karl Herholz, Satoshi Minoshima, Christopher C Rowe, Osama Sabri, John Seibyl, et al. Snm practice guideline for dopamine transporter imaging with 123i-ioflupane spect 1.0. *Journal of Nuclear Medicine*, 53(1):154, 2012.
11. Kimberly D Seifert and Jonathan I Wiener. The impact of datscan on the diagnosis and management of movement disorders: A retrospective study. *American journal of neurodegenerative disease*, 2(1):29, 2013.
12. Francisco Jesús Martínez-Murcia, Juan Manuel Górriz, Javier Ramírez, IA Illán, Andrés Ortiz, Parkinson's Progression Markers Initiative, et al. Automatic detection of parkinsonism using significance measures and component analysis in datscan imaging. *Neurocomputing*, 126:58–70, 2014.
13. Eduardo Tolosa, Thierry Vander Borght, Emilio Moreno, and DaTSCAN Clinically Uncertain Parkinsonian Syndromes Study Group. Accuracy of datscan (123i-ioflupane) spect in diagnosis of patients with clinically uncertain parkinsonism: 2-year follow-up of an open-label study. *Movement Disorders*, 22(16):2346–2351, 2007.
14. SJ Ortega Lozano, MD Martinez del Valle Torres, E Ramos Moreno, S Sanz Viedma, T Amrani Raissouni, andJM Jiménez-Hoyuela. Quantitative evaluation of spect with fp-cit. importance of the reference area. *Revista española de medicina nuclear (English Edition)*, 29(5):246–250, 2010.
15. David J Towey, Peter G Bain, and Kuldip S Nijran. Automatic classification of 123i-fp-cit (datscan) spect images. *Nuclear medicine communications*, 32(8):699–707, 2011.
16. IAa Illán, JMa Górriz, Ja Ramírez, Fa Segovia, JMb Jiménez-Hoyuela, and SJ Ortega Lozano. Automatic assistance to parkinsons disease diagnosis in datscan spect imaging. *Medical physics*, 39(10):5971–5980, 2012.
17. F Segovia, JM Górriz, J Ramírez, I Alvarez, JM Jiménez-Hoyuela, and SJ Ortega. Improved parkinsonism diagnosis using a partial least squares based approach. *Medical physics*, 39(7Part1):4395–4403, 2012.
18. Barbara Palumbo, Mario Luca Fravolini, Tommaso Buresta, Filippo Pompili, Nevio Forini, Pasquale Nigro, Paolo Calabresi, and Nicola Tambasco. Diagnostic accuracy of parkinson disease by support vector machine (svm) analysis of 123i-fp-cit brain spect data: implications of putaminal findings and age. *Medicine*, 93(27), 2014.

19. R Prashanth, Sumantra Dutta Roy, Pravat K Mandal, and Shantanu Ghosh. Automatic classification and
20. prediction models for early parkinsons disease diagnosis from spect imaging. *Expert Systems with Applications*, 41(7):3333–3342, 2014.
21. R Prashanth, Sumantra Dutta Roy, Pravat K Mandal, and Shantanu Ghosh. High-accuracy detection of early parkinson's disease through multimodal features and machine learning. *International journal of medical informatics*, 90:13–21, 2016.
22. Antonio Augimeri, Andrea Cherubini, Giuseppe Lucio Cascini, Domenico Galea, Maria Eugenia Caligiuri, Gaetano Barbagallo, Gennarina Arabia, and Aldo Quattrone. Cadacomputer-aided datscan analysis. *EJNMMI physics*, 3(1):4, 2016.
23. Haijun Lei, Zhongwei Huang, Jian Zhang, Zhang Yang, Ee-Leng Tan, Feng Zhou, and Baiying Lei. Joint detection and clinical score prediction in parkinson's disease via multi-modal sparse learning. *Expert Systems with Applications*, 80:284–296, 2017.
24. Francisco PM Oliveira, Diogo Borges Faria, Durval C Costa, Miguel Castelo-Branco, and João Manuel RS Tavares. Extraction, selection and comparison of features for an effective automated computer-aided diagnosis of parkinsons disease based on [123 i] fp-cit spect images. *European journal of nuclear medicine and molecularimaging*, 45(6):1052–1062, 2018.
25. PB Schiilkop, Chris Burgest, and Vladimir Vapnik. Extracting support data for a given task. In *Proceedingsof the 1st international conference on knowledge discovery & data mining*, pages 252–257, 1995.
26. Bernardete Ribeiro. Support vector machines for quality monitoring in a plastic injection molding process.
27. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 35(3):401–410,2005.
28. Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14 (3):199–222, 2004..
29. Begüm Demir and Lorenzo Bruzzone. A multiple criteria active learning method for support vector regression.
30. *Pattern recognition*, 47(7):2558–2567, 2014.
31. Chih-Chung Chang. " libsvm: a library for support vector machines," acm transactions on intelligent systemsand technology, 2: 27: 1–27: 27, 2011. *http://www. csie. ntu. edu. tw/˜ cjlin/libsvm*, 2, 2011.
32. Kenneth Marek, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Chris Coffey, Karl Kieburtz, Emily Flagg, Sohini Chowdhury, et al. The parkinson progression marker initiative (ppmi). *Progress in neurobiology*, 95(4):629–635, 2011.
33. Dongning Guo, Shlomo Shamai, and Sergio Verdú. Mutual information and minimum mean-square error in gaussian channels. *IEEE Transactions on Information Theory*, 51(4):1261–1282, 2005.
34. Viv Bewick, Liz Cheek, and Jonathan Ball. Statistics review 14: Logistic regression. *Critical care*, 9(1):112,2005.
35. Stephan Dreiseitl and Lucila Ohno-Machado. Logistic regression and artificial neural network classificationmodels: a methodology review. *Journal of biomedical informatics*, 35(5-6):352–359, 2002.