Identifying Parkinson’s Disease through Speech Patterns

Abstract

Parkinson’s disease is a wide-spread, devastating neurodegenerative disorder that affects a patient’s motor skills, which contributes to altered speech production. To predict whether a person has Parkinson’s disease based on their vocal patterns, we use data from The National Center of Voice and Speech to create support vector classifier and support vector machine models, which classify people as either having or not having Parkinson’s disease based on various vocal metrics obtained from a recording of the person’s speech. From these models, we find that the support vector machine with a radial kernel works best with a testing accuracy of 86% and a testing ROC AUC of 88.3%. Given the small sample size, class imbalance and lack of interpretability, future areas of research could involve more data collection and use of more interpretable classification models.

Introduction

Parkinson's disease (PD) is a neurodegenerative disorder that affects mostly dopamine-producing ("dopaminergic") neurons in an area of the brain called substantia nigra [3]. According to the Centers for Disease Control and Prevention, it is the second most common neurodegenerative disease after Alzheimer's disease. [4] Each year, about 60,000 Americans are diagnosed with PD [3]. It is further estimated that, by 2020, nearly 1 million Americans will be diagnosed with PD [3]. Being able to accurately estimate the number of people in different localities living with PD helps organizations like the Parkinson's Foundation attract the attention of relevant authorities to look into the disease [3]. Machine learning methods have proven to be reliable in providing accurate estimates in other fields and, thus, can be applied in predicting whether a person has PD or not. Different algorithms have already been researched and applied to the problem, such as Classification and Regression Trees (CART), Support Vector Machines (SVM) and Artificial Neural Networks (ANN) [5]. Additionally, the SVM method used achieved an accuracy of 79.98% without feature selection and 93.84% with feature selection [5]. This shows that SVM is a robust algorithm to apply to the problem. We, thus, wish to investigate the application of machine learning - specifically, support vector machines - in the classification of people according to whether they have PD or not.

Methods

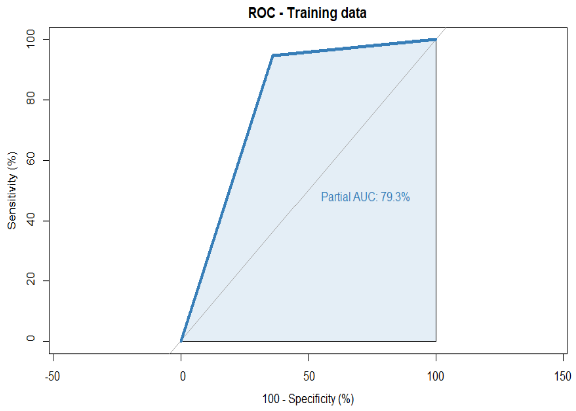
Data. This data was collected through a collaboration between University of Oxford Professor Max Little and the National Center for Voice and Speech in Denver, Colorado in 2009. [1] Voice recordings of 31 different participants were taken at the National Center for Voice and Speech, with each participant providing roughly 6 different recordings for a total of 195 recordings of "sustained vowel phonations." A sustained vowel phonation is one kind of vocal test used to assess potential impairment in which the participant produces a specific vowel sound and tries to maintain that pitch for as long as they can. [2] The participants are between 46 and 85 years old. Of the 31 participants, 23 of them were diagnosed with Parkinson's disease. Of those with PD, the time since their diagnosis ranges from 0 to 28 years. [2] The speech signal recordings were analyzed for various signal properties through the software Multi-Dimensional Voice Program (MDVP). [2]

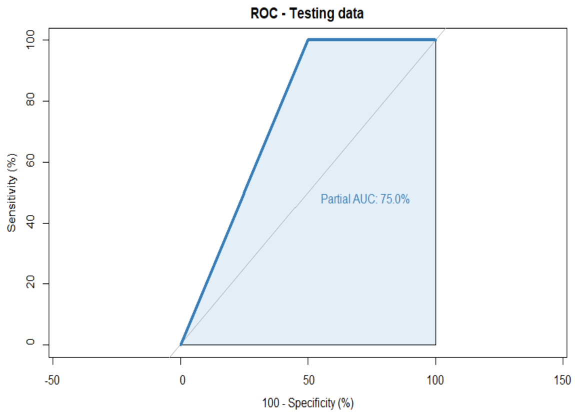
Variables. This dataset contains 195 observations and 22 predictor variables. The response variable, status, is categorical and represents whether or not the participant has PD. The first variable is the name of the participant. The other 22 predictor variables represent different wave measurements of the voice recording signal, which are used to assess potential voice disorders. These include the natural frequency of the signal (F0), the highest frequency of the recording (Fhi), the lowest frequency of the recording (Flo), jitter as a percentage, absolute jitter, relative amplitude perturbation (RAP), period perturbation quotient (PPQ), jitter DDP, shimmer, shimmer in decibels, 3, 5, and 11-point amplitude perturbation quotients (APQ3, APQ5, APQ), shimmer DDA, noise to harmonics ratio (NHR), harmonics to noise ratio (HNR), recurrence period density entropy (RPDE), detrended fluctuation analysis (DFA), correlation dimension (D2), and pitch period entropy (PPE). Jitter is the extent of variation in frequency between vocal cycles, while shimmer is the extent of variation in amplitude between vocal cycles.

Analytic Methods. We begin by randomly splitting our dataset into a training set with 150 observations and a testing set with 45 observations. Then, using the training data and a cost of 100, we create a support vector classifier model. A support vector classifier model is a maximal margin classifier that allows for softer margin. Varying the hyperparameter known as cost determines how soft the margin is - a bigger cost leads to a harder margin. We create 4 different models by varying the cost from really small to really large to get a sense of the range of values to use in cross-validation. The costs used are 1e-06, 1e-03, 1e02 and 1e05. We also test for whether we need scaling or not using models with the 2nd cost and comparing their accuracies. We then conduct cross-validation with said ranges and obtain the best model from which we create the training and test confusion matrix and extract classification metrics such as accuracy. Finally, we conduct cross-validation for both polynomial and radial kernel support vector machines aiming to find the best degree and gamma respectively. Degree and gamma are hyperparameters that account for the dimensionality and variance of the vector dot products respectively.

Results

The support vector classifier model with a cost of 100 has 38 support vectors, a training accuracy of 87.33%, and a testing accuracy of 86.67%. The training and testing ROC curves for this model are shown below. The training and test confusion matrices for the other 3 models can be found in the appendix.





Without scaling, the model with cost = 0.001 achieves 84% training accuracy and 82.222% test accuracy. The model that has scaling achieves 76% training accuracy and 73.33% test accuracy.

After creating and running the 4 models, we decided on the range 1e-2 to 1e10 for support vector machine tuning. The best model has a cost 0.01 and 63 support vectors with $\beta_0 = 5.83979$. The corresponding βvector can be found in the appendix. The model has a training accuracy of 84.67% and a test accuracy of 80%..

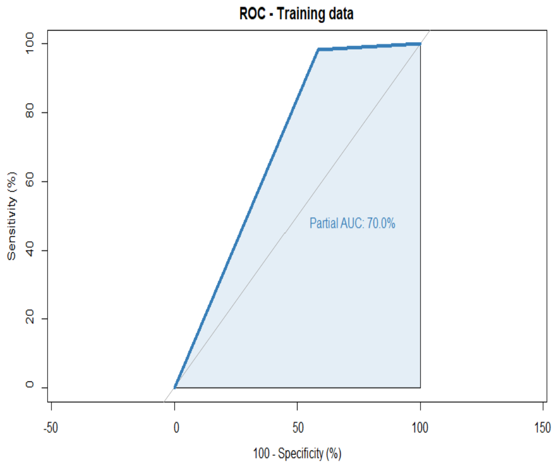
Training Confusion Matrix: 0.98 sens

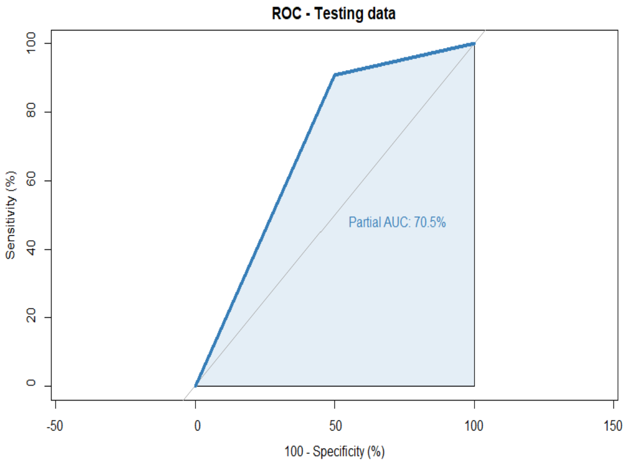
	Actual No PD	Actual PD
Predicted No PD	15	2
Predicted PD	21	112

Testing Confusion Matrix: 0.91 sens

	Actual No PD	Actual PD
Predicted No PD	6	3
Predicted PD	6	30

The training and testing ROC curves for the support vector classifier model with a cost of 0.01, as determined by cross-validation, are shown below. The ROC AUC for the training data is 70%, and the ROC AUC for the testing data is 70.5%.





The best model from the polynomial kernel cross-validation has degree = 2, cost = 0.01 and 49 support vectors. It achieves a training accuracy of 90% and a test accuracy of 84%. The best model from the radial kernel cross-validation has gamma = 1e-04, cost = 52,631,578 and 37 support vectors. It achieves 100% training accuracy and 86.67% testing accuracy. The confusion matrices as well as the ROC curves can be found in the appendix.

Discussion/Conclusion

In terms of classification accuracy, the radial kernel model fit found through cross validation is the best model with 100% training accuracy and 86.67% testing accuracy. All models exhibit comparable and relatively high test accuracies, an indicator that they generalize rather well. Additionally, scaling the data in our case seems to not be advantageous as it reduces the classification accuracy. Lower degrees and lower gamma values perform better for non-linear kernel models fitted to this dataset. The latter, particularly, seems to indicate that higher variance when increasing dimensionality improves classification accuracy.

Regarding the context of our data, we would want to be extremely confident when classifying a person as having Parkinsons. Thus, the model with best performance could be the one that maximizes sensitivity or the true positive rate. Only two models (*SVC, cost = 1e5, scale = False* and *radial model*) record test sensitivities below 90%. Other models have sensitivities within a similar range, with some even having perfect sensitivities on both training and testing data. This occurrence as a result of the model having no false negatives and, sometimes, no true negatives either. The models that exhibit this condition are linear kernel models with relatively low costs which could imply that softer margins worsens classification accuracy for this dataset.

Limitations/Future Work

One major limitation of using support vector classifiers and support vector machines is the lack of interpretability. These models do not provide information on how different predictor variables affect whether a patient has Parkinson's disease, nor the significance of the predictors. Thus, if we were interested in finding which verbal tests provide the most information toward diagnosing Parkinson's disease, we would select a different kind of model. Another limitation is the small sample size. With only 195 vocal recordings coming from only 32 participants, this dataset is relatively small, so it may be easier to generalize results with a larger sample size.

Future research can explore classifying people with PD using a dataset with more observations. It also seems worth considering other types of models, such as classification trees or logistic regression models in order to identify which vocal metrics provide the most important information for diagnosing PD. It's also important to remember that decreased speech production abilities is only one of many effects of Parkinson's disease, so it may also be beneficial to examine other factors, such as reaction times or working memory, in addition to vocal patterns. The advantage of using vocal patterns is that these speech production tests are non-invasive and fairly easy to obtain. However, there may be other factors that can produce even more accurate classification models.

The more we learn about Parkinson's disease and the more accurately we can diagnose it, the more we can help those suffering with PD.

References

[1] Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection | BioMedical Engineering OnLine | Full Text (biomedcentral.com)

Little, M.A., McSharry, P.E., Roberts, S.J. *et al.* Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection. *BioMed Eng OnLine* 6, 23 (2007). https://doi.org/10.1186/1475-925X-6-23

[2] Suitability of dysphonia measurements for telemonitoring of Parkinson's disease (maxlittle.net)

Max A. Little[1], *Member IEEE*, Patrick E. McSharry[1], *Senior Member IEEE*, Eric J. Hunter[2], Jennifer Spielman[2], Lorraine O. Ramig[2,3]

[1]Systems Analysis, Modelling and Prediction Group, University of Oxford, UK. [2]National Center for Voice and Speech, The Denver Center for the Performing Arts, Denver, Colorado, US. [3]Department of Speech, Language and Hearing Science, University of Colorado at Boulder, Colorado, US.

[3] Parkinsons' Foundation

Parkinson's Foundation. "Parkinson's Foundation | Better Lives. Together." Parkinson's Foundation, 14 Apr. 2021, www.parkinson.org.

[4] Genetics, coffee consumption and Parkinson's disease

CDC. "Genetics, Coffee Consumption, and Parkinson's Disease | CDC." Centers for Disease Control and Prevention, 29 Jan. 2013, www.cdc.gov/genomics/hugenet/casestudy/parkinson/parkcoffee_view.htm.

[5] Early Diagnosis of Parkinson's Disease Using Machine Learning Algorithms

Karapinar Senturk, Zehra. "Early Diagnosis of Parkinson's Disease Using Machine Learning Algorithms." Medical Hypotheses, vol. 138, 2020, p. 109603. Crossref, doi:10.1016/j.mehy.2020.109603.

Appendix

Descriptive Stats Summary of entire dataset

MDVP.Fo.Hz.	MDVP.Fhi.Hz.	MDVP.Flo.Hz.	MDVP.Jitter...	MDVP.Jitter.Abs.	MDVP.RAP
Min. : 88.33	Min. :102.1	Min. : 65.48	Min. :0.001680	Min. :7.000e-06	Min. :0.000680
1st Qu.:117.57	1st Qu.:134.9	1st Qu.: 84.29	1st Qu.:0.003460	1st Qu.:2.000e-05	1st Qu.:0.001660
Median :148.79	Median :175.8	Median :104.31	Median :0.004940	Median :3.000e-05	Median :0.002500
Mean :154.23	Mean :197.1	Mean :116.32	Mean :0.006220	Mean :4.396e-05	Mean :0.003306
3rd Qu.:182.77	3rd Qu.:224.2	3rd Qu.:140.02	3rd Qu.:0.007365	3rd Qu.:6.000e-05	3rd Qu.:0.003835
Max. :260.11	Max. :592.0	Max. :239.17	Max. :0.033160	Max. :2.600e-04	Max. :0.021440

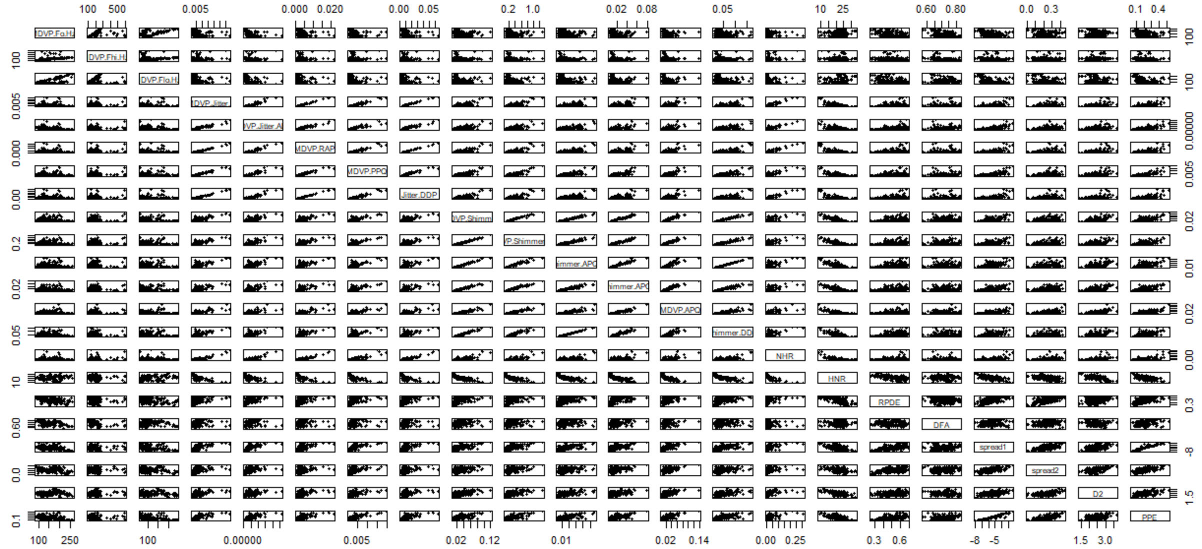
MDVP.PPQ	Jitter.DDP	MDVP.Shimmer	MDVP.Shimmer.dB.	Shimmer.APQ3	Shimmer.APQ5
Min. :0.000920	Min. :0.002040	Min. :0.00954	Min. :0.0850	Min. :0.004550	Min. :0.00570
1st Qu.:0.001860	1st Qu.:0.004985	1st Qu.:0.01650	1st Qu.:0.1485	1st Qu.:0.008245	1st Qu.:0.00958
Median :0.002690	Median :0.007490	Median :0.02297	Median :0.2210	Median :0.012790	Median :0.01347
Mean :0.003446	Mean :0.009920	Mean :0.02971	Mean :0.2823	Mean :0.015664	Mean :0.01788
3rd Qu.:0.003955	3rd Qu.:0.011505	3rd Qu.:0.03789	3rd Qu.:0.3500	3rd Qu.:0.020265	3rd Qu.:0.02238
Max. :0.019580	Max. :0.064330	Max. :0.11908	Max. :1.3020	Max. :0.056470	Max. :0.07940

MDVP.APQ	Shimmer.DDA	NHR	HNR	RPDE	DFA
Min. :0.00719	Min. :0.01364	Min. :0.000650	Min. : 8.441	Min. :0.2566	Min. :0.5743

1st Qu.:0.01308	1st Qu.:0.02474	1st Qu.:0.005925	1st Qu.:19.198	1st Qu.:0.4213	1st Qu.:0.6748
Median :0.01826	Median :0.03836	Median :0.011660	Median :22.085	Median :0.4960	Median :0.7223
Mean :0.02408	Mean :0.04699	Mean :0.024847	Mean :21.886	Mean :0.4985	Mean :0.7181
3rd Qu.:0.02940	3rd Qu.:0.06080	3rd Qu.:0.025640	3rd Qu.:25.076	3rd Qu.:0.5876	3rd Qu.:0.7619
Max. :0.13778	Max. :0.16942	Max. :0.314820	Max. :33.047	Max. :0.6852	Max. :0.8253

spread1	spread2	D2	PPE
Min. :-7.965	Min. :0.006274	Min. :1.423	Min. :0.04454
1st Qu.:-6.450	1st Qu.:0.174350	1st Qu.:2.099	1st Qu.:0.13745
Median :-5.721	Median :0.218885	Median :2.362	Median :0.19405
Mean :-5.684	Mean :0.226510	Mean :2.382	Mean :0.20655
3rd Qu.:-5.046	3rd Qu.:0.279234	3rd Qu.:2.636	3rd Qu.:0.25298
Max. :-2.434	Max. :0.450493	Max. :3.671	Max. :0.52737

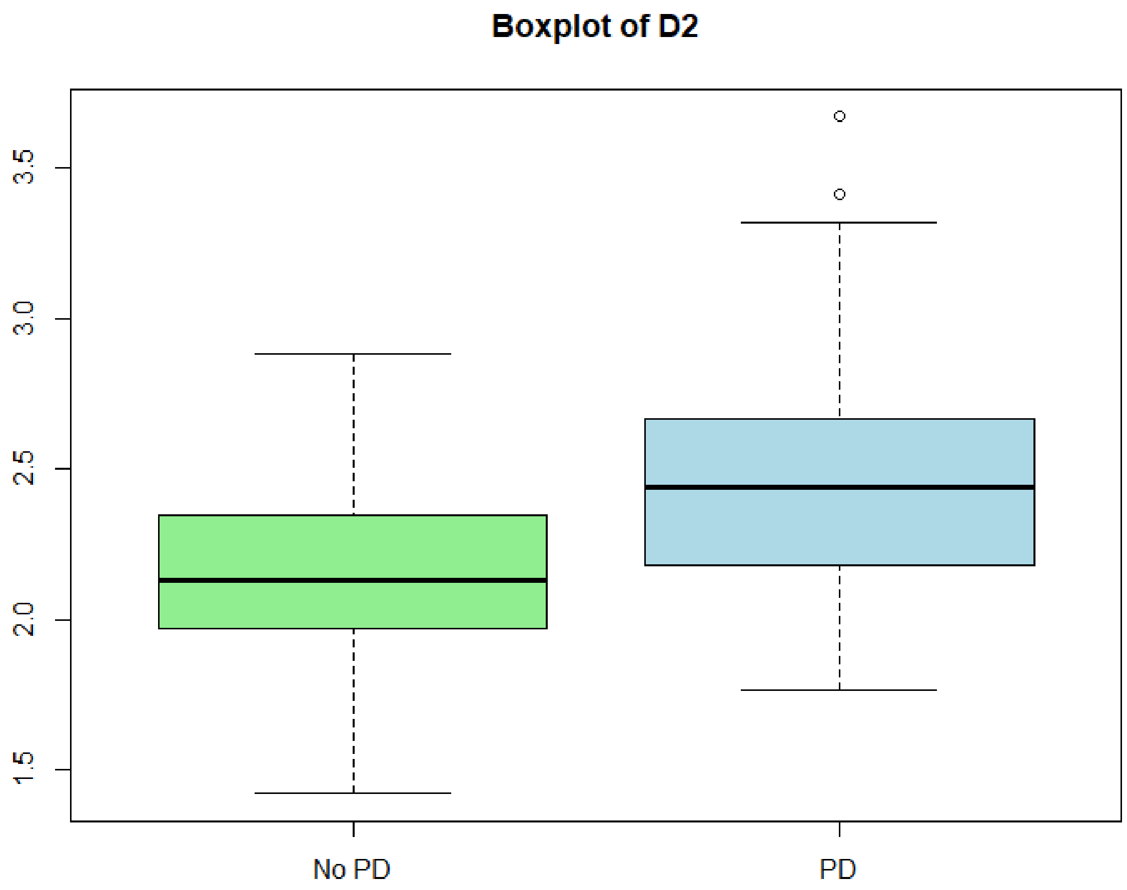
Whole dataset Scatterplots

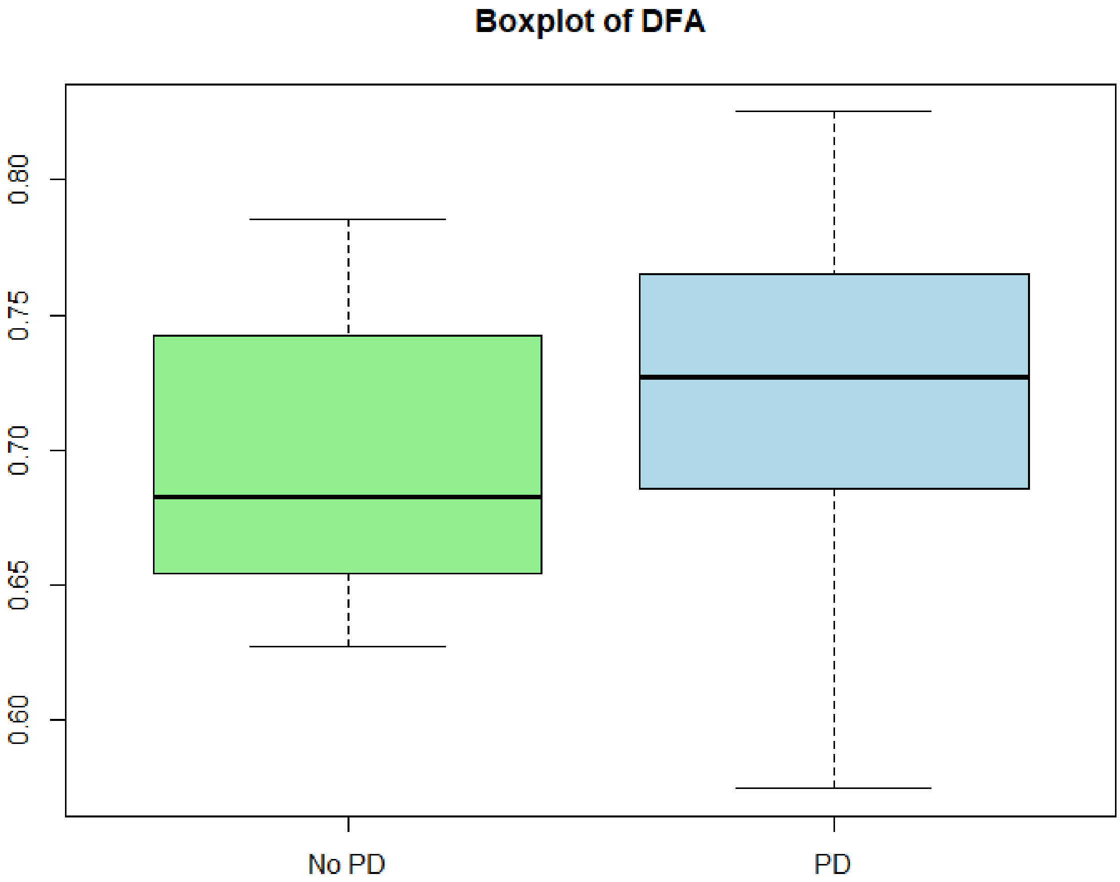


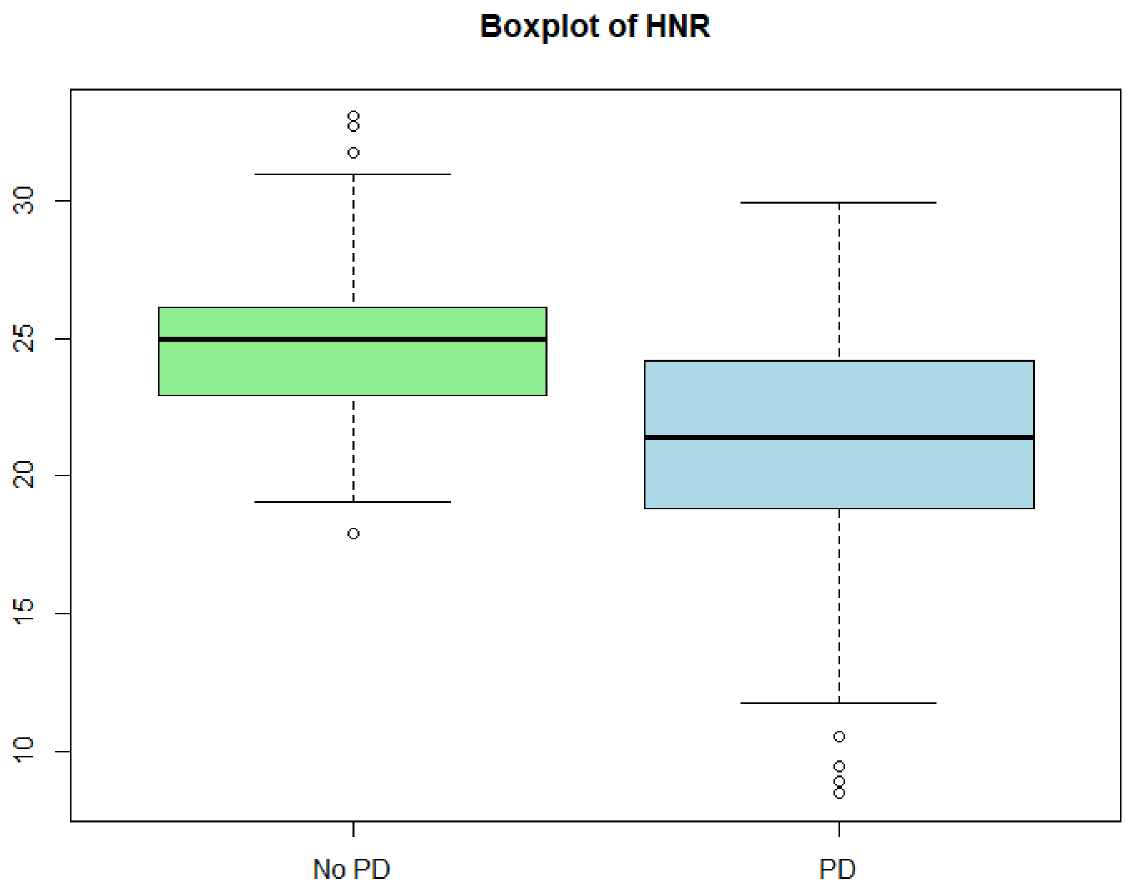
Target data descriptive stats

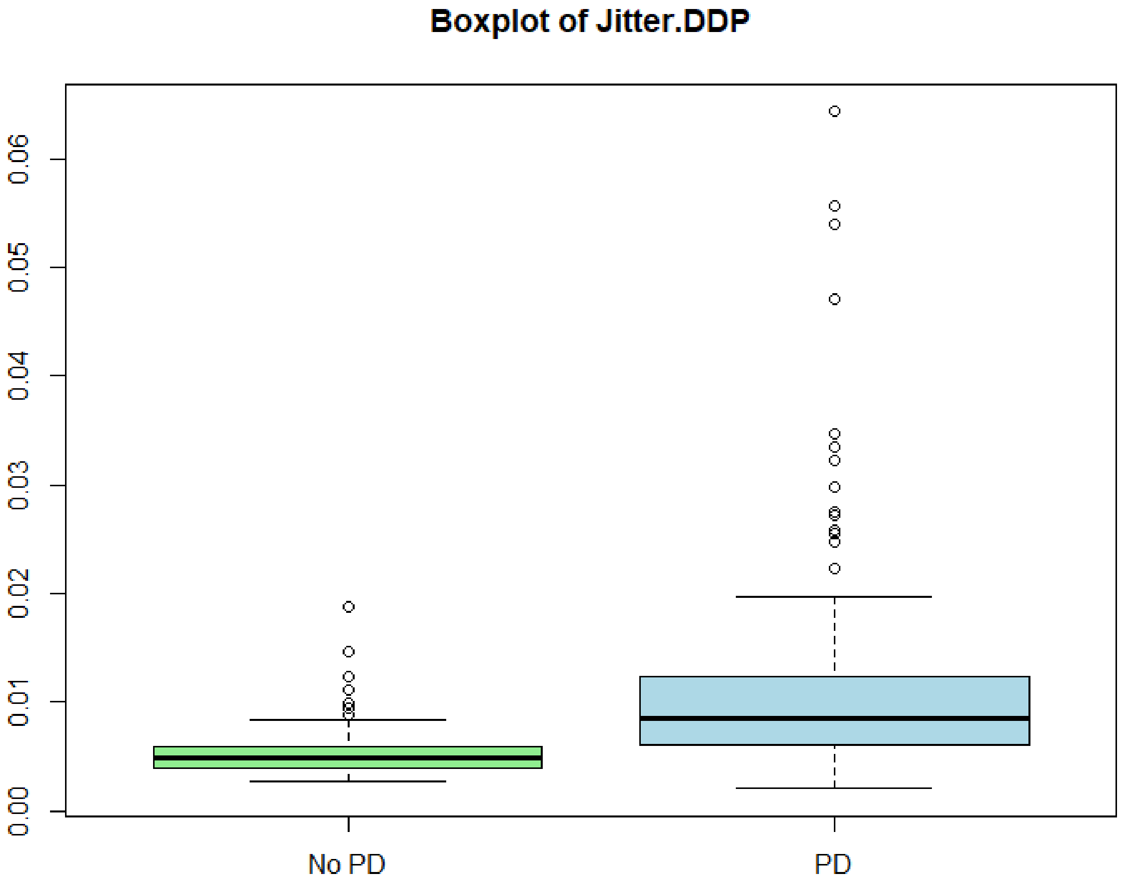
	No PD	Has PD
Number of cases	48	147

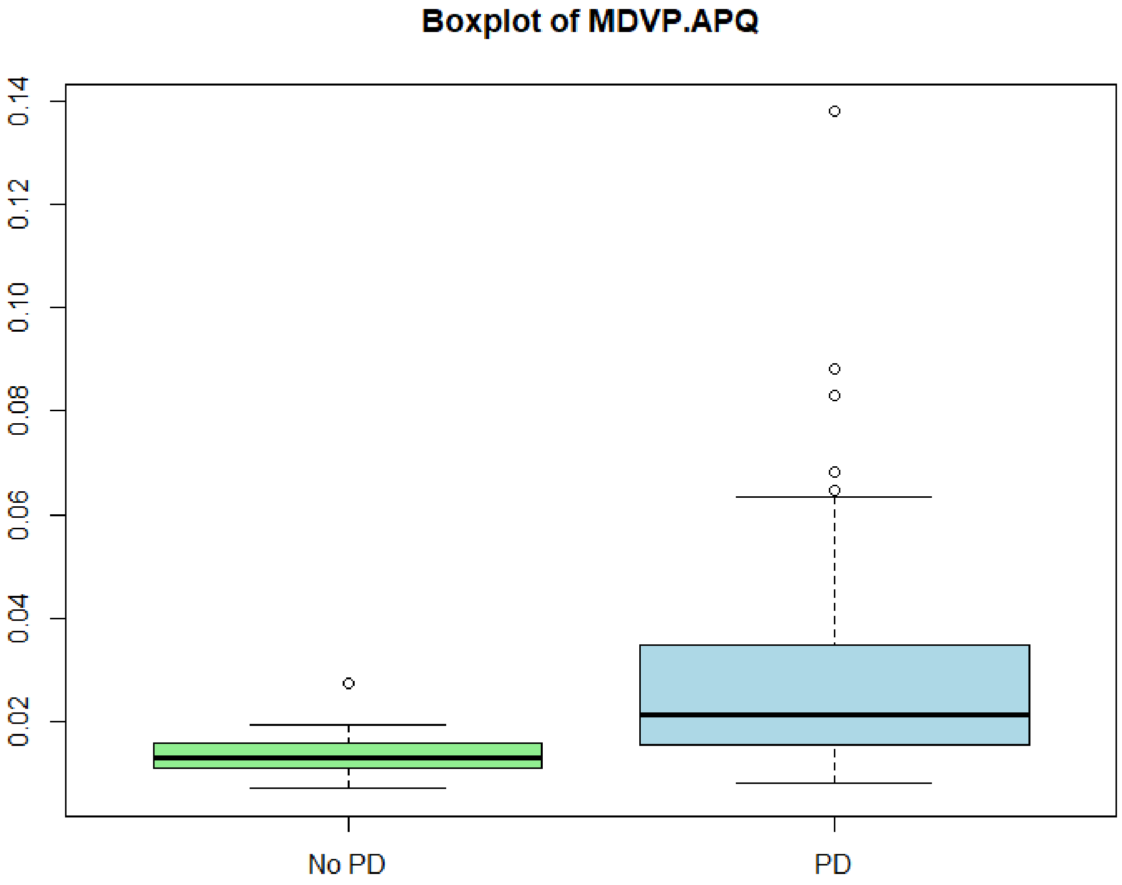
Boxplots of each numerical variable comparing PD and No PD groups distributions

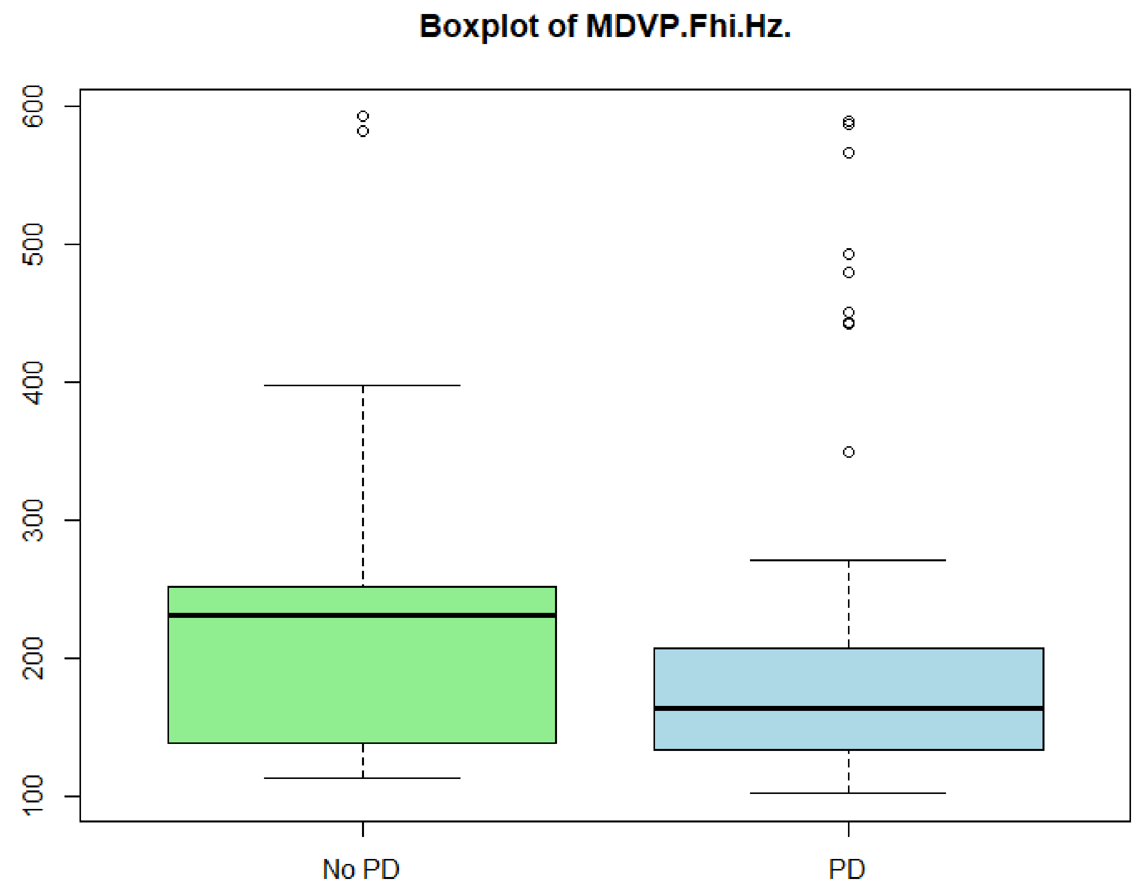

Boxplot of D2
3.5
3.0
2.5
2.0
1.5
No PD
PD

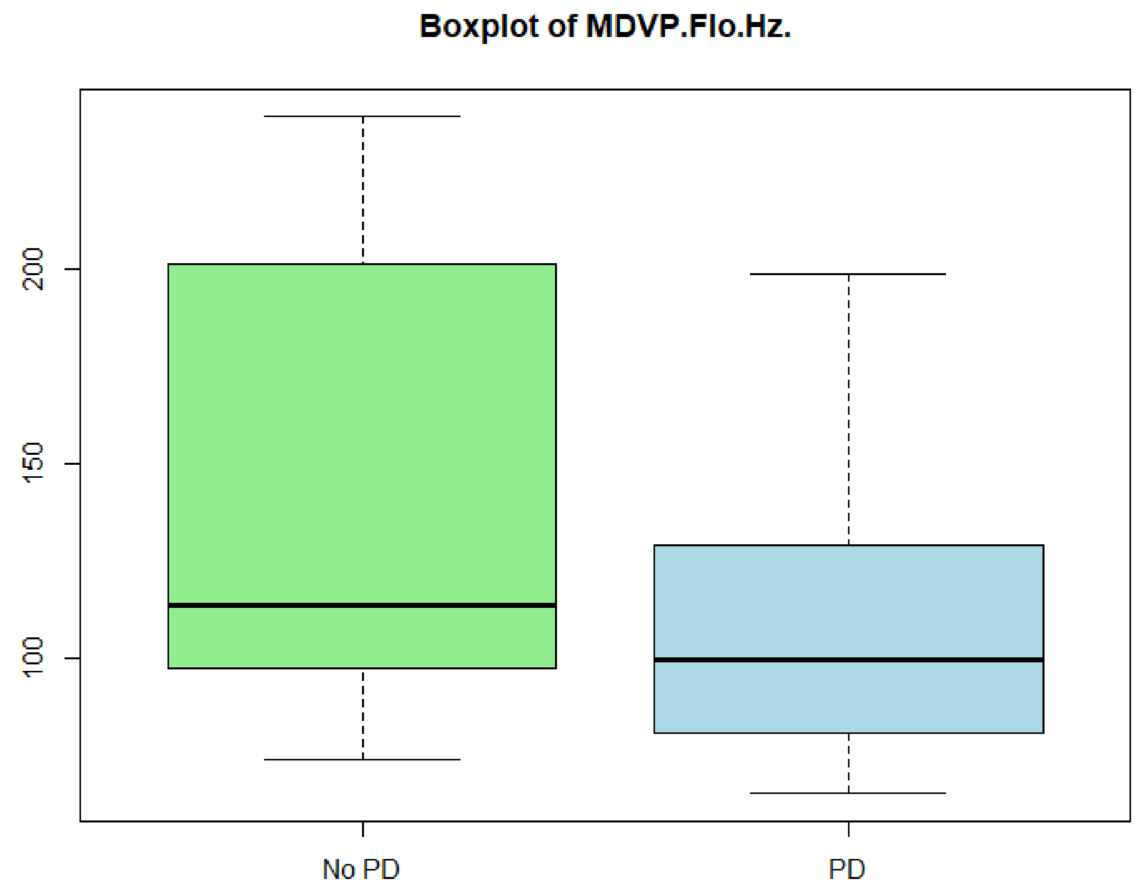

Boxplot of DFA
0.60
0.65
0.70
0.75
0.80
No PD
PD

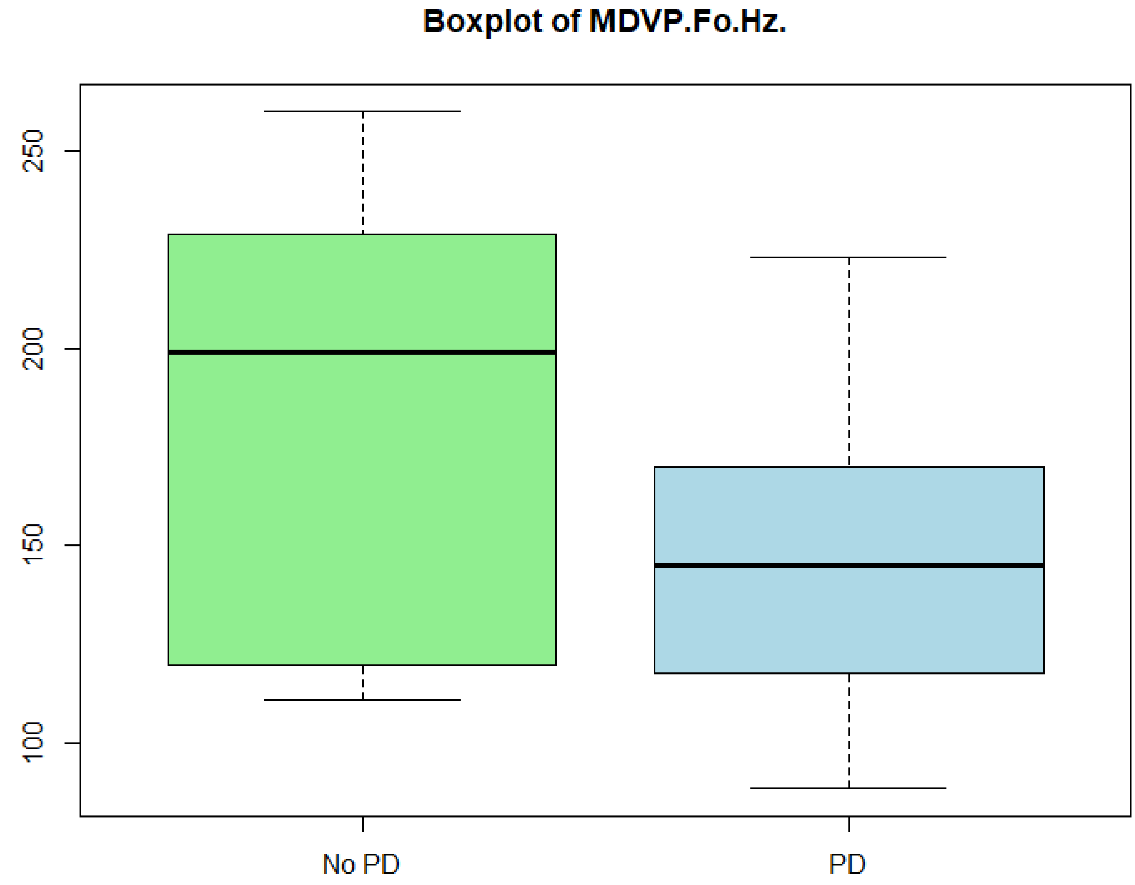

Boxplot of HNR
30
25
20
15
10
No PD
PD


Boxplot of Jitter.DDP
0.00
0.01
0.02
0.03
0.04
0.05
0.06
No PD
PD

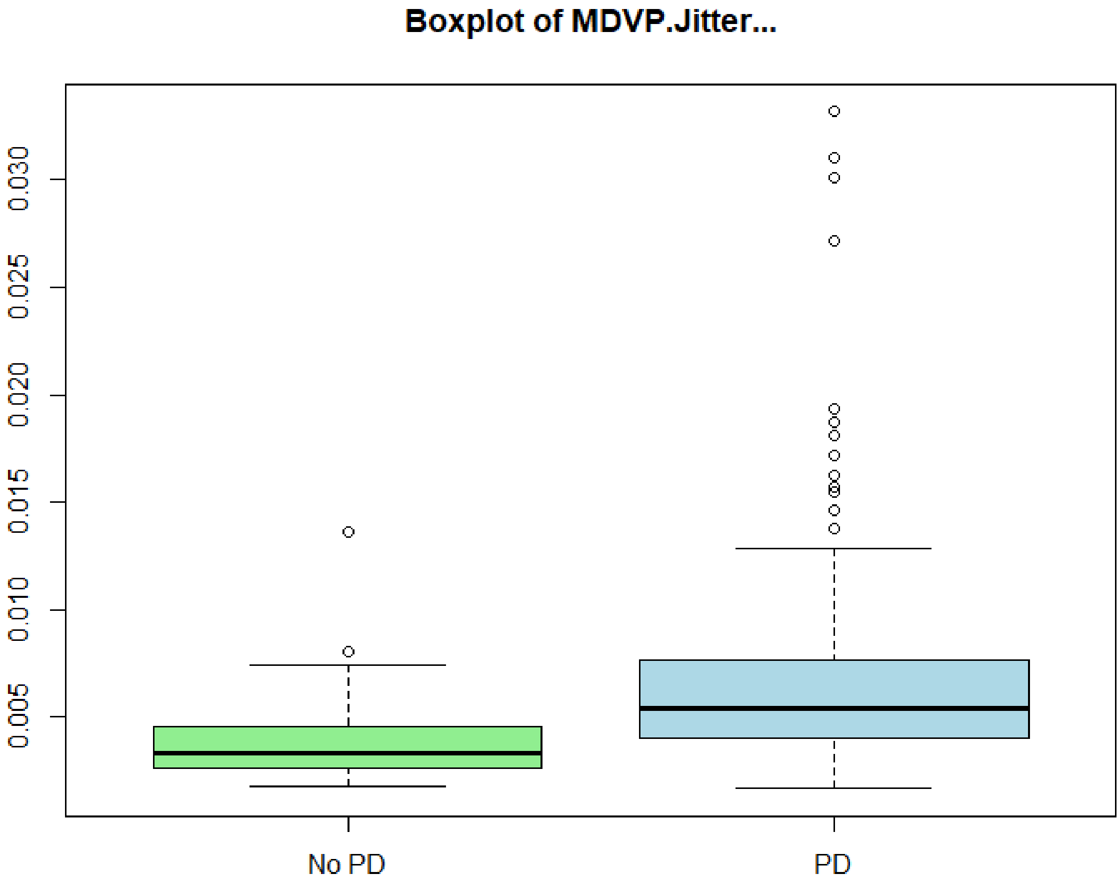

Boxplot of MDVP.APQ
0.14
0.12
0.10
0.08
0.06
0.04
0.02
No PD
PD

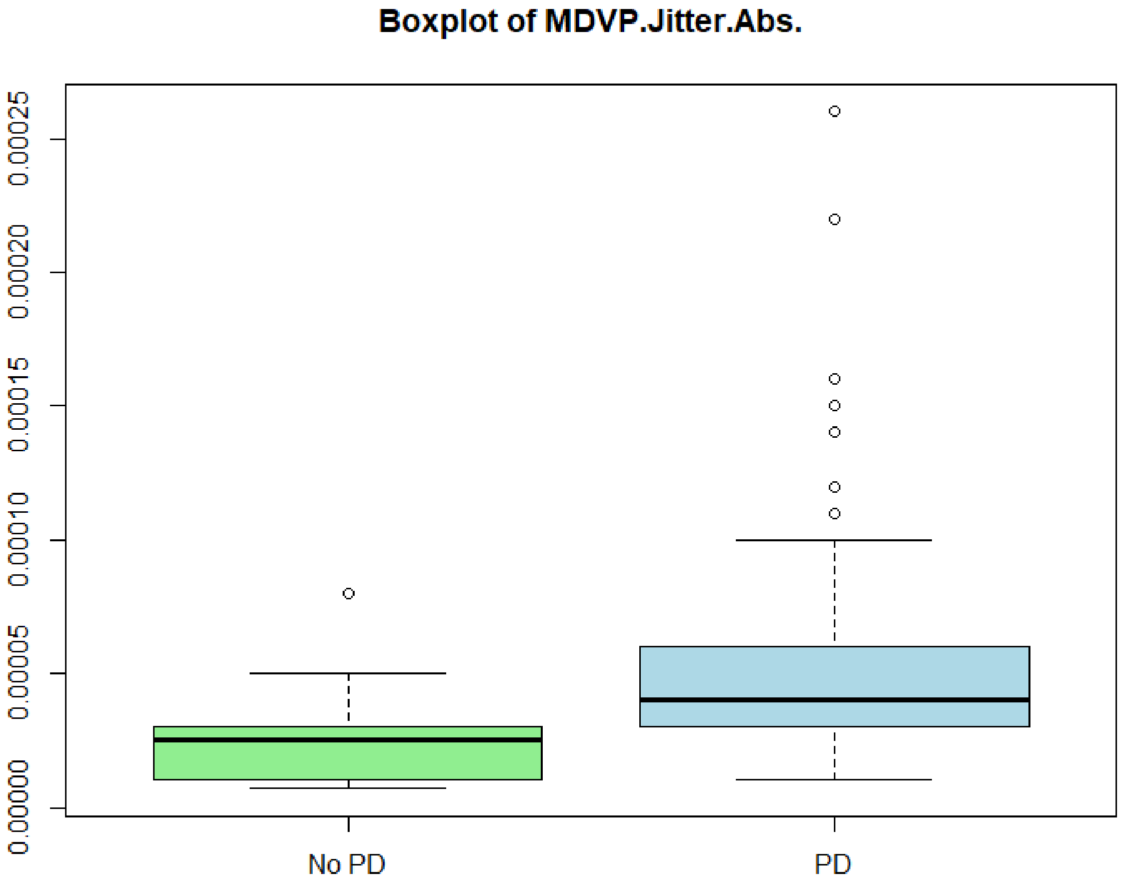

Boxplot of MDVP.Fhi.Hz.
100
200
300
400
500
600
No PD
PD

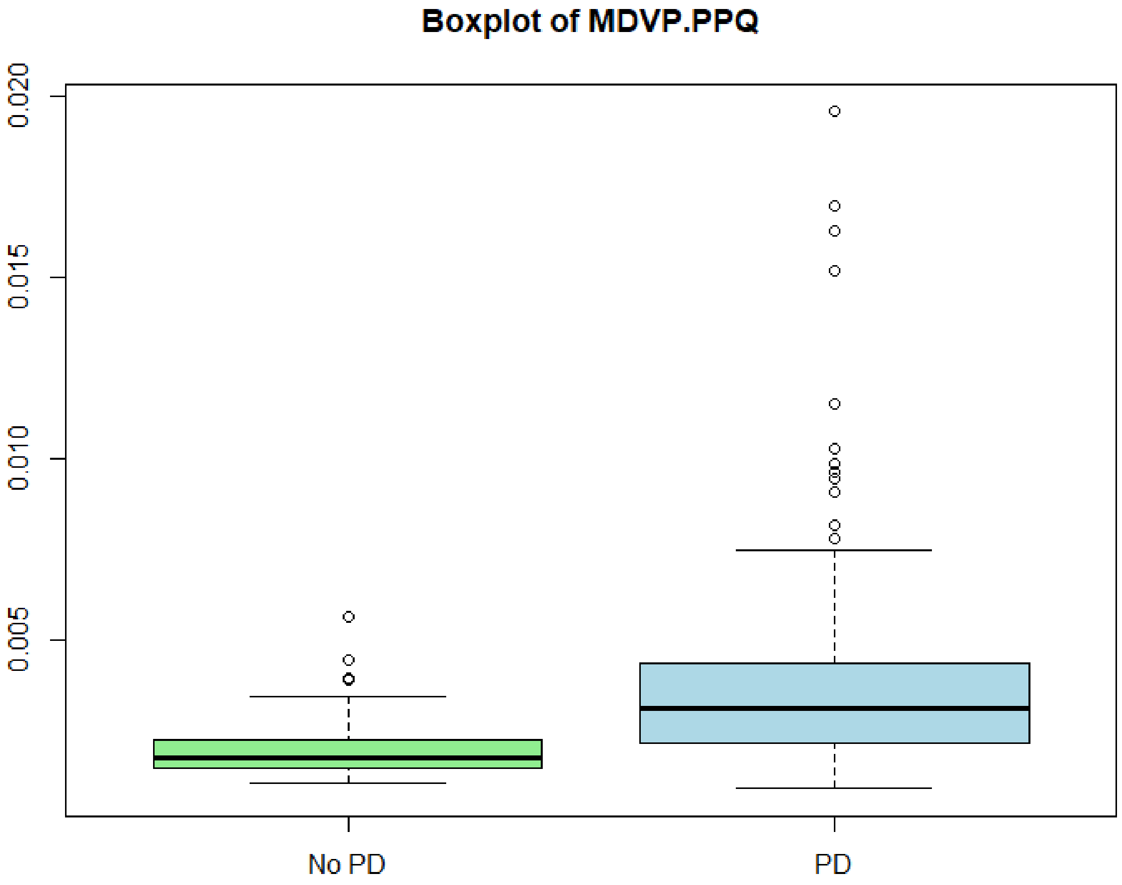

Boxplot of MDVP.Flo.Hz.
200
150
100
No PD
PD

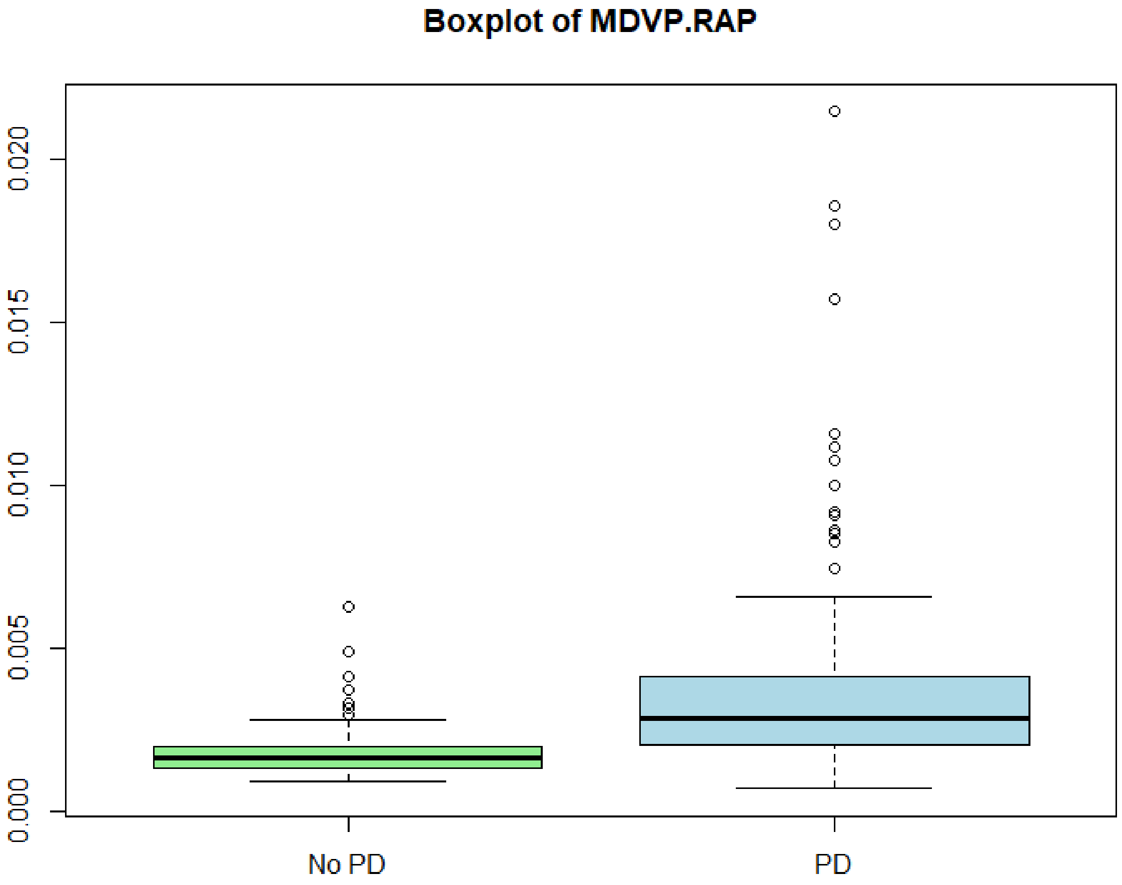


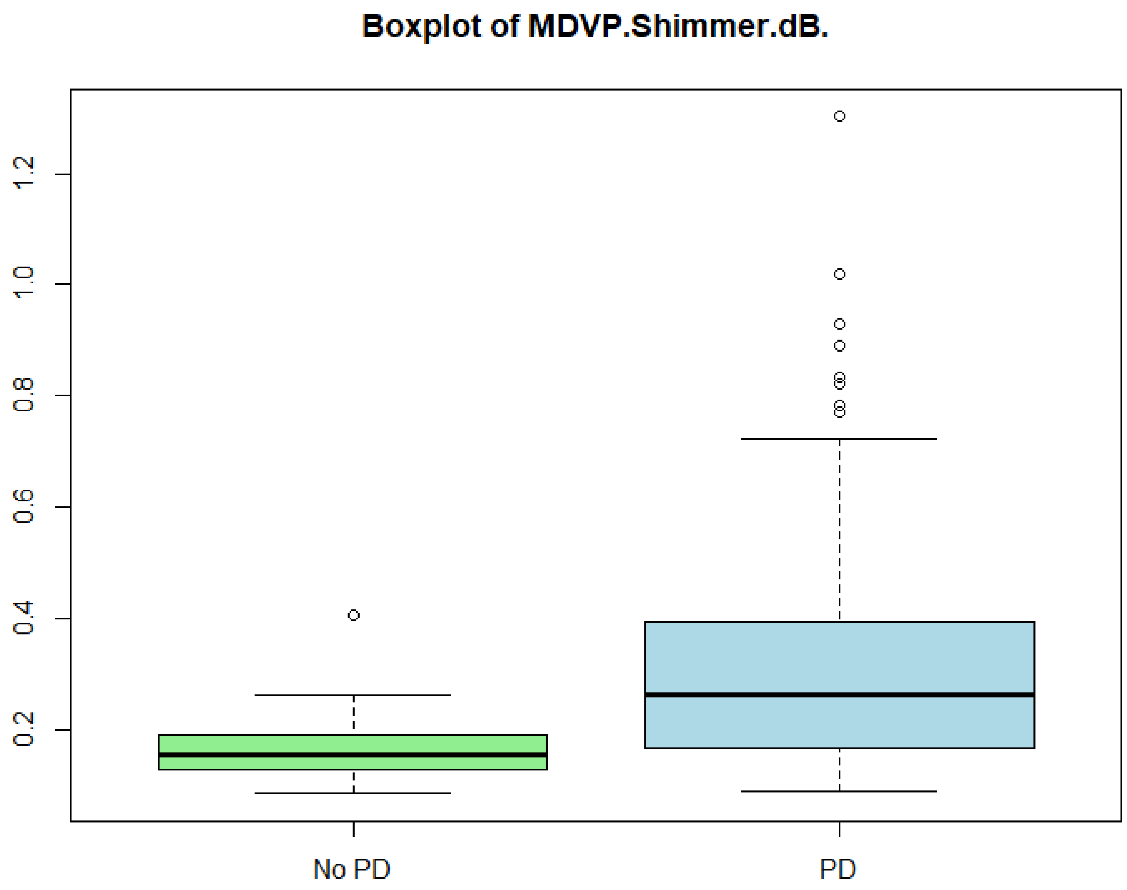
Boxplot of MDVP.Fo.Hz.

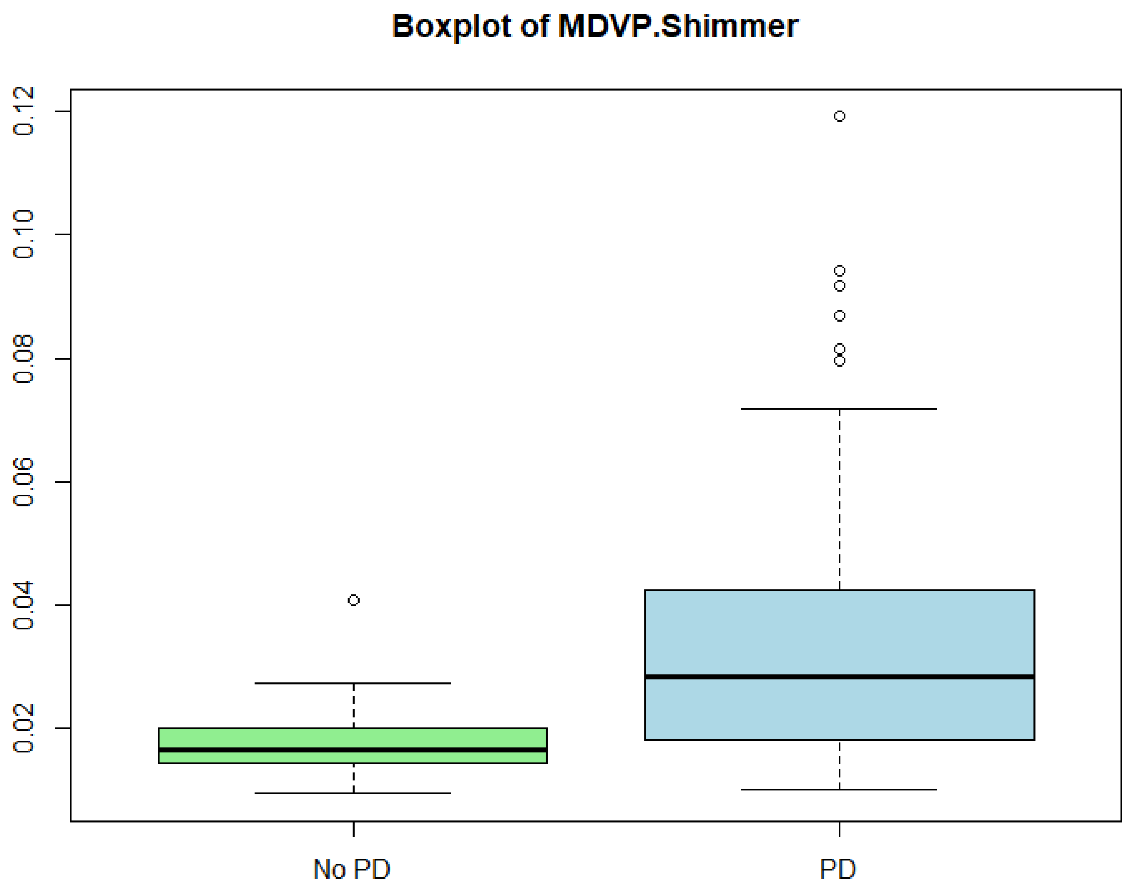

Boxplot of MDVP.Jitter...
0.005
0.010
0.015
0.020
0.025
0.030
No PD
PD

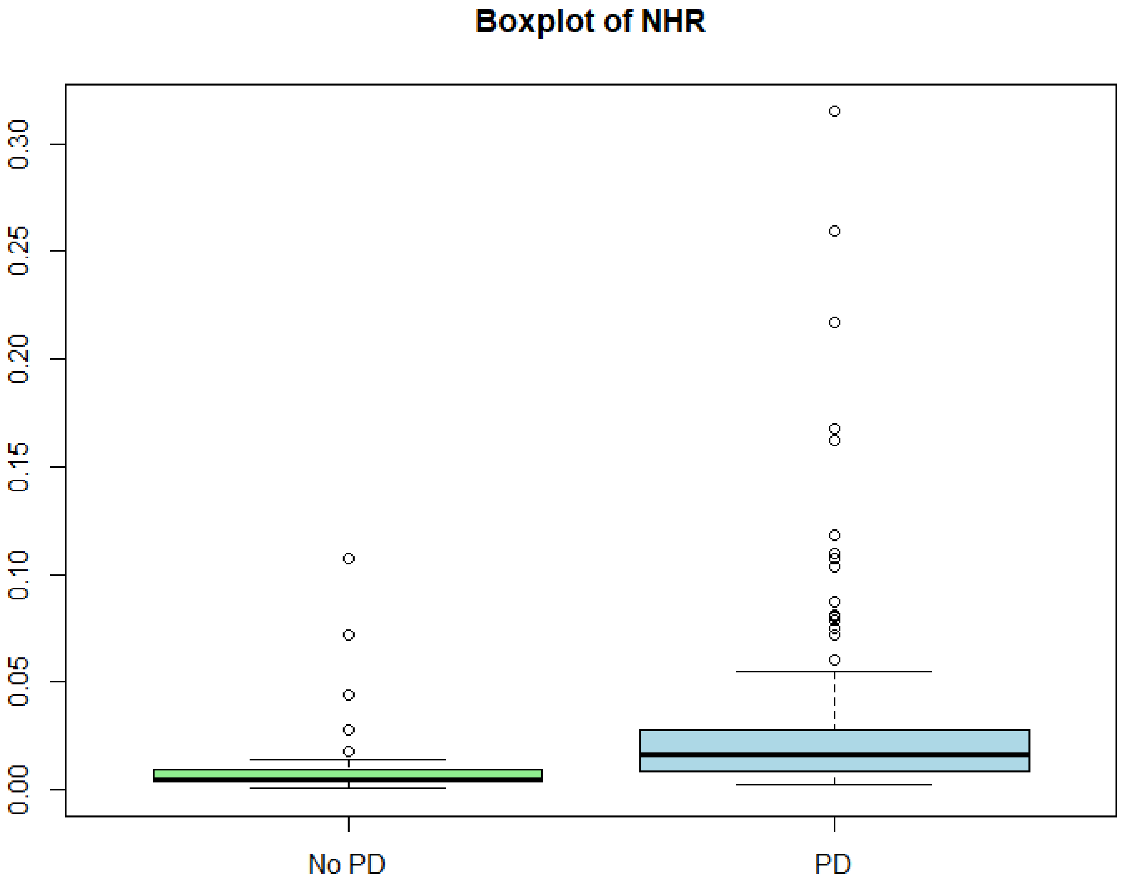

Boxplot of MDVP.Jitter.Abs.
0.00000
0.00005
0.00010
0.00015
0.00020
0.00025
No PD
PD

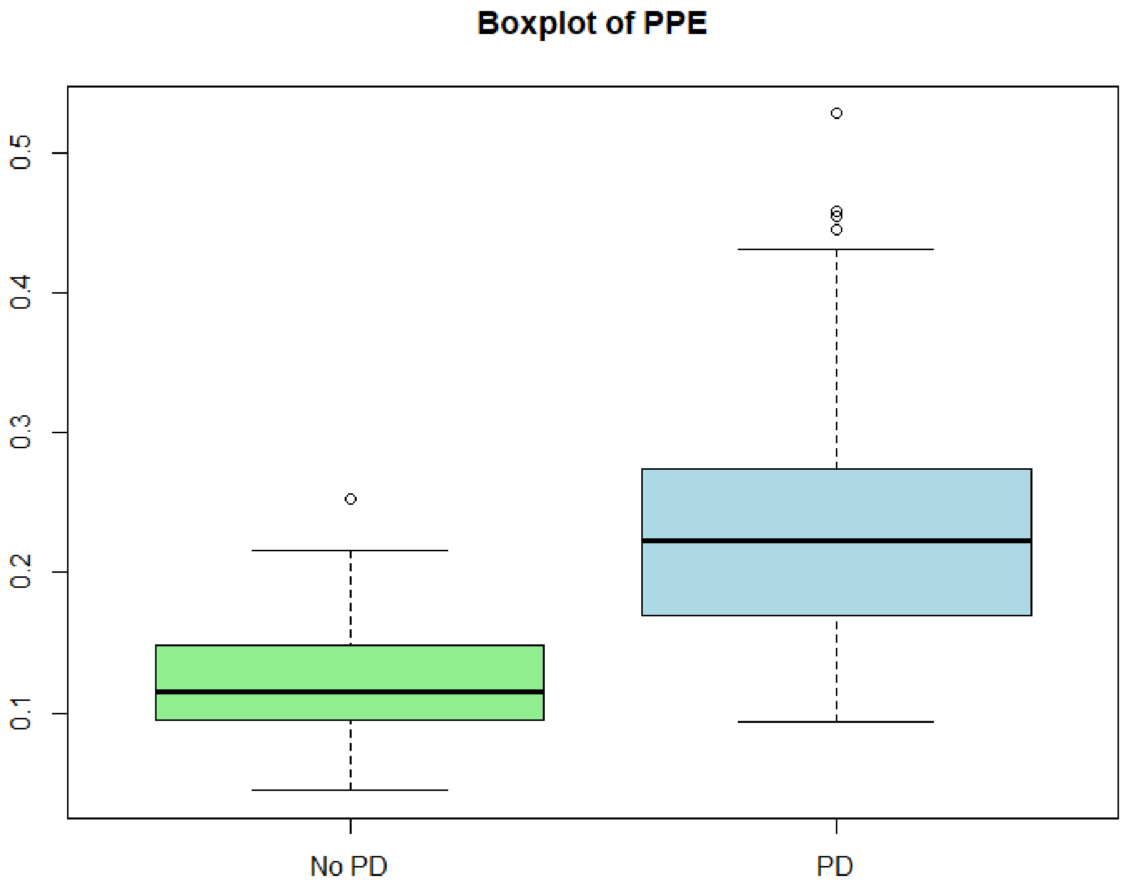

Boxplot of MDVP.PPQ
0.020
0.015
0.010
0.005
No PD
PD

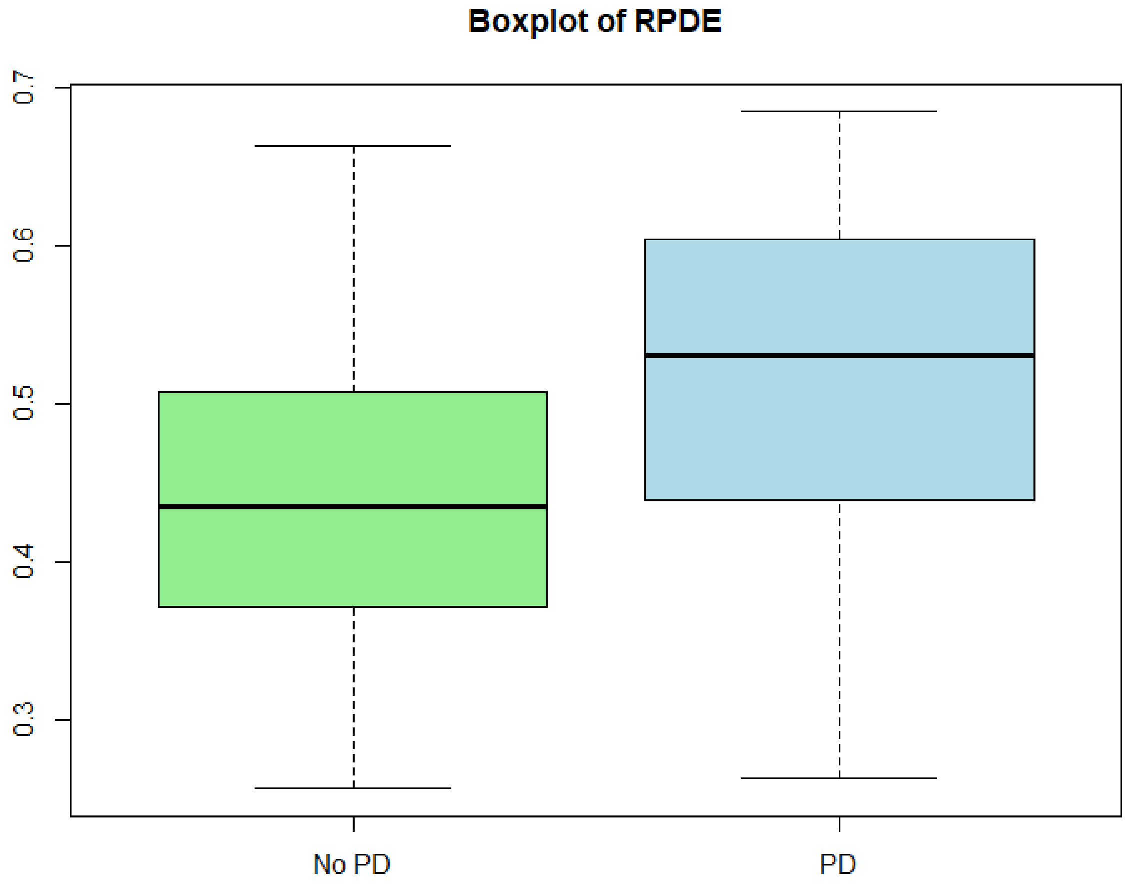

Boxplot of MDVP.RAP
0.000
0.005
0.010
0.015
0.020
No PD
PD

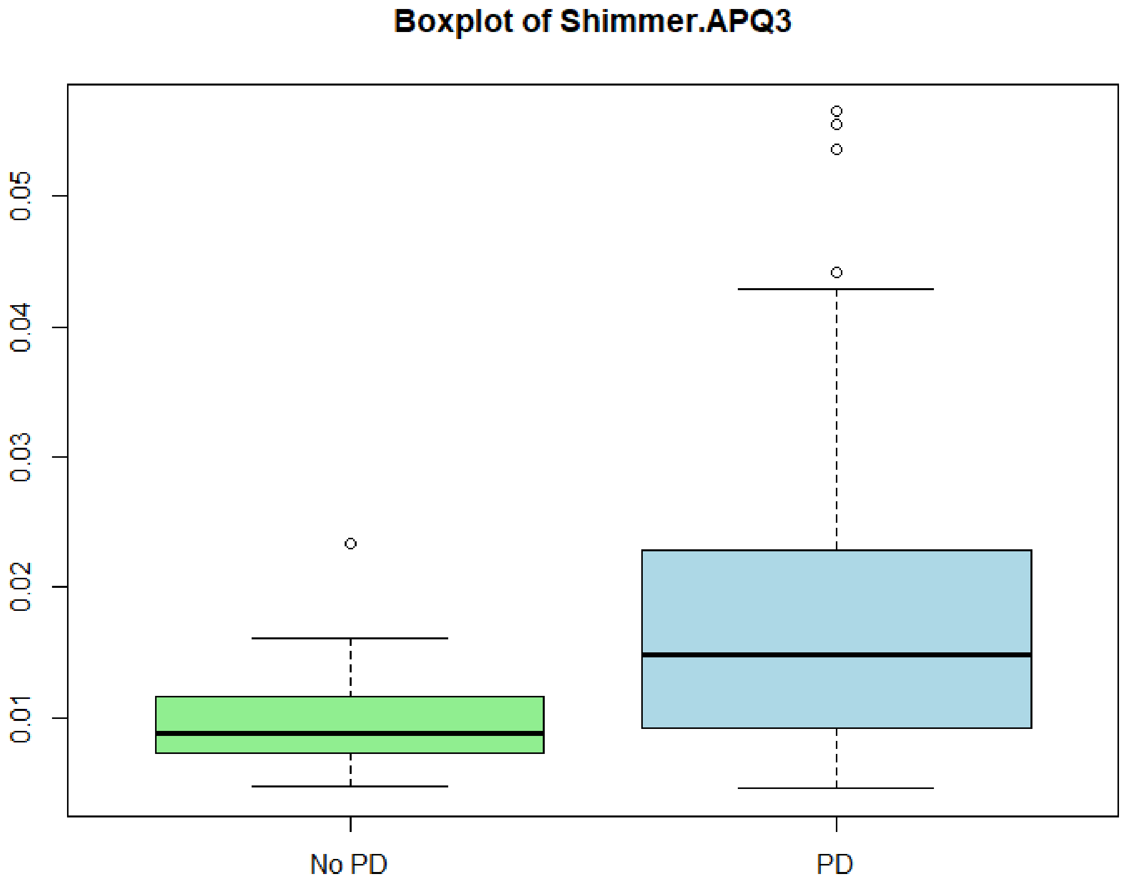

Boxplot of MDVP.Shimmer.dB.
1.2
1.0
0.8
0.6
0.4
0.2
No PD
PD

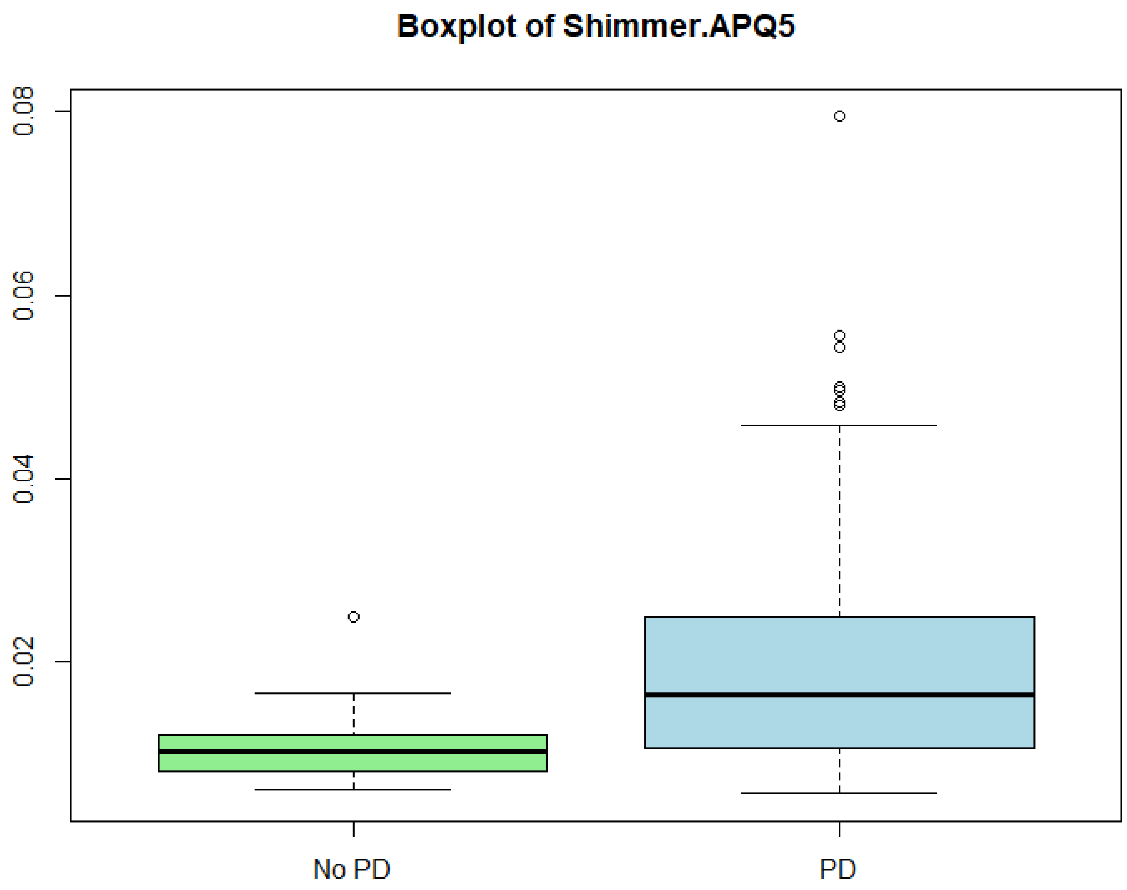

Boxplot of MDVP.Shimmer
0.02
0.04
0.06
0.08
0.10
0.12
No PD
PD

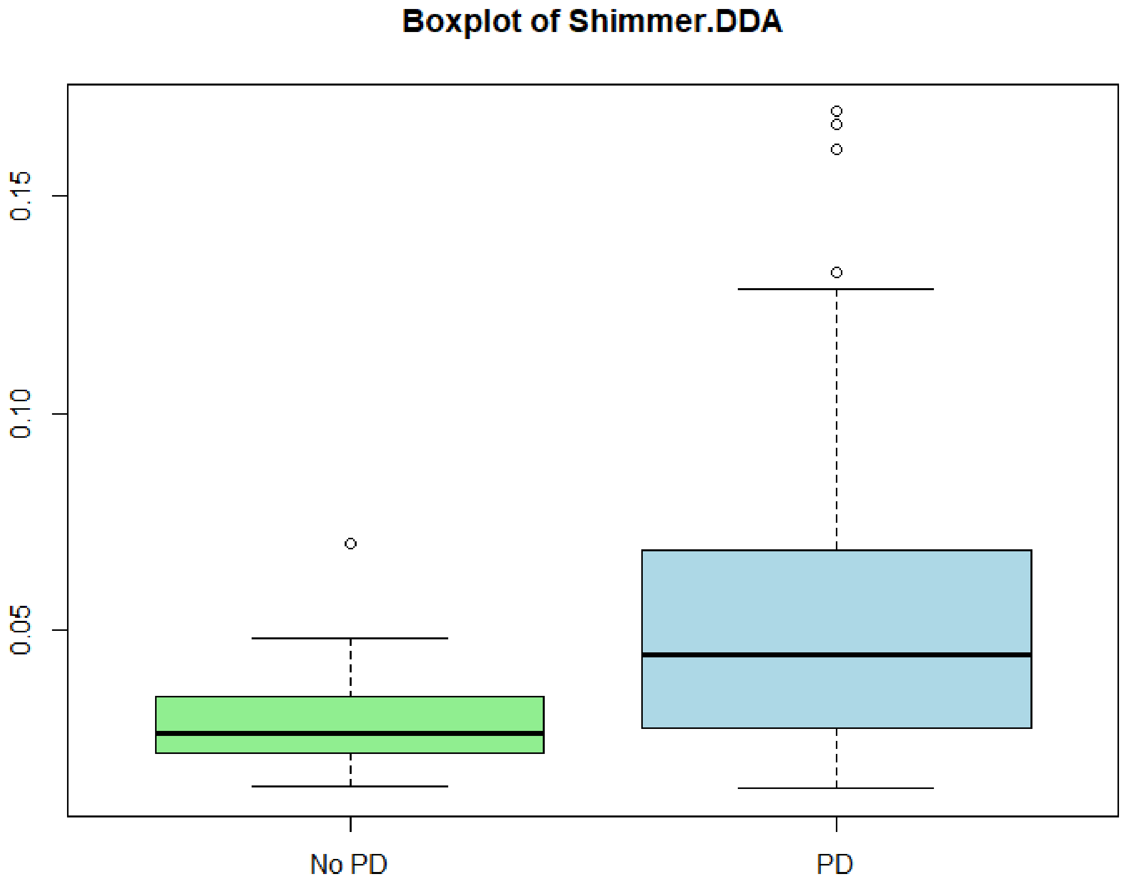

Boxplot of NHR
0.00
0.05
0.10
0.15
0.20
0.25
0.30
No PD
PD

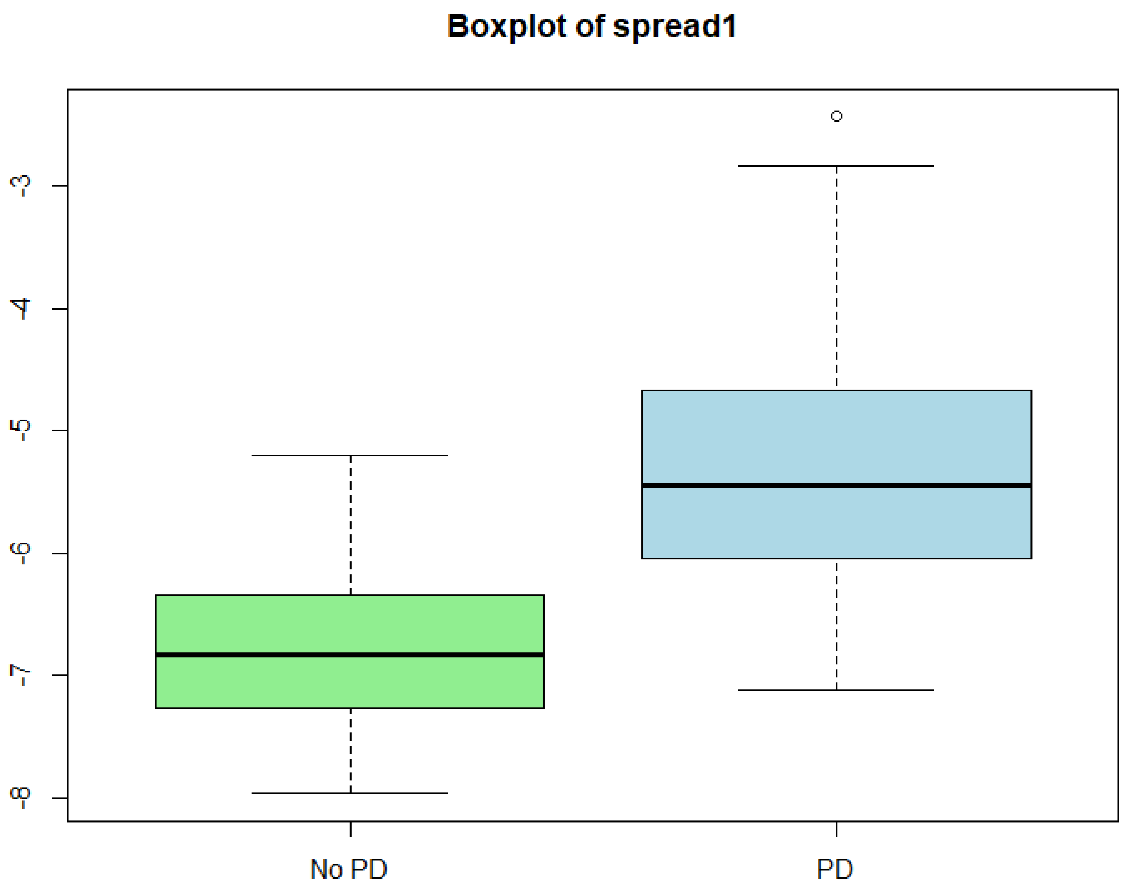

Boxplot of PPE
0.5
0.4
0.3
0.2
0.1
No PD
PD

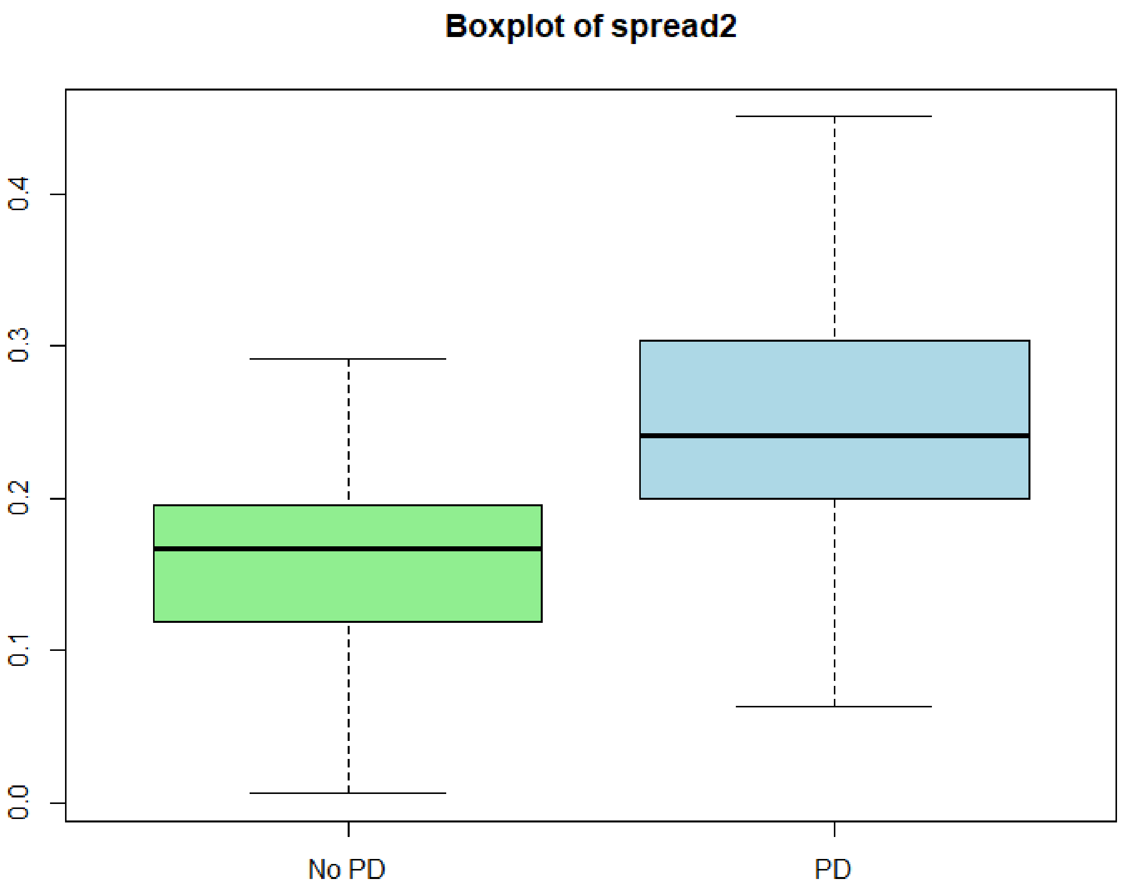

Boxplot of RPDE
0.3
0.4
0.5
0.6
0.7
No PD
PD


Boxplot of Shimmer.APQ3
0.01
0.02
0.03
0.04
0.05
No PD
PD


Boxplot of Shimmer.APQ5
0.08
0.06
0.04
0.02
No PD
PD


Boxplot of Shimmer.DDA
0.05
0.10
0.15
No PD
PD


Boxplot of spread1
-3
-4
-5
-6
-7
-8
No PD
PD



Support Vector Classifier with cost = 0.001, no scaling

Train Confusion Matrix: 0.97 sens

	Actual No PD	Actual PD
Predicted No PD	15	3
Predicted PD	21	111

Test Confusion Matrix: 0.93 sens

	Actual No PD	Actual PD
Predicted No PD	6	2
Predicted PD	6	31

Support Vector Classifier with cost = 0.001, scale = True

Train Confusion Matrix: 1 sens

	Actual No PD	Actual PD
Predicted No PD	0	0
Predicted PD	36	114

Test Confusion Matrix: 1 sens

	Actual No PD	Actual PD
Predicted No PD	0	0
Predicted PD	12	33

SVC with cost = 100, no scaling

Training: 0.94 sens

	Actual No PD	Actual PD
Predicted No PD	23	6
Predicted PD	13	108

Testing: 1 sens

	Actual No PD	Actual PD
Predicted No PD	6	0

Predicted PD	6	33

SVC, cost = 1e5, scale = False

Train: 0.90 sens

	Actual No PD	Actual PD
Predicted No PD	22	11
Predicted PD	14	103

Test: 0.84 sens

	Actual No PD	Actual PD
Predicted No PD	7	5
Predicted PD	5	28

SVC, cost = 1e-06, scale = False

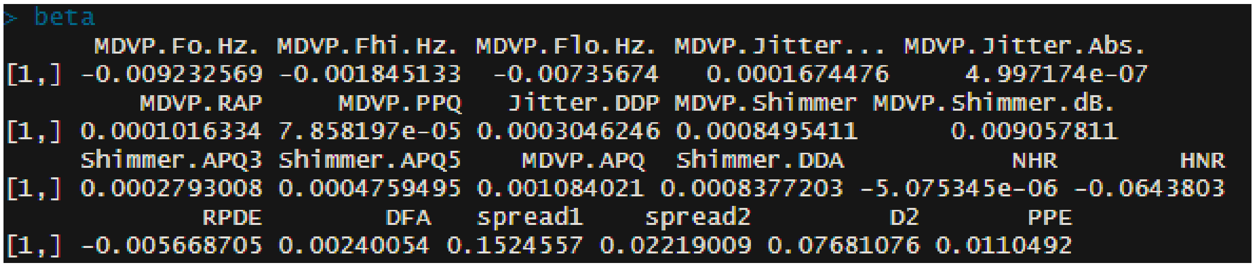
Train: 1 sens

	Actual No PD	Actual PD
Predicted No PD	0	0
Predicted PD	36	114

Test: 1 sens

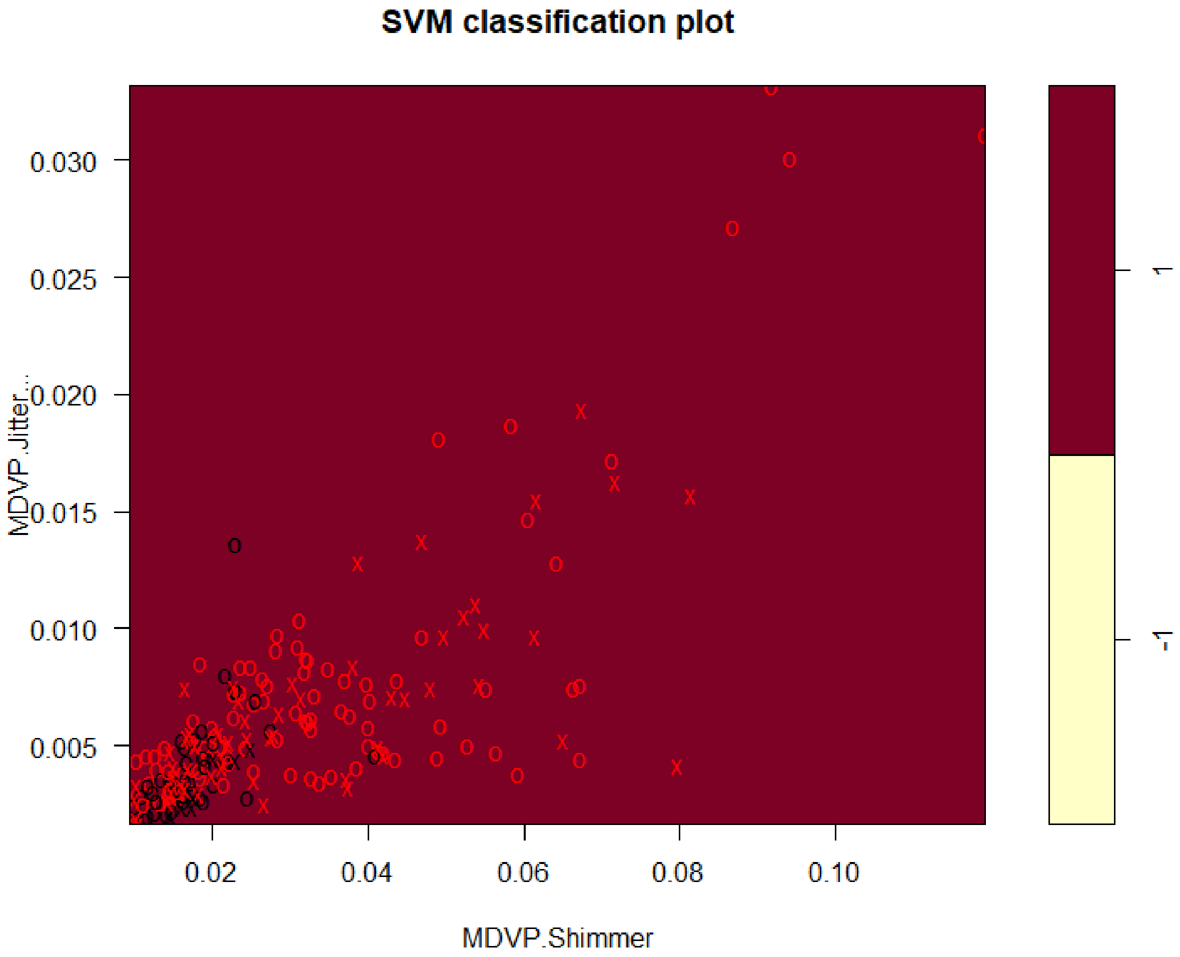
	Actual No PD	Actual PD
Predicted No PD	0	0
Predicted PD	12	33

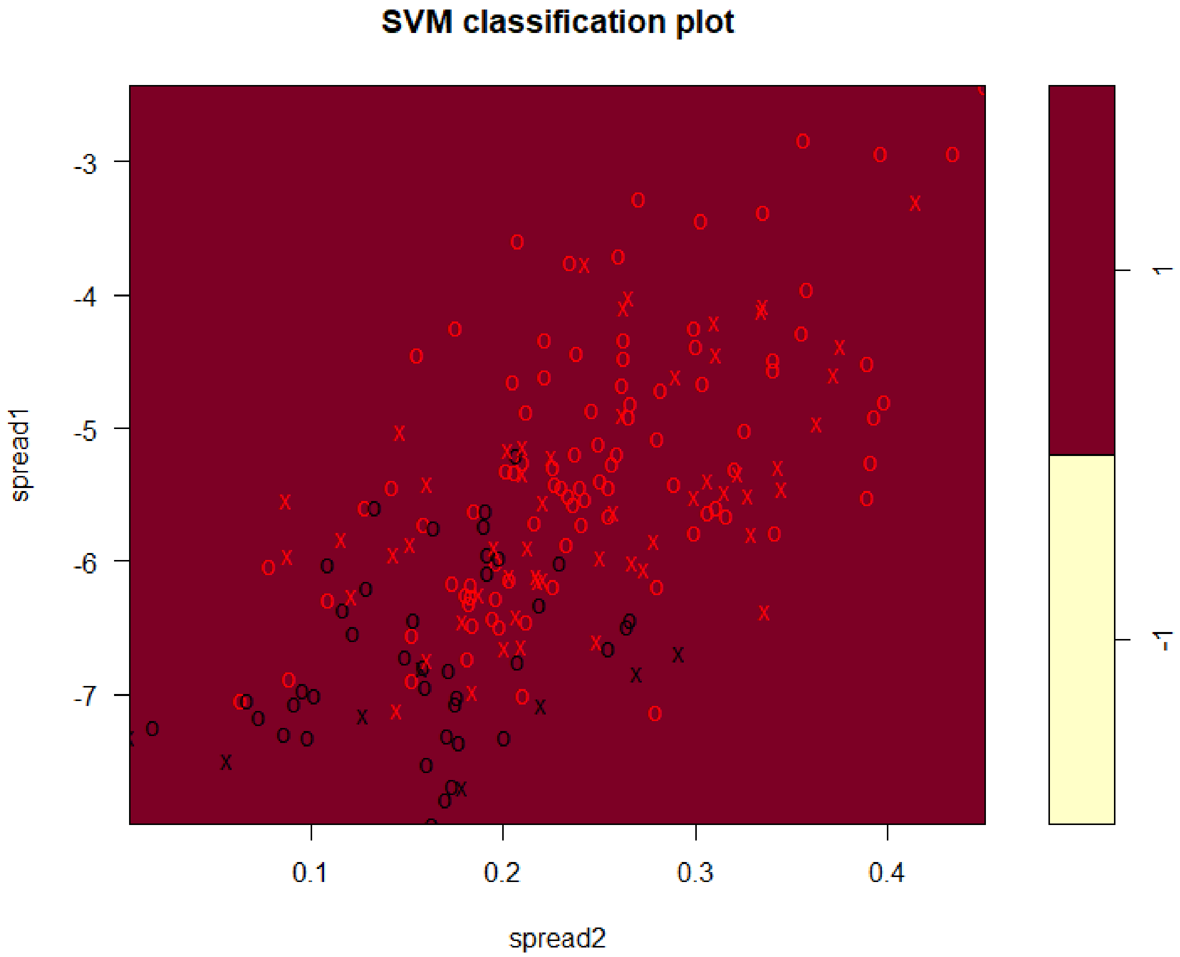
Support Vector Classifier Cross-Validation Best Model Beta Vector



```
> beta
      MDVP.Fo.Hz. MDVP.Fhi.Hz. MDVP.Flo.Hz. MDVP.Jitter... MDVP.Jitter.Abs.
[1,] -0.009232569 -0.001845133  -0.00735674   0.0001674476     4.997174e-07
         MDVP.RAP     MDVP.PPQ   Jitter.DDP MDVP.Shimmer MDVP.Shimmer.dB.
[1,] 0.0001016334 7.858197e-05 0.0003046246 0.0008495411      0.009057811
     Shimmer.APQ3 Shimmer.APQ5    MDVP.APQ  Shimmer.DDA           NHR        HNR
[1,] 0.0002793008 0.0004759495 0.001084021 0.0008377203 -5.075345e-06 -0.0643803
             RPDE        DFA   spread1    spread2         D2       PPE
[1,] -0.005668705 0.00240054 0.1524557 0.02219009 0.07681076 0.0110492
```

Examples of Decision Boundaries by Best Model


SVM classification plot
MDVP.Jitter...
0.030
0.025
0.020
0.015
0.010
0.005
MDVP.Shimmer
0.02
0.04
0.06
0.08
0.10
1
-1



Polynomial Kernel Cross Validation Best Model Matrices and ROC curves

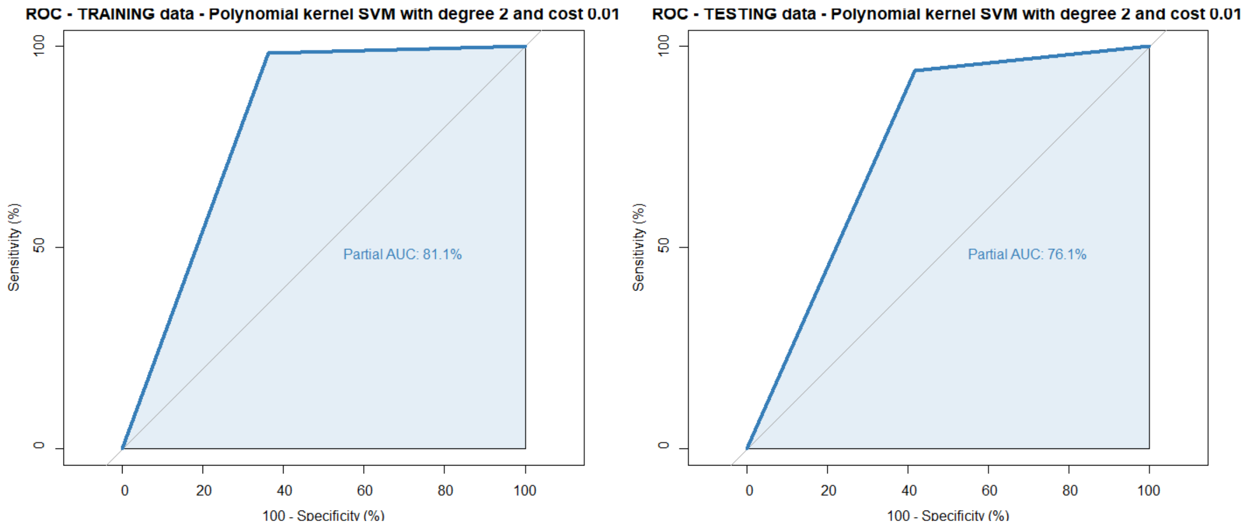
Train: 0.98 sens

	Actual No PD	Actual PD
Predicted No PD	23	2
Predicted PD	13	112

Test: 0.93 sens

	Actual No PD	Actual PD
Predicted No PD	7	2

Predicted PD	5	31



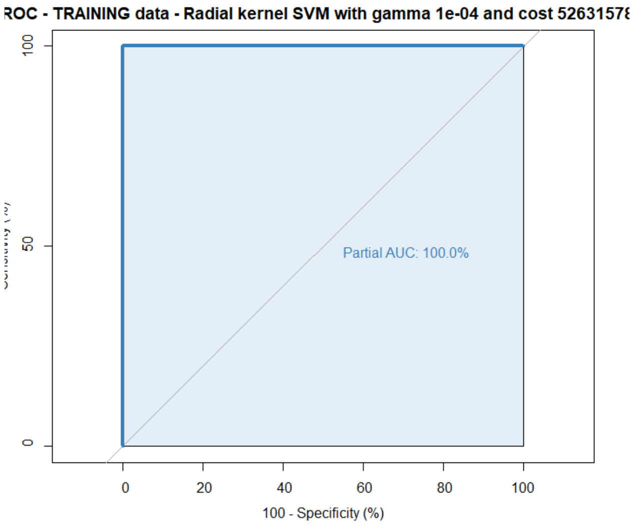
Radial Kernel Cross Validation Best Model Matrices and ROC curves

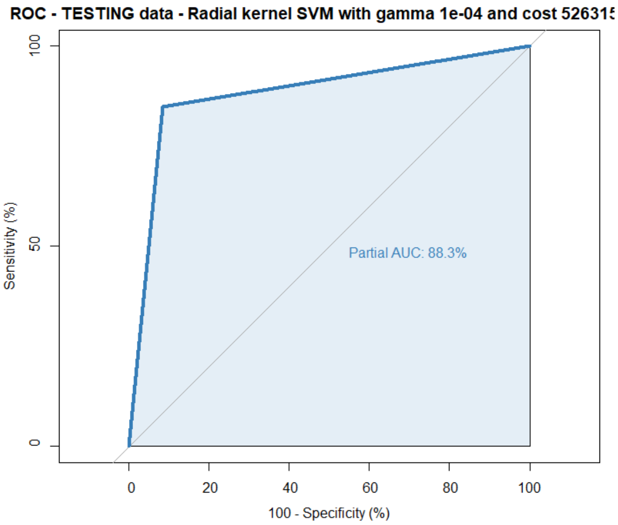
Train: 1 sens

	Actual No PD	Actual PD
Predicted No PD	36	0
Predicted PD	0	114

Test: 0.85 sens

	Actual No PD	Actual PD
Predicted No PD	11	5
Predicted PD	1	28


ROC - TRAINING data - Radial kernel SVM with gamma 1e-04 and cost 52631578
Partial AUC: 100.0%
100
50
0
0
20
40
60
80
100
100 - Specificity (%)


ROC - TESTING data - Radial kernel SVM with gamma 1e-04 and cost 526315
Sensitivity (%)
Partial AUC: 88.3%
100
50
0
0
20
40
60
80
100
100 - Specificity (%)