Optimized machine learning methods for prediction of cognitive outcome in Parkinson's disease

Mohammad R. Salmanpour[a], Mojtaba Shamsaei[a], Abdollah Saberi[b], Saeed Setayeshi[a], Ivan S. Klyuzhin[c], Vesna Sossi[d], Arman Rahmim[e,d,f,*]

[a] *Department of Energy Engineering and Physics, Amirkabir University of Technology, Tehran, Iran*
[b] *Department of Computer Engineering, Islamic Azad University, Tehran, Iran*
[c] *Department of Medicine, University of British Columbia, Vancouver, BC, Canada*
[d] *Department of Physics & Astronomy, University of British Columbia, Vancouver, BC, Canada*
[e] *Department of Radiology, University of British Columbia, Vancouver, BC, Canada*
[f] *Department of Radiology, Johns Hopkins University, Baltimore, MD, USA*

ARTICLE INFO

ABSTRACT

Background: Given the increasing recognition of the significance of non-motor symptoms in Parkinson's disease, we investigate the optimal use of machine learning methods for the prediction of the Montreal Cognitive Assessment (MoCA) score at year 4 from longitudinal data obtained at years 0 and 1.

Methods: We selected n = 184 PD subjects from the Parkinson's Progressive Marker Initiative (PPMI) database (93 features). A range of robust predictor algorithms (accompanied with automated machine learning hyperparameter tuning) and feature subset selector algorithms (FSSAs) were selected. We utilized 65%, 5% and 30% of patients in each arrangement for training, training validation and final testing respectively (10 randomized arrangements). For further testing, we enrolled 308 additional patients.

Results: First, we employed 10 predictor algorithms, provided with all 93 features; an error of 1.83 ± 0.13 was obtained by LASSOLAR (Least Absolute Shrinkage and Selection Operator - Least Angle Regression). Subsequently, we used feature subset selection followed by predictor algorithms. GA (Genetic Algorithm) selected 18 features; subsequently LOLIMOT (Local Linear Model Trees) reached an error of 1.70 ± 0.10. DE (Differential evolution) also selected 18 features and coupled with Thiel-Sen regression arrived at a similar performance. NSGAII (Non-dominated sorting genetic algorithm) yielded the best performance: it selected six vital features, which combined with LOLIMOT reached an error of 1.68 ± 0.12. Finally, using this last approach on independent test data, we reached an error of 1.65.

Conclusion: By employing appropriate optimization tools (including automated hyperparameter tuning), it is possible to improve prediction of cognitive outcome. Overall, we conclude that optimal utilization of FSSAs and predictor algorithms can produce very good prediction of cognitive outcome in PD patients.

1. Introduction

Parkinson's disease (PD) is a chronic, progressive neurodegenerative disease for which there is currently no cure [1–3]. It is the second-most common neurodegenerative disorder after Alzheimer's disease, affecting 2–3% of the population over 65 years of age [4,5]. PD is characterized by loss of nigrostriatal dopaminergic neurons [6], resulting in a range of primarily motor symptoms (resting tremor, cogwheel rigidity, bradykinesia, postural instability and others). In addition, there is increasing recognition of the importance of other non-motor symptoms in PD [7]; a common non-motor symptom that can have a major

impact on the quality of life is cognitive decline [8], which is rather common in PD patients. The Montreal Cognitive Assessment (MoCA), was designed as a rapid screening instrument for cognitive dysfunction and is used in many studies of PD patients [9,10]. It assesses different cognitive domains, including attention and concentration, executive functions, memory, language, visuoconstructional skills, conceptual thinking, calculations, and orientation [9].

There is ongoing active research towards improved prediction of outcome in PD patients [11–16]. This is especially important as one aims to properly adapt and power clinical trial studies of novel disease modifying therapies. Past efforts have primarily focused on prediction

---

of motor outcome [13,57]. Cognitive decline might however be independent of motor decline in PD [17]. While older age has been shown as highly predictive of cognitive decline and dementia, memory concerns have recently become prevalent in younger age groups, resulting in referral of many patients from primary care to professional memory clinics for assessment [18]. Past efforts to predict cognitive decline in PD have been primarily based on using simple linear regression [19,20]], (see the discussion section). In the present work, we thus set to investigate a range of advanced machine learning methods for optimal prediction of early cognitive decline in de-novo PD patients.

Machine learning algorithms allow improved task performance without being explicitly programmed [21]. Approaches based on machine learning aim to build classification or prediction algorithms automatically by capturing statistically robust patterns present in the analyzed data. Most predictor algorithms are not able to work with a large number of input features, and thus it is necessary to select the optimal few features to be used as inputs. Furthermore, using only the most relevant features may improve the prediction accuracy. The process of feature selection can be performed either manually or automatically using feature subset selector algorithms (FSSAs). Our present efforts include search for optimal combination of machine learning methods (incorporating static and dynamic algorithms) and FSSAs for the task of predicting cognitive outcome in PD patients, quantified as the MoCA score. A range of predictor algorithms were selected amongst various families of learner and regressor algorithms, and in addition, a range of FSSAs were considered in combination with predictor algorithms, to optimize prediction of the outcome. Using clinical data measured at baseline (year 0) and year 1, we set to predict MoCA score at year 4.

2. Materials and methods

2.1. Longitudinal patient data

Two datasets extracted from the PPMI database (www.ppmi-info.org/data) were analyzed. In the first dataset, a wide range of potential predictor features were considered, as summarized in Supplemental Table 1. A total of 184 PD subjects (123 males, 61 females; average age in year 0: 67.9 ± 9.60, range (39,91)) were included in this data set, for which all 93 features were available, as shown in Supplemental Table 1.

The second larger patient dataset only needed to include predictor features that were determined to be important in our analysis of the first dataset (namely features 9, 10, 16, 28, 35 and 51 in Supplemental Table 1), as elaborated below. Using a smaller number of features enables the use of a larger number of PD subjects from the PPMI database as not all patients had all the 93 features we used originally in our study. However, in the next stage, we were able to select a larger set of patients who had the vital 6 features that we identified in our original study of the 93 features. Thus, following this first step, and identifying the vital predictive features, we were able to utilize a larger number of patients for further validation of our work specifically, we next evaluated 466 subjects (including 308 PD and 158 healthy controls). We performed analysis on the entire set (n = 466) as well as the PD only subset (n = 308). The 466 subjects consisted of 302 males, 164 females; average age in year 0: 67.5 ± 10.3, range (39,91). The 308 subset of PD subjects consisted of 206 males, 102 females; average age in year 0: 67.5 ± 9.90, range (39,91).

The MoCA score at year 4 was the predicted outcome measure. The MoCA consists of a series of tests; the combined score ranges from 0 to 30 with 30 corresponding to no cognitive impairment. In the first dataset (N = 184 subjects), the mean MoCA score was 26.5 ± 3.40; range (13,30). In the second dataset (N = 466), the mean MoCA score was 26.8 ± 3.20; range (11,30). In the PD-only subset (N = 308), the mean MoCA score was 26.5 ± 3.50; range (11,30). Fig. 1 shows the distribution of datasets. For improved robustness of clinical measures,
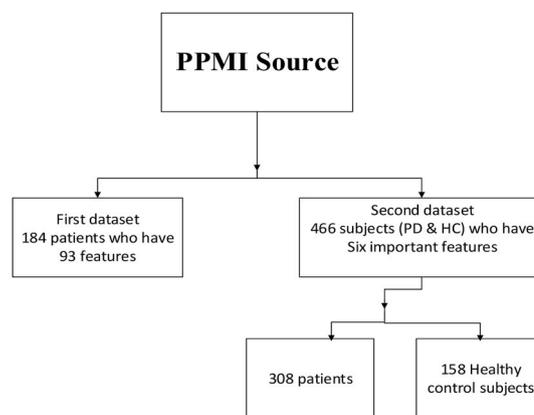

Fig. 1. Diagram of datasets.

some scores (e.g. UPDRS I, II, III) were averaged if available within ± 6 months. All applied algorithms except LASSOLAR, BRR, DTC, PAR and Thiel-Sen Regression were implemented in Matlab R 2016 b platform. The remaining algorithms were implemented in the Python 3.7.2 platform.

2.2. Machine learning methods

Two group of algorithms were employed: 1) Predictor algorithms; and 2) Feature Subset Selector algorithms (FSSAs).

2.2.1. Utilizing automated machine learning hyperparameter tuning to adjust parameters of predictor algorithms

A range of optimal predictor algorithms were selected amongst various families of learner and regressor algorithms. These are all listed in the Supplement (part 2). Specifically, we selected 10 predictor algorithms: 1) LOLIMOT (Local Linear Model Trees) [22,23], 2) RBF (Radial basis Function) [24], 3) MLP-BP (Multilayer Perceptron-Back propagation) [25,26], 4) LASSOLAR (Least Absolute Shrinkage and Selection Operator – Least Angle Regression) [27,28], 5) RFA (Random Forest Algorithm) [29,30], 6) RNN (Recurrent Neural Network) [31,32], 7) BRR (Bayesian Ridge Regression) [33–35], 8) DTC (Decision Tree Classification) [36–38], 9) PAR (Passive Aggressive Regression) [39–41], and 10) Thiel-Sen Regression [42–44]. In this work, we automatically adjusted intrinsic parameters such as the number of neurons and number of layers in the predictor algorithms via automated machine learning hyperparameter tuning. Such hyperparameter tuning was used in various algorithms such as LOLIMOT, RBF, RNN, MLP-BP, RFA so that algorithm parameters were automatically optimized given one of the data arrangements prior to formal training/validation/testing processes. Automated tuning, implemented with our own in-house code, performs an error minimization search scheme, seeking to optimize the hyperparameters starting with random initialization, pursuing a systematic trial-and-error search scheme for tuning the parameters.

2.2.2. Utilizing FSSAs for feature selection

6 FSSAs were employed and compared to select the most effective features (see Supplement part 3 for more details): 1) GA (Genetic Algorithm) [45,46], 2) ACO (Ants Colony Optimization) [47,48], 3) PSO (Particle Swarm Optimization) [49,50], 4) SA (Simulated Annealing) [51,52], 5) DE (Deferential Evolution) [53,54], and 6) NSGAII (Non-dominated Sorting Genetic Algorithm) [55,56]. All algorithms aimed to minimize the prediction error by selecting the best combination of features, while NSGAII additionally aimed to reduce number of features.

2.3. Analysis procedure

Evaluation was performed in two stages. In the first stage, we utilized the dataset with 184 subjects who had all 93 features in Supplemental Table 1. We generated 10 randomized arrangements of the dataset. In each run, we randomly allocated 65% of patients to training, 5% to training validation in order to minimize overfitting from the training step, and 30% to final testing (thus repeating this process 10 times). Subsequently, the mean absolute error and standard deviation of the MoCA score prediction for each method were reported and compared. These were computed from the final test sets in the randomized arrangements for more reliable and appropriate assessment of our results.

In the second stage, we first performed systematic feature selection via FSSAs of the dataset with 184 patients, identifying the most vital features for prediction of outcome. We then created an expanded dataset of patients that contained these selected vital features. As such, this enabled us to arrive at two expanded datasets: 466 subjects (PD and HC subjects) and 308 patients (PD only). For each, a single randomized arrangement was considered for additional validation of our work (~65% for training, ~5% for training validation and ~30% for final testing), while we ensured that the new final test sets in the expanded datasets did not include any subjects in the original set (184 patients) as used in first stage efforts, for completely independent testing. The mean absolute error from the final test sets were then reported.

3. Result

3.1. First stage analysis including the complete set of input features

The mean absolute errors of the tested predictor algorithms obtained when all features (93 features) were applied to 10 predictor algorithms are plotted in Fig. 2. The LASSOLAR predictor algorithm significantly outperformed the other algorithms (p-value < 0.05; nonparametric Friedman test), reaching a mean absolute error of 1.83 ± 0.13; the poorest result was reached by the RBF (6.54 ± 0.48). Several other algorithms also performed relatively well, such as BRR which reached absolute error of 2.10 ± 0.19. A plot of the predicted MoCA scores versus the actual scores in one of the final test arrangements of LASSOLAR is shown in Fig. 3. In the plot, r is the Pearson correlation coefficient between true and predicted outcomes.
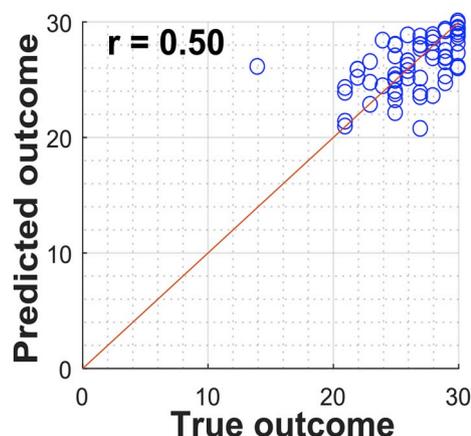

Fig. 3. A typical performance of LASSOLAR for prediction of MoCA outcome in 55 final test patients. X axis shows the actual scores and Y axis depicts the predicted MoCA scores. R is the Pearson product-moment correlation coefficient between true vs. predicted outcomes.

Table 1
Feature subsets selected by 6 Feature Subset Selector Algorithms (FSSAs).

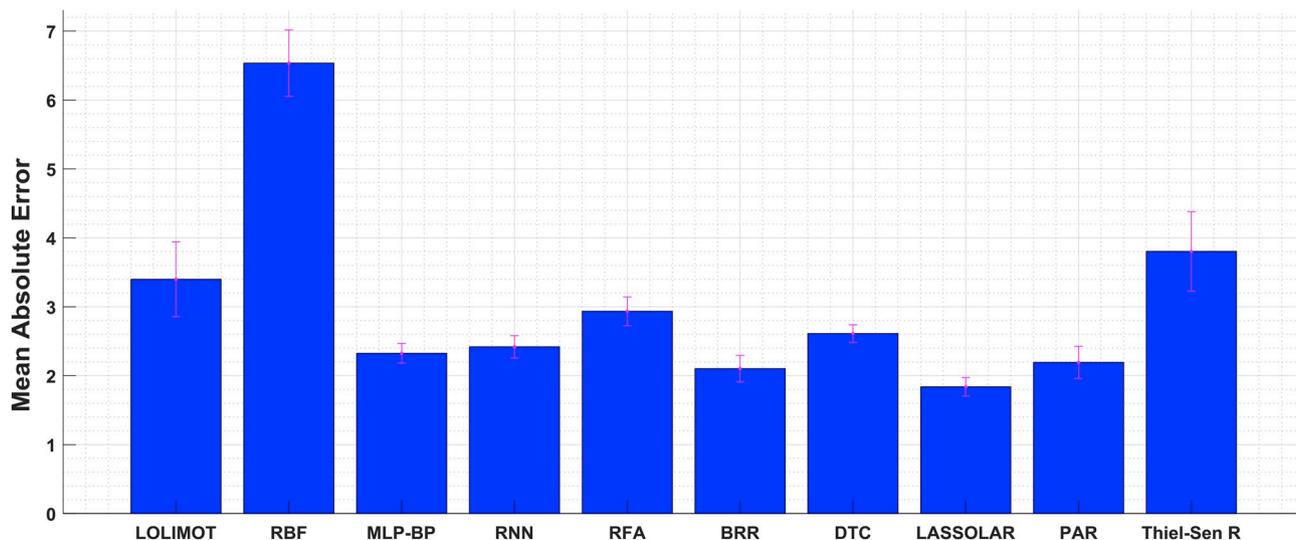|	GA	SA	DE	PSO	ACO	NSGAII
Selected features	7	1	8	2	2	9
	9	8	9	3	5	10
	10	9	10	8	9	16
	15	10	15	10	10	28
	16	15	16	15	15	35
	20	16	28	23	29	51
	28	24	35	24	30	
	35	27	40	40	38	
	44	40	51	43	40	
	51	42	54	51	43	
	53	44	56	52	51	
	56	51	58	58	56	
	60	56	60	64	72	
	68	61	65	67	73	
	69	69	66	71	78	
	72	73	72	73	84	
	80	79	73	77	85	
	90	90	77	84	90	

Fig. 2. Mean absolute errors in units of MoCA score of the tested predictor algorithms computed from 10 randomized data arrangements. The error bars represent the standard deviation. X axis lists the predictor algorithms and Y axis shows their mean absolute errors.
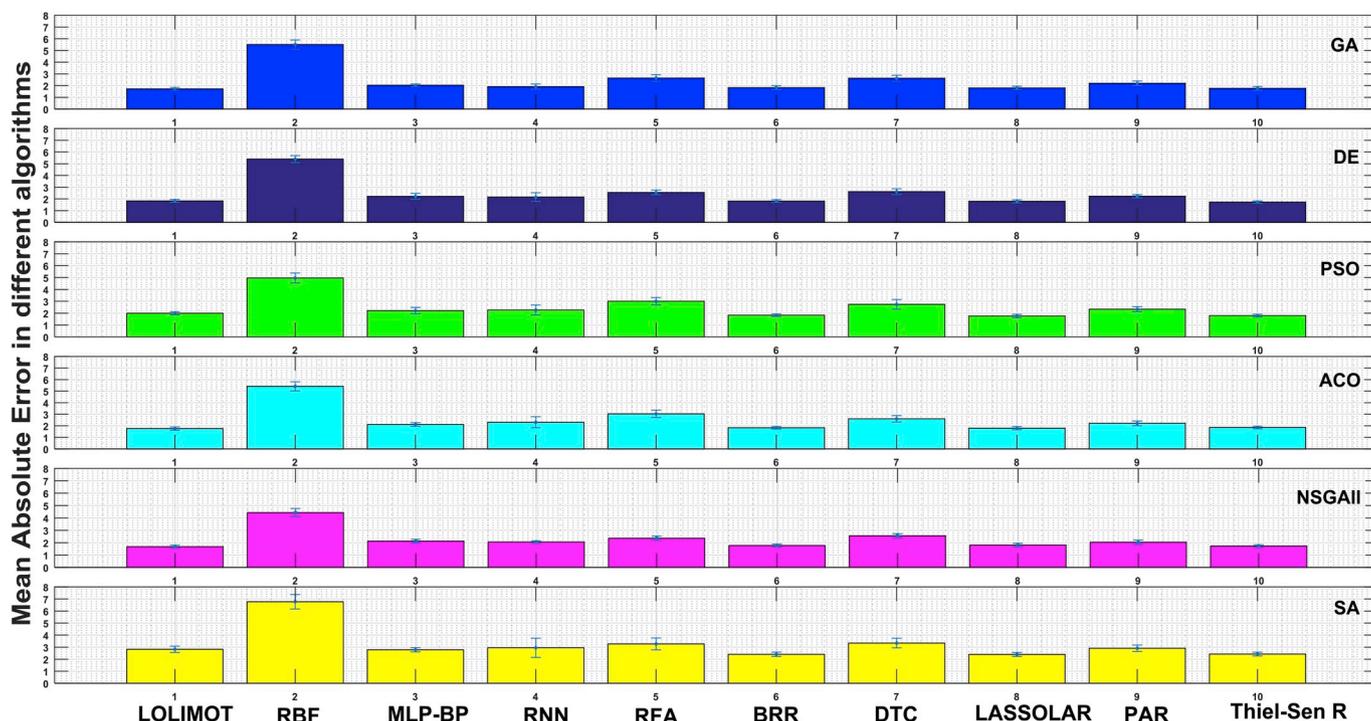
Fig. 4. Performance plots for application of the 6 FSSAs followed by 10 predictor algorithms. The error bars represent the standard deviation. In each part, X axis lists the predictor algorithms and Y axis shows their mean absolute errors.

3.2. Second stage analysis including selected sets of input features by FSSAs

In our subsequent efforts, pre-selection of features was performed via FSSAs prior to application to predictor algorithms. The selected subsets of features by each of the 6 FSSAs are shown in Table 1. The selected features are described in Supplemental Table 1.

All features selected via the FSSAs were then applied to the 10 predictor algorithms, and the results are shown in Fig. 4. It was seen that all FSSAs, with the exception of SA, reached acceptable results, i.e. the mean absolute errors in the predictor algorithms reached ~1.70. A plot of the predicted MoCA scores versus the actual scores in one of the final test arrangements of LOLIMOT is shown in Fig. 5. The errors thus obtained using all FSSAs except ACO were significantly lower ($p < 0.05$; paired t-test) compared to the lowest error obtained by the first approach (1.83 ± 0.13), while using non-parametric Friedman test, only errors of NSGAII and DE were significantly lower ($p < 0.05$)

compared to the lowest error obtained by the first approach.

NSGAII was able to find the most optimal feature combinations from the 93 features. It selected 6 features as most important, namely features 9, 10, 16, 28, 35 and 51, corresponding to: (i,ii) MoCA years 0 and 1, (iii) REM (Sleep Behavior Disorder Questionnaire) year 1, (iv,v) LNS (Letter Number Sequencing) Number 4 year 0 and Number 3 year 1, and (vi) STAIA (State-Trait Anxiety Inventory for Adults) year 0. The experiments and results with NSGAII are further elaborated in the Supplement (part 4). It is a worth noting that LOLIMOT does not work very well with very large number of inputs (thus did not perform best in part A), but is an excellent algorithm for a reasonable (< 20) number of inputs [57], as we discuss more below.

In addition, using three predictor algorithms such as LOLIMOT, MLP-BP and RFA, we performed a prediction task using solely the MoCA score at year zero. The comparison between those results and results of NSGAII (6 vital features) are shown in Table 2.

Overall, integrating the vital six predictive features, allowed us to reach significantly lower errors (all p-values < 0.01 using paired t-test; and < 0.05 using nonparametric Friedman test).

3.3. Additional independent testing

We subsequently created additional independent sets of patients having the most important features (e.g. only 184 subjects had all 93 features as mentioned above, but a total of 466 subjects, i.e. 282 additional, had the 6 identified optimal features). This allowed additional
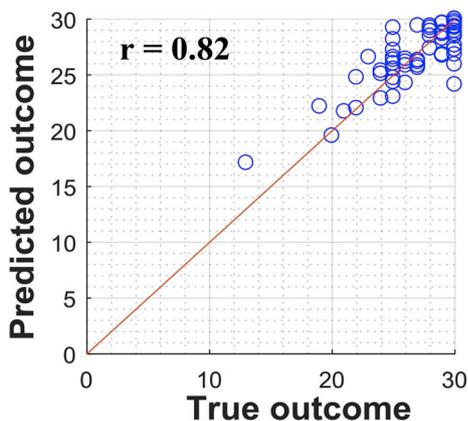

Fig. 5. A typical predictive performance of LOLIMOT (following FSSA pre-selection of features) for 55 patients. X axis shows the actual scores and Y axis indicates the predicted MoCA scores. R is the Pearson product-moment correlation coefficient between true vs. predicted outcomes.

Table 2
Performance table of results for prediction task when using MoCA in year zero alone vs. using all six vital features.

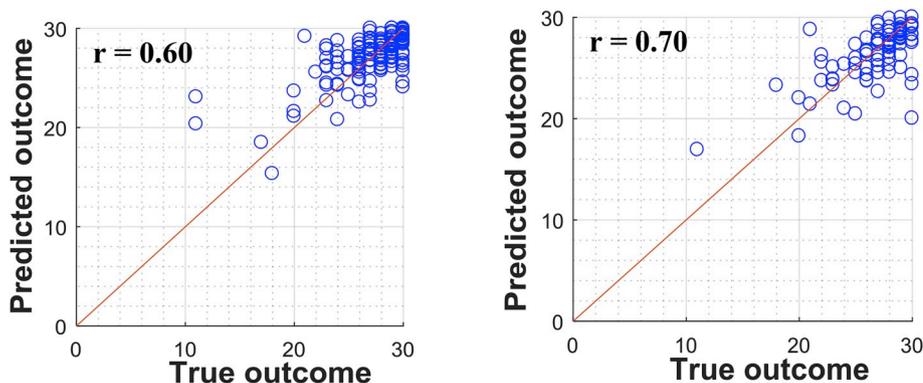|	Mean Absolute Error (only MoCA - year zero)	Mean Absolute Error (using all six vital features)
LOLIMOT	2.27 ± 0.20	1.68 ± 0.12
MLP-BP	2.47 ± 0.20	2.12 ± 0.15
RFA	2.87 ± 0.25	2.36 ± 0.17

Fig. 6. Predictive performance of LOLIMOT (following NAGAII preselection of features) for (left) 140 subjects including new PD & HC, and (right) 93 new PD subjects. X axis shows the actual scores and Y axis depicts the predicted MoCA scores. R is the Pearson product-moment correlation coefficient between the true and predicted outcomes.

validation of our work. In the expanded set, we created a single arrangement for training, training validation and final test (~65% for train, ~5% for training validation and ~30% for final test), while we ensured that the new final test set only included newly included patients which were not in main dataset, to obtain a completely independent testing. This resulted in an error of 1.73. Similarly, we utilized only PD subjects (n = 308), again ensuring that the final test set only included newly included patients (which were not in main dataset); the absolute error obtained in the independent final test set was 1.65. These are shown in Fig. 6.

4. Discussion

Prediction of disease progression in PD patients holds significant value. Individual knowledge of disease progression can help to make proper social and occupational decisions in association with future physical functioning of recently diagnosed patients. The analysis proposed in this work can help detect variables relevant to identification of disease progression and thus guide the design and interpretation of clinical trials involving neuroprotective and symptomatic therapy [58]. This is particularly relevant as PD progression is heterogeneous; Hely et al. [59] showed that despite the employment of different targeted treatment strategies, during 10 years of observation, 9 of 126 patients progressed to confinement to bed or a wheelchair unless aided, whereas 13 patients remained without significant functional restrictions. Such variability leaves us with the challenge of predicting the future course of PD in individuals and groups of patients.

Camargo et al. [60] showed that MoCA can be a good screening test for cognitive function in PD. There are number of benefits [61] for using MoCA that consists of it 1) being a quick measure of global cognitive function with short administration time compared with other tests, 2) covering a wide range of function in the cognitive domain, 3) having sensitivity to milder cognitive deficits in PD, 4) capturing executive dysfunction, 5) being widely adapted and utilized both clinically and in clinical research, 6) being widely adopted in many disease states, and 7) having alternate variants proposed for multiple languages and a more specific subtest weighting for PD. At the same time, there are some limitations including cultural and language sensitivity and different cut-offs for illiterate patients.

Another cognitive test, namely MMSE (Mini–Mental State Examination) can also be effective. In a study of 50 PD patients, it was shown to outperform MoCA (AUC = 0.936 vs. 0.906) though this was with respect to another cognitive scale namely SCOPA-Cog taken as gold reference [9]. A MoCA score of < 21 was comparable with a MMSE score of < 26, and both were correlated with age rather than severity of motor symptoms, thus it was inferable that cognitive decline might be independent of motor decline in PD [17]. While older age was shown as highly predictive of cognitive decline and dementia, memory concerns have recently become prevalent in younger age groups, resulting in referral of many patients from primary care to professional

memory clinics for assessment [18].

Given that MoCA, compared to the Mini-Mental State Exam (MMSE), had a high discriminatory power in detecting mild cognitive impairment (PD-MCI) and a MoCA score of ≤ 26 provided a sensitivity of 93.1% for the diagnosis of PD-MCI in a longitudinal study over 2 years, MoCA test in year zero was a reliable feature in predicting cognitive decline in early PD [19]. Schrag et al. [20] performed univariate and multivariate linear analyses on baseline data for 390 patients with Parkinson's disease, evaluating cognitive decline using (i) MoCA score changes between baseline and year 2, (ii) MoCA scores at year 2 as dependent variables, and (iii) diagnosis of cognitive impairment at year 2. They showed good predictive performance using multivariate analyses. We note that our present work considers a more challenging task of predicting outcome in year 4, and in our experience, usage of more sophisticated (non-linear) models was necessary. Furthermore, because of our year 4 MoCA prediction task, we did not have a large set of patients with features such as CSF amyloid to t-tau ratio, APOE status and DAT SPECT, to allow us to perform direct comparison with the above-mentioned year 2 prediction study. As more data becomes available, we are motivated to incorporate these additional features for improved prediction tasks.

In this work, we explored two general methods, both of which resulted in very good prediction of the MoCA score in year 4. In the first effort, we employed 10 predictor algorithms that were robust predictors among different families of machine learning methods, which included static and dynamics algorithms. When all features were applied to the predictor algorithms, the lowest mean absolute error ~1.83 ± 0.13 was obtained by LASSOLAR.

While many predictor algorithms are not able to work with very large input features, LASSOLAR is designed to work with large inputs features. To eliminate the challenge arising from having many input variables, we then employed FSSAs for pre-selecting more effect features. As shown in Fig. 4, all FSSAs except SA selected acceptable combinations, so that the mean absolute errors decreased to 1.68 ± 0.12. GA selected 18 features, and upon application of the selected features, LOLIMOT reached a mean absolute error 1.70 ± 0.10, though some other predictor algorithms such as RNN, LASSOLAR, BRR and Thiel-Sen R also reached approximately similar results relative to LOLIMOT. DE also selected 18 among 93 features, and subsequently, prediction via Thiel-Sen R reached a mean absolute error 1.70 ± 0.10; other algorithms such as LOLIMOT, LASSOLAR and BRR also resulted in approximately similar results compared to Thiel-Sen R. PSO selected number features similar to GA, and with those selected features applied to all predictor algorithms, LASSOLAR reached a mean absolute error of 1.76 ± 0.15. ACO selected features similar to GA, so that when those applied to predictor algorithms, LOLIMOT resulted in an error of 1.76 ± 0.14, although some other algorithms such as Thiel-Sen R, LASSOLAR and BRR performed nearly similarly. The errors thus obtained using FSSAs, with the exception of using the latter (ACO-based), were significantly lower (p < 0.05; paired *t*-test) compared to the

lowest error obtained by the first approach (1.83 ± 0.13), while using non-parametric Friedman test, only performance of NSGAII and DE were significantly better (p < 0.05) compared to the lowest error obtained by the first approach.

As shown in Table 1, NSGAII selected the most optimal combinations among all FSSAs. It selected six features as vital features for outcome prediction. According to Table 1 and Fig. 4, NSGAII acted differently compared to all FSSAs by selecting fewer features. Overall, features 9, 10, 16, 28, 35 and 51 ((i,ii) MoCA years 0 and 1, (iii) REM (Sleep Behavior Disorder Questionnaire) year 1, (iv,v) LNS (Letter Number Sequencing) Number 4 year 0 and Number 3 year 1, and (vi) STAIA (State-Trait Anxiety Inventory for Adults) year 0 were seen to be most predictive of outcome. FSSAs such as GA and DE also selected these six vital features, and predictor algorithms associated with them reached a mean absolute error of ~1.70. As shown in Table 2, we reached errors 2.27 ± 0.20, 2.47 ± 0.20 and 2.87 ± 0.25 when we employed LOLIMOT, MLP-BP and RFA applied on sole MoCA score in year zero respectively. Hence, worse performance resulted only using MoCA score in year zero compared with more important six features selected by NSGAII.

The predicted feature, MoCA in year 4, assesses a range of cognitive domains including attention and concentration, executive functions, memory, language, visuoconstructional skills, conceptual thinking, calculations, and orientation. We first elaborate the selected features: Features 9 and 10 are MoCA scores in year zero and one, which are naturally selected as the prediction task itself is for MoCA in year 4. Feature 16 is REM Sleep Behavior Disorder Questionnaire year one (in fact, PD subjects that suffer from REM sleep disorder have been shown to have higher abnormalities in the cholinergic system which has been associated with higher prevalence of cognitive impairment [62]). Subject self-rating instrument assesses sleep behavior with short questions. Questions are framed to gather information about current behaviors. Feature 28 is Letter Number Sequencing-NUM 4 year zero and Feature 35 is Letter Number Sequencing -NUM 3 year one. For Letter Number Sequencing test, the subject is read a combination of numbers and letters and is asked to recall the numbers first in ascending order and then the letters in alphabetical order. Feature 51 is State-Trait Anxiety Inventory for Adults. Although many of the items attempt measuring "anxiety", approximately half of them inquire about negative characteristics, (e.g., feeling "tense," "frightened," or "upset"). There is also evidence that anxiety and cognitive impairment are linked [63]. Overall, we believe, in this data-driven discovery framework, that these non-motor features all demonstrate different aspects of cognitive behavior.

Finally, we performed a final validation test involving new patients, on the optimal feature combination arrived at by NSGAII. Specifically, we compared results obtained on the main prior test (184 patients who had all 93 features as used in main dataset) with results obtained on an expanded set (308 PD patients; obtained by selecting new patients who had the optimally needed vital features). In this expanded set, for training validation (5% of all patients) and final test (30% of all patients), we only used new patients so that all 184 common patients were only included in training data (65% of whole patients). By using 93 new patients, we reached an absolute error of ~1.65. This result is approximately similar to the results of the main final test; i.e. absolute error ~1.68 ± 0.12.

We also performed a second validation test by combination of PD and HC subjects. For this additional independent test, we selected 466 subject that had the most vital features selected by NSGAII. We applied 65% training (303 subjects; including the 184 prior patients used in initial dataset and 119 subjects of new PD and HC cases), 5% training validation and 30% final testing (using only new subjects), arriving at an absolute prediction error of ~1.73. This result is similar to results obtained for the other datasets. Overall, the results of independent tests were consistent with one another, confirming our findings on optimal prediction of cognitive outcome in PD patients. We also performed

additional validation testing for other combinations selected via NSGAII (see supplement, part 4). In comparison, all mean absolute errors for new patients were less or equal to mean absolutes errors of main prior test (the lowering is attributed to a larger training set in new expanded patient set), further confirming our findings.

The limited size of a dataset is a limiting factor in outcome prediction; as such, to maximize our numbers, we had to select a set of 184 patients for which imaging data was not available for all patients and thus could not evaluate the impact of imaging. Another approach we used to tackle this limitation was, after we had identified the vital six predictive features, to look for additional patients which had these vital features, further expanding our dataset, and allowing additional independent testing. In our work, we used features subset selection algorithms to reduce the number of features (for size reduction) to avoid overfitting, although it was possible to use utilize extraction algorithms such as PCA which we hope to explore in future work. At the same time, we believe feature selection might be more clinically informative than feature extraction, providing insights as to which features are most important, whereas in feature extraction, features are combined and transformed into new dimensions and thus may not be easily interpretable.

5. Conclusion

This work has demonstrated that by employing appropriate machine learning tools, including automated machine learning hyperparameter tuning, it is possible to attain very good prediction of cognitive outcome. Our work involved prediction of MoCA in year 4. It was shown that LASSOLAR was able to work with a significant number of inputs directly, reaching the smallest mean absolute error of 1.83 ± 0.13. In another approach, FSSAs were first utilized, pre-selecting optimal features prior to application of machine learning algorithms. Improved results, reaching a mean absolute error of ~1.68 ± 0.12 were obtained. For selection of optimal combinations of features by FSSAs, NSGAII performed most favorably, selecting 6 vital features for prediction of cognitive outcome, namely MoCA years 0 and 1; REM (Sleep Behavior Disorder Questionnaire) year 1; LNS (Letter Number Sequencing) Number 4 year 0 and Number 3 year 1; STAIA (State-Trait Anxiety Inventory for Adults) year 0. Additional independent testing was performed, on larger datasets containing the most predictive features, confirming our findings. Overall, we conclude that optimal utilization of feature selection and predictor algorithms can produce excellent prediction of cognitive outcome in PD patients.

Data and codes availability

All codes (included predictor algorithms and Feature Subset Selector Algorithms) and all datasets are publicly shared at: https://github.com/salmanpoor/Machine-Learning-Algorithms_M.R.Salmanpour-A.Saberimanesh-A.Rahmim.

Conflict of interest

The authors have no relevant conflicts of interest to disclose.

Summary

There are presently no disease modifying therapies for Parkinson's disease (PD), and appropriately powered clinical trials for discovery of such therapies are vital. Past efforts have primarily focused on prediction of motor outcome in PD. At the same time, cognitive decline is an important manifestation of PD in a large subset of patients and is a source of considerable burden. The present work utilizes the Montreal Cognitive Assessment (MoCA; range 0–30) as a measure of cognitive ability, and aims to predict outcome in year 4 from baseline (year 0) as well as year 1 data. The work probes a range of predictor machines such

as LOLIMOT, RNN, RBF, MLP_BP, PAR, BRR, LASSOLAR, Theil Sen R, RFA and DTC. We first analyzed a set of 184 patients with 93 potentially predictive features. We utilized 65% of patients for training, 5% for training validation, and 30% for final testing, performing 10 different randomized arrangements. In a first approach, we made direct use of the 93 features in each of the predictor machines, including use of automated machine learning hyperparameter tuning. Subsequently, LASSOLAR was shown to obtain the lowest mean absolute error of 1.83 ± 0.13. In the second part, 6 Feature Subset Selector Algorithms (FSSAs), namely GA, DE, SA, PSO, ACO, NSGAII were applied to pre-select features prior to application of predictor algorithms. Using the FSSAs, combined with LOLIMOT, enabled absolute error of ~1.70. Finally, additional independent tests were performed. We accumulated 466 subjects (308 PD + 158 healthy controls (HC) who had the most important features as identified above using FSSAs. Predictions in their test sets (which only included new patients) resulted in an absolute errors of 1.73 and 1.65 for PD + HC and PD-only sets, respectively, further confirming our findings. Overall, we demonstrate excellent prediction of cognitive outcome when utilizing appropriate and optimal feature selection and prediction algorithms.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2019.103347.

References

[1] T. Simuni, K. Sethi, Nonmotor manifestations of Parkinson's disease, Ann. Neurol. 64 (2008) 65–80.

[2] E. Wolter, Non-motor extranigral signs and symptoms in Parkinson's disease, Park. Relat. Disord. 3 (2009) 6–12.

[3] K. Bayulkem, G. Lopez, Non-motor fluctuations in Parkinson's disease: clinical spectrum and classification, J. Neurol. Sci. 289 (2010) 89_92.

[4] D. Dai, Y. Wang, et al., Polymorphisms of DRD2 and DRD3 genes and Parkinson's disease: a meta-analysis, Biomedica 2 (2) (2014) 275–281.

[5] T. Lebouvier, T. Chaumette, et al., The second brain and Parkinson's disease, EGN 30 (5) (2009) 735–741.

[6] P. McNamara, E. Stavitsky, et al., Side of motor symptom onset and pain complaints in Parkinson's disease, Int. J. Geriatr. Psychiatry 25 (2010) 519e24.

[7] N. Bohnen, R. Albin, The cholinergic system and Parkinson disease, Behav. Brain Res. 2 (221) (2011) 564–573.

[8] J.G. GabrielHou, E.C. Lai, NON-MOTOR SYMPTOMS OF PARKINSON'S DISEASE, Int. J. Gerontol. 2 (1) (2007) 53–64.

[9] C.H. Ferreira Camargo, E. d. Souza Tolentino, et al., Comparison of the use of screening tools evaluating cognitive impairment in patients with Parkinson's disease, Dement Neuropsychol 10 (4) (2016) 344–350.

[10] Parkinson Progression Marker Initiative, (2011).

[11] A. Nieuwboer, W.D. Weerdt, B. René Dom, Prediction of outcome of physiotherapy in advanced Parkinson's disease, SAGE Journals 16 (8) (2002) 886–893.

[12] S. Grill, J. Weuve, M. Weisskopf, Predicting outcomes in Parkinson's disease: comparison of simple motor performance measures and the Unified Parkinson's Disease Rating Scale-III. J. Parkinson's Dis. 3 (2011) 287–298.

[13] A. Rahmim, P. Huang, et al., Improved prediction of outcome in Parkinson's disease using radiomics analysis of longitudinal DAT SPECT images, Neuroimage: Clinic 16 (2017) 539–544.

[14] D. Arnaldi, F. De Carli, et al., Prediction of cognitive worsening in de novo Parkinson's disease: clinical use of biomarkers, Mov. Disord. 32 (2017) 1738–1747.

[15] I. Fyfe, Prediction of cognitive decline in PD, Nat. Rev. Neurol. 14 (2018) 213–317.

[16] C. Gao, H. Sun, et al., Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson's disease, Sci. Rep. (2018) 1–21.

[17] K. Ohta, T. Osada, et al., Comparing the montreal cognitive assessment with mini-mental state examination in Japanese Parkinson's disease patients, Neurology and Clinical Neuroscience 2 (2014) 44–49.

[18] A. Wojtowicz, A.J. Larner, Diagnostic test accuracy of cognitive screeners in older people, Prog. Neurol. Psychiatr. 21 (1) (2017) 17–21.

[19] N. Kandiah, A. Zhang, et al., Montreal Cognitive Assessment for the screening and prediction of cognitive decline in early Parkinson's disease, Park. Relat. Disord. 20 (11) (2014) 1145–1148.

[20] A. Schrag, U. Siddiqui, et al., Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed Parkinson's disease: a cohort study, Lancet Neurol. 16 (2017) 66–75.

[21] Y. Singh, P.K. Bhatia, O. Sangwan, A REVIEW OF STUDIES ON MACHINE LEARNING TECHNIQUES, Int. J. Comput. Sci. Secur. 1 (1) (2007) 70–84.

[22] O. Nelles, A. Fink, R. Isermann, Local linear model Trees (LOLIMOT) toolbox for nonlinear system identification, science Direct (IFAC System Identification) 33 (15) (2000) 845–850.

[23] J. Mart′nez-Morales, E. Palacios, Modeling of internal combustion engine emissions by, SciVerse Science Direct 3 (2012) 251–258.

[24] Y. Arora, A. Singhal, A. Bansal, A study of applications of RBF network, Int. J. Comput. Appl. 94 (2) (2014) 17–20.

[25] S. Alsmadi, M. khalil, et al., Back propagation algorithm: the best algorithm, IJCSNS International Journal of Computer Science and Network Security 9 (4) (2009) 378–383.

[26] D. Rumelhart, E. Geoffrey, et al., Leaner Representations by back-Propagating errors, Nature 323 (9) (1986) 533–536.

[27] V. Fonti, Feature Selection Using LASSO, VU Amsterdam, Amsterdam, 2017.

[28] B. Efron, T. Hastie, et al., Least angle regression, Ann. Stat. 32 (2004) 407–499.

[29] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.

[30] A. Jehad, R. Khan, N. Ahmad, Random forests and decision Trees, IJCSI International Journal of Computer Science Issues 9 (5) (2012) 272–278.

[31] A. Townley, M. Ilchmann, et al., Existence and learning of oscillations in recurrent neural networks, IEEE Trans. Neural Netw. 11 (1) (2000) 205–214.

[32] N. Maknickiene, V. Rutkauskas, et al., Investigation of financial market prediction by recurrent neural network, Innovative Infotechnologies for Science, Business and Education 11 (2) (2011) 3–8.

[33] A. Efendi, A simulation study on Bayesian Ridge regression models for several collinearity levels, AIP Conference Proceedings, 2017.

[34] C.M. Bishop, P.J.K.B.S. Michael Jordan (Ed.), Pattern Recognition and Machine Learning, 1th ed., Springer Science + Business Media, LLC, 233 Spring Street, New York, 2006New York, NY 10013, USA.

[35] G. Karabatsos, Fast Marginal Likelihood Estimation of the Ridge Parameter(s) in Ridge Regression and Generalized Ridge Regression for Big Data, Statistics-, 2015, pp. 1–44.

[36] M. RodneyOD, P. Goodman, Decision tree design using information theory, Knowl. Acquis. 2 (1990) 1–19.

[37] S. Chourasia, Survey paper on improved methods of ID3 decision tree, International Journal of Scientific and Research Publications 3 (12) (2013) 1–4.

[38] D. Denison, C. Holmes, et al., Bayesian Methods for Nonlinear Classification and Regression, John Wiley and Sons, New York, 2002.

[39] K. Crammer, O. Dekel, et al., On-line passive-aggressive algorithms, J. Mach. Learn. Res. 7 (2006) 551–585.

[40] J. Lu, P. Zhao, C.H. Steven, Online passive aggressive active learning and its, JMLR: Workshop and Conference Proceedings 39 (2014) 266–282.

[41] M. Blondel, Y. Kubo, N. Ueda, Online passive-aggressive algorithms for non-negative matrix factorization and completion, Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics, vol. 33, PMLR, 2014, pp. 96–104.

[42] S.H. Shah, A. Rashid, et al., A comparative study of ordinary Least squares regression and theil-sen regression through simulation in the presence of outliers, Lasbela, U. J.Sci.Techl V (2016) 137–142.

[43] Theil, A rank-invariant method of linear and polynomial regression analysis. I, II, III, Nederl. Akad. Wetensch 53 (1950) 386–392 521–525, 1397–1412.

[44] P. Sen, Estimates of the regression coefficient based on kendall's tau, J. Am. Stat. Assoc. 63 (324) (1968) 1379–1389.

[45] J. McCall, Genetic algorithms for modelling and optimisation, J. Comput. Appl. Math. 184 (2004) 205–222.

[46] M. Mitchell, Genetic algorithms: an overview, Complexity 1 (1) (1995) 31–39.

[47] C. Blum, Ant colony optimization: introduction and recent trends, Phys. Life Rev. 2 (2005) 353–373.

[48] P. Sivakumar, K. Elakia, A survey of ant colony optimization, Int. J. Adv. Res. Comput. Sci. Softw. Eng. 6 (3) (2016) 574–578.

[49] Q. Bai, Analysis of particle Swarm optimization algorithm, Comput. Inf. Sci. 3 (2010) 180–184.

[50] S. Singh, A review on particle Swarm optimization algorithm, Int. J. Sci. Eng. Res. 5 (4) (2014) 551–553.

[51] W. Dolan, P. Cummings, M. LeVan, Process optimization via simulated, AIChE J. 35 (1989) 725–736.

[52] S. Kirkpatrick, C. Gelatt, M. Vecchi, Optimization by simulated annealing, Science, New Series 220 (1983) 671–680.

[53] D. KARABOGA, S. OKDEM, A simple and global optimization algorithm for, Turk J Elec Engin 12 (2004) 53–60.

[54] A. Musrrat, M. Pant, A. Abraham, Simplex differential evolution, Acta Polytechnica Hungarica 6 (2009) 95–115.

[55] D. Kalyanmoy, A. Associate, et al., A fast and elitist multiobjective genetic algorithm, IEEE Trans. Evol. Comput. 6 (2002) 182–197.

[56] Y. Yusoff, M. Ngadiman, A. Mohd Zain, Overview of NSGA-II for optimizing machining process parameters, Procedia Engineering 15 (2011) 3978–3983.

[57] M.R. Salmanpour, M. Shamsaeei, et al., Machine learning methods for optimal prediction of outcome in Parkinson's disease, IEEE Nucl. Sci. Symp Conf. Record,

Sedney, 2018.

[58] C. Marras, P. Rochom, A.E. Lang, Predicting motor decline and disability in Parkinson disease, Arch. Neurol. 59 (2002) 1724–1728.

[59] M. Hely, J. Morris, et al., The sydny multicentre stuly of Parkinson's Disease : progression and mortality at 10 years, J Neurol Neurosurg Psychiarty 67 (1999) 300–3007.

[60] C. Camargo, E. Tolentino, et al., Comparison of the use of screening tools, Dement Neuropsychol 10 (4) (2016) 344–350.

[61] M. Skorvanek, J.G. Goldman, et al., Global scales for cognitive screening in Parkinson's disease: critique and recommendations, Mov. Disord. 33 (2) (2017) 208–218.

[62] H. Shimada, S. Hirano, et al., Mapping of brain acetylcholinesterase alterations in Lewy body disease by PET, Neurology 73 (4) (2009) 4 273-278.

[63] C.F. Delphin-Combe, A. Bathsavanis, et al., complaint., Relationship between anxiety and cognitive performance in an elderly population with a cognitive, Eur. J. Neurol. 23 (7) (2016) 1210–1217.