Machine Learning Methods for Optimal Prediction of Outcome in Parkinson's Disease

M. R. Salmanpour, M. Shamsaei, A. Saberi, S. Setayeshi, E. Taherinezhad, I. S. Klyuzhin, J. Tang, V. Sossi, A. Rahmim

ABSTRACT-In the present work, we systematically probe a range of predictor machines (11 machines), and aim to find the best combinations of features to result in improvements in prediction of outcome in PD. First, we created 32 combinations of 18 conventional features experimentally and selected 4 arrangements of 204 PD subjects. The combinations were applied to the various predictor machines), thereby absolute error of the best combination reached 4.3 (in prediction of MDS-UPDRS-III motor performance in year 4). This is in comparison to previous works that attained errors of around 9. In second part, subset selector machines were used for selecting the best combinations between all features, and GA and ACO selector machines selected the best combinations, further lowering error when combined with LOLIMOT for prediction. Selected features by GA and ACO (UPDRS I-Year 1, UPDRS III-Year 1, left putamen Uptake-Year 1, Age, Gender) had positive effect on prediction of outcome and mean absolute error reached 4.15. Moreover, other subset selector machines also reached acceptable results; mean absolute errors in some predictor machines were below 4.7. Overall, LOLIMOT was seen as the best predictor machine, and GA and ACO as the best feature subset selector machines. Furthermore, MDS-UPDRS III-years 0 and 1, MDS-UPDRS II- year 1, MDS-UPDRS I- years 0 and 1, age and DAT SPECT putamen as well as caudate uptake - year 1 were seen as most important predictors of outcome.

I. INTRODUCTION

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder after Alzheimer's disease, affecting over 1% of individuals over the age of 60 [1]. PD is a progressive, degenerative movement disorder, characterized by neuronal loss in the substantia nigra with the loss of dopaminergic terminals in the basal ganglia [2,3,4]. It is characterized by a series of motor and non-motor symptoms such as resting tremor, rigidity, bradykinesia, postural instability and autonomic dysfunction [5]. There is not any proven disease modifying therapies for PD, hence, there is an essential need to establish biomarkers of disease progression is vital [6] e.g. an aim of the Parkinson's Progressive Marker Initiative (PPMI) [7]. Furthermore, there is significant interest in prognostication of disease outcome, to properly adapt and power clinical trial studies, as applied to appropriate patients. Stratification of PD based on expected prognosis would allow better designs of disease modifying trials, with greater power to ascertain efficacy [8,9]. There is increasing research into prediction of outcome in PD [8,10,11,12,13,14] in order to properly adapt and power clinical trial studies of novel disease modifying therapies. An aim is to identify and utilize the most relevant features to predict outcome [15].

In the present work, we aim to utilize longitudinal information for improved prediction of outcome as well as identification of optimal combinations of features. We aim to predict motor assessment (MDS-UPDRS-III) at year 4 from data at baseline (year 0) and year 1, by optimal combination of machine learning methods, including static machines, dynamics machines and subset-selector machines [16].

II. MATERIALS AND METHODS

Data were extracted from the PPMI database (www.ppmi-info.org/data) [17]. The movement disorder society unified Parkinson's disease rating scale (MDS-UPDRS) – part III (motor) in year 4 was used as outcome. We considered 18 features as predictors: (1-6) MDS-UPDRS, parts I, II and III in year 0 and 1, (7-8) demographics (age, sex), (9-16) DAT SPECT images (putamen as well as caudate uptake, both left and right) in years 0 and 1, and (17-18) disease duration (DD), taken with respect to time of diagnosis (DD-diag.) as well as time of appearance of symptoms (DD-sympt.). For consistency, we only included patients that were off drug. These selection criteria resulted in 204 PD subjects (149 males, 55 females; average age in year 67.58±10.03, range [39, 91]), with widely distributed year 4 outcome UPDRS-III: 31.4±10.6; range [8,77]. For better consistency, MDS-UPDRS scores were averaged if within ±6 months.

Manuscript received Dec 19, 2018. This work was supported by the Michael J. Fox Foundation, including use of data available from the PPMI—a public-private partnership—funded by The Michael J. Fox Foundation for Parkinson's Research and funding partners (listed at www.ppmi-info.org/fundingpartners). This work was also supported by the Natural Sciences and Engineering Research Council of Canada.

M. R. Salmanpour is with Amirkabir University of Technology (Tehran Polytechnic), Iran (e-mail: m.salmanpoor66@gmail.com, mohamad91@aut.ac.ir).

M. Shamsaei is with Amirkabir University of Technology (Tehran Polytechnic), Iran (e-mail: pysham@aut.ac.ir).

S. Setayeshi is with Amirkabir University of Technology (Tehran Polytechnic), Iran (e-mail: setayesh@aut.ac.ir).

A. Saberi is with Central Tehran Branch, Islamic Azad University, Iran (e-mail: saberimanesh62@gmail.com).

E. Taherinezhad is with Kharazmi University, Tehran, Iran (e-mail: elahefereshte20@gmail.com).

Ivan S. Klyuzhin is with the University of British Columbia, Vancouver, Canada (e-mail: ivansk@physics.ubc.ca).

J. Tang is with Oakland University, Rochester, USA. (e-mail: jingtang@gmail.com).

V. Sossi is with the University of British Columbia, Vancouver, Canada (e-mail: vesna@phas.ubc.ca).

A. Rahmim is with the University of British Columbia, Vancouver, Canada (e-mail: arman.rahmim@ubc.ca).

978-1-5386-8495-5/18/$31.00 ©2018 IEEE

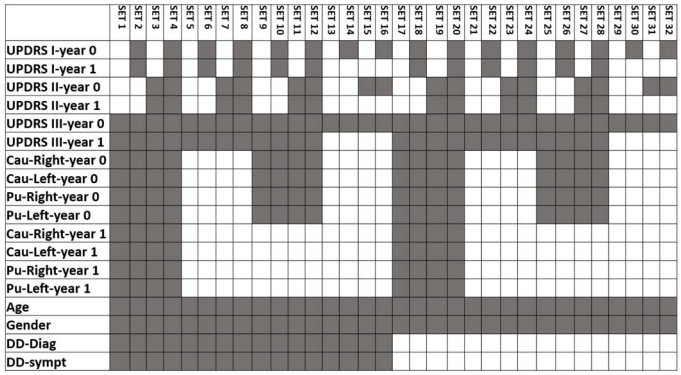
To derive optimal feature combinations, we pursue two approaches, as elaborated next.

A. Making the combinations manually

First, 32 combinations of the 18 features were experimentally created (Table 1), and each combination was assessed using different predictor machines. In this work, a range of predictor algorithms were selected amongst various families of learner and regressor algorithms including: 1) LOLIMOT (Local Linear Model Trees) [18,19], 2) RBF (Radial Basis Functions) [20], 3) MLP-BP (Multilayer perceptron-Backpropagation) [21,22], 4) LASSOLARs (Least Absolute Shrinkage and Selection Operator- Least Angle Regression) [23,24], 5) Random Forest Algorithm [25,26], 6) RNN (Recurrent Neural Network) [27,28,29], 7) BRR (Bayesian Ridge Regression) [30,31], 8) DTC (Decision Tree Classification) [32,33], 9) PAR (Passive Aggressive Regressor) [34,35,43], 10) Thiel-Sen Regressor [36,37,38,39], and 11) ANFIS (Adaptive neuro fuzzy inference system) [40,44].

Overall, 65% of all data were used for training, 5% for validation and 30% for testing. Four different randomized arrangements of the original data were generated. Thus each combination was run four times in order to train and plot the mean absolute errors and their standard deviation. Q-learning (a reinforcement learning technique) was used in most machines so that machine parameters were automatically optimized [45].

Table 1. various combinations between eighteen features



B. Selecting features through feature selector machines

Alternatively, we used systematic feature selection prior to prediction. Specifically, 6 subset selector machines were employed and assessed. All machines aimed to minimize error by selecting the best combination, while NSGAII also aimed to reduce number of features in the set. The machines used for feature subset selection were: 1) GA (Genetic Algorithm) [46,47], 2) ACO (Ant Colony Optimization algorithm) [48,49], 3) PSO (Particle Swarm Optimization algorithm) [50,51], 4) SA (Simulated Annealing) [52,53], and 5) DE (Differential Evolution algorithm) [54,55], and 6) NSGAII (Nondominated sorting genetic algorithm) [56,57].

Subsequently, the selected features were tested on the 11 predictor machines.

III. RESULTS AND DISCUSSION

In our first efforts (A), 32 manual combinations were experimentally created and were applied to the predictor machines. Results are shown in Fig. 1 (ANFIS was excluded for use as predictor machine because it is appropriate for combinations with lower number of features). As shown in Fig. 1, LOLIMOT resulted in best performance over a wide range of combinations. The best results for LOLIMOT were achieved for sets 18, 21 and 22, and some machines also reached similar results for those sets, but were less consistent when including other range of features. Overall, in prediction of year 4 MDS-UPDRS-III (outcome range [8-77]), absolute errors as low as 4.32 were achieved. Such prediction performance far exceeds results in other prior works where absolute errors of the order of 9 were obtained when using similar features [8]. Best results were observed when features MDS-UPDRS I (year 0 and 1), MDS-UPDRS-III (years 0 and 1), putamen as well as caudate uptake (both left and right; years 0 and 1), age and gender were in the sets. Conventional DAT SPECT images were seen to have some (but not large) effect. Some combinations such as sets 1, 3, 6, 7, 8, 19, 20, 23 and 24 also had good results, although they were less consistent than sets 18, 21 and 22.

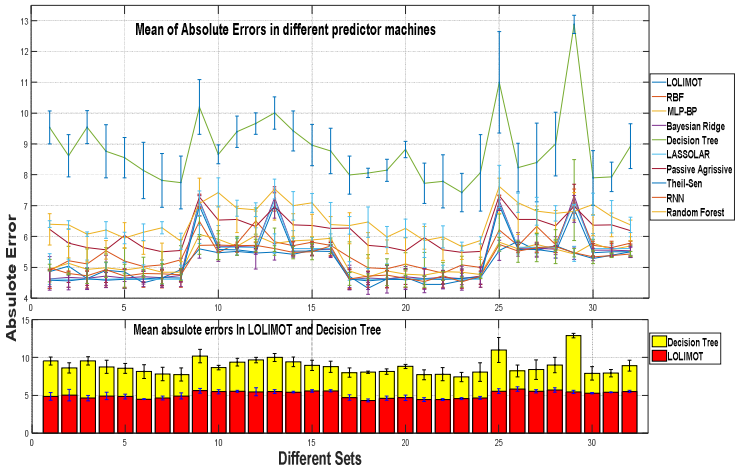


Fig. 1. (Top) Absolute errors in prediction of outcome from 10 predictor machines, each from 32 combinations shown in Table 1. ANFIS was excluded from results due to significantly poorer performance. (Bottom) Amongst the 10 machines, the best and worst performers are shown as bar plots.

Fig. 2 shows one of fitting curves of LOLIMOT results (one of 4 arrangements) where the Y axis is the predicted outcome and X axis reflects the true outcome. Overall, LOLIMOT was seen as the best predictor machine.

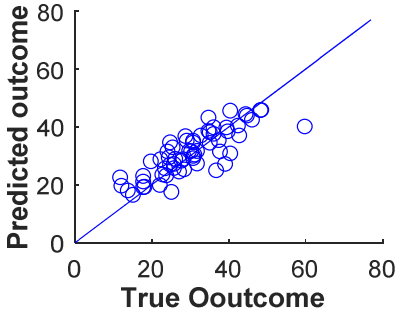


Fig. 2. Plot of outcome prediction from LOLIMOT making use of features in set 18.

In the second set of our efforts (B), we utilized preprocessing, using 6 feature subset selector algorithms. The features selected through these different machines are shown in Table 2.

Table 2. Selected features via different subset selector machines

Selectors	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5
GA	UPDRS I-Year one	UPDRS III-Year one	Left putamen uptake-Year one	Age	Gender
ACO	UPDRSI-Year one	UPDRSIII-Year one	Left putamen uptake-Year one	Age	Gender
PSO	UPDRSII-year zero	UPDRSIII-Year one	Left putamen uptake-Year one	DD-diag	Gender
SA	UPDRSII-year zero	UPDRSIII-Year one	Left putamen uptake-year zero	DD-diag	Gender
DE	UPDRSII-year zero	UPDRSIII-Year one	Left putamen uptake-Year one	DD-diag	Gender
NSGAII	UPDRS II- year zero	UPDRS III-year one	Age	Gender	----------

These featured combinations selected via the selector machines were then applied to the 11 predictor machines as mentioned in Sec. II-A. The results are plotted in Fig. 3. Features selected by subset feature selector machines enabled slightly better results than obtained in the previous section. Most subset feature selector machines selected combinations leading to prediction errors less than 5.

The performances were best for features selected by GA and ACO (UPDRS I-Year one, UPDRS III-Year one, Left putamen uptake -Year one, Age, Gender). In the case of LOLIMOT predictor machine, lowest mean absolute error of 4.15 was reached, though other machines such as LASSOLAR, Bayesian Ridge, Theil Sen Regressor and RNN also performed well. Plot of the results of GA and ACO as applied to LOLITMOT are shown in Fig. 4. Overall, GA and ACO were seen to perform as best feature subset selector machines.

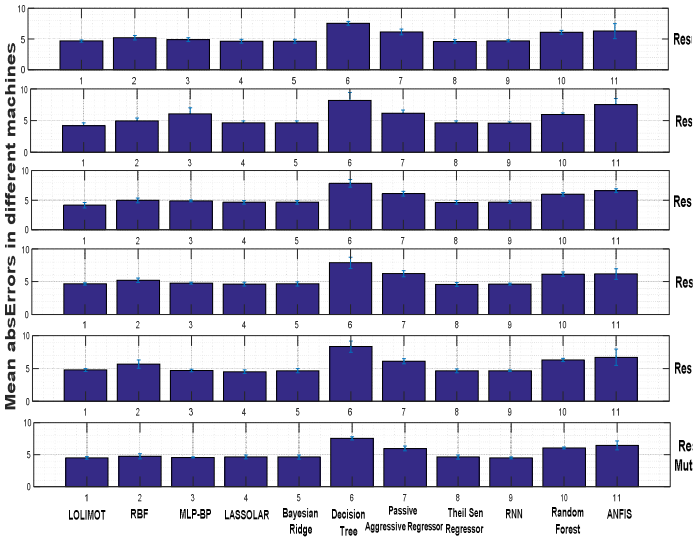


Fig. 3. 11 predictor machines applied to features selected by 6 selector machines.

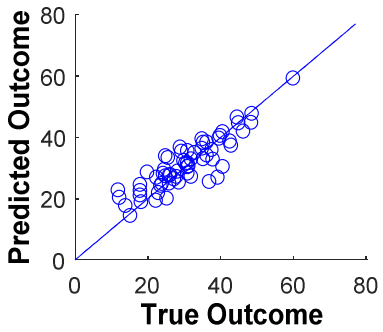


Fig. 4. Plot of outcome prediction by LOLIMOT with features selected by GA or ACO.

Importantly, we saw that not using imaging information lowers performance to around 4.5 (in prediction error), which is likely not a clinically significant degradation. In our other investigations we also found that conventional imaging measures do not correlate well with clinical measures [58] nor improve prediction [8]. However, radiomics analysis of DAT SPECT images, going beyond conventional imaging measures, was seen to provide significant improvements in both tasks. Further, in our recent ongoing efforts involving deep learning based prediction of outcome [59,60], significant improvements were observed, involving implicit discovery of patterns in images. In other words, our present research indicates that there is a need to move beyond conventional imaging metrics for improved prediction of outcome.

IV. CONCLUSIONS

This work explored a range of predictor machines and also aimed to find the best combinations of features to result in improvements in prediction of outcome in PD. Overall, we demonstrated that for our patient studies, absolute errors of the order of ~ 4 could be reached in prediction of motor outcome (UPDRS-III), in comparison to prior works reaching errors of the error of 9 when using readily available features, underlining the importance of optimization (e.g. using Q learning and appropriate feature selection as pursued in this work).

REFERENCE

[1] Z. M. ChenShi, "Exploring the Effects of GeneticVariantson Clinical Profiles of Parkinson's Disease Assessed by the Unified Parkinson's Disease RatingScaleandtheHoehn–YahrStage," *PLOSONE,* pp. 1-11, 2016.

[2] D. Brooks, V. Ibanez, and et al, "Differing Patterns of Striatal F-18 Dopa Uptake in Parkinsons-Disease, Multiple System Atrophy, and Progressive Supranuclear Palsy. Ann Neurol," *Ann Neurol,* vol. 28, pp. 547-555, 1990.

[3] E. Garnett, A. Lang and et al, "A Rostrocaudal Gradient for Aromatic Acid Decarboxylase in the Human Striatum," *Canadian Journal of Neurological Sciences,* no. 14, pp. 444-447, 1987.

[4] A. Stoessl, . W. Martin and et al, "Advances in imaging in Parkinson's disease," *Lancet Neurology,* vol. 10, pp. 987-1001, 2011.

[5] JankovicJ, ".Parkinson'sdisease:clinicalfeaturesanddiagnosis," *JNeurolNeurosurgPsychiatry,* vol. 79, no. 4, pp. 368-376, 2008.

[6] K. Marek, D. Jennings and et al, "Biomarkers for Parkison's disease: tools to assess Parkinson's disease onset and progression," *Ann Neurol,* no. 64, pp. S111-S121, 2008.

[7] Parkinson Progression Marker Initiative, "The Parkinson Progression Marker Initiative (PPMI)," *Prog Neurobiol,* no. 95, pp. 629-635, 2011.

[8] A. Rahmim, P. Huang and et al., "Improved prediction of outcome in Parkinson's disease using radiomics analysis of longitudinal DAT SPECT images," *NeuroImage: Clinical,* vol. 16, pp. 539-544, 2017.

[9] B. Post, J. Speelman and R. de Haan, "Clinical heterogeneity in newly diagnosed," *Journal of neurology,* no. 255, pp. 716-722, 2008.

[10] I. Fyfe, "Prediction of cognitive decline in PD," *Nature Reviews Neurology,* no. 14, pp. 213-317, 2018.

[11] D. Arnaldi, F. De Carli and et al, "Prediction of cognitive worsening in de novo Parkinson's disease: Clinical use of biomarkers.," *Mov Disord,* vol. 32, pp. 1738-1747, 2017.

[12] C. Gao and et al, "Model-based and Model-free Machine Learning Techniques for Diagnostic Prediction and Classification of Clinical Outcomes in Parkinson's Disease," *SCIENTIFIC RePorTS,* pp. 1-21, 2018.

[13] A. Nieuwboer, W. D. Weerdt and et al, "Prediction of outcome of physiotherapy in advanced Parkinson's disease," *SAGE Journals,* vol. 16, no. 8, pp. 886-893, 2002.

[14] S. Grill, J. Weuve and . M. Weisskopf, "Predicting Outcomes in Parkinson's Disease: Comparison of Simple Motor Performance Measures and the Unified Parkinson's Disease Rating Scale-III," *Journal of Parkinson's Disease,* vol. 1, pp. 287-298, 2011.

[15] A. K. Tiwari, "MACHINE LEARNING BASED APPROACHES FOR PREDICTION OF PARKINSON'S DISEASE," *Machine Learning and Applications: An International Journal (MLAIJ),* no. 3, pp. 33-39, 2016.

[16] Y. Singh, P. Bhatia and et al, "A REVIEW OF STUDIES ON MACHINE LEARNING TECHNIQUES," *International Journal of Computer Science and Security,* vol. 1, no. 1, pp. 70-84, 2007.

[17] "Parkinson Progression Marker Initiative," 2011.

[18] N. Oliver, . F. Alexander and I. Rolf , "Local Linear Model Trees (LOLIMOT) Toolbox for Nonlinear System Identification," *science Direct (IFAC System Identification),* vol. 33, no. 15, pp. 845-850, 2000.

[19] J. D. Mart´ınez-Morales and E. V. Palacios, "Modeling of internal combustion engine emissions by," *SciVerse Science Direct,* vol. 3, pp. 251-258, 2012.

[20] Y. Arora, . A. Singhal and A. Bansal, "A Study of Applications of RBF Network," *International Journal of Computer Applications,* vol. 94, no. 2, pp. 17-20, 2014.

[21] S. Alsmadi, . M. khalil and et al, "Back Propagation Algorithm: The Best Algorithm," *IJCSNS International Journal of Computer Science and Network Security,* vol. 9, no. 4, pp. 378-383, 2009.

[22] D. E. Rumelhart, E. Geoffrey and et al, "Leaner Representations By back-Propagating errors," *Nature,* vol. 323, no. 9, pp. 533-536, 1986.

[23] B. Efron, T. Hastie and et al, "Least angle regression," *The Annals of Statistics,* vol. 32, pp. 407-499, 2004.

[24] V. Fonti, "Feature Selection using LASSO," VU Amsterdam, Amsterdam, 2017.

[25] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, p. 5–32, 2001.

[26] A. Jehad, R. Khan and M. Ahmad Nasir , "Random Forests and Decision Trees," *IJCSI International Journal of Computer Science Issues,* vol. 9, no. 5, pp. 272-278, 2012.

[27] A. Townley, M. Ilchmann and et al, "Existence and Learning of Oscillations in Recurrent Neural Networks," *IEEE TRANSACTIONS ON NEURAL NETWORKS,* vol. 11, no. 1, pp. 205-214, 2000.

[28] M. Nijole and . V. Aleksan, "Investigation of financial market prediction by recurrent neural network," *Innovative Infotechnologies for Science, Business and Education,* vol. 11, no. 2, pp. 3-8, 2011.

[29] N. Maknickiene, V. Rutkauskas and Aleksandr, "Investigation of financial market prediction by recurrent neural network," *Innovative Infotechnologies for Science, Business and Education,* vol. 11, no. 2, pp. 3-8, 2011.

[30] A. E. Efendi, "A simulation study on Bayesian Ridge regression models for several collinearity levels," in *AIP Conference Proceedings*, 2017.

[31] C. M. Bishop, Pattern Recognition and, 1th ed., P. J. K. B. S. Michael Jordan, Ed., New York: Springer Science+Business Media, LLC,, 2007.

[32] M. G. RodneyOD and P. Smyth, "Decision tree design using information theory," *Knowledge Acquisition,* vol. 2, pp. 1-19, 1990.

[33] S. Chourasia, "Survey paper on improved methods of ID3 decision tree," *International Journal of Scientific and Research Publications,* vol. 3, no. 12, pp. 1-4, 2013.

[34] J. Lu, P. Zhao, SCH. Hoi, "Online Passive Aggressive Active Learning and its," *JMLR: Workshop and Conference Proceedings,* vol. 39, pp. 266-282, 2014.

[35] M. Blondel, Y. Kubo and N. Ueda, "Online Passive-Aggressive Algorithms for Non-Negative Matrix Factorization and Completion," *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics, PMLR,* vol. 33, pp. 96-104, 2014.

[36] S. H. Shah, A. Rehman and T. Rashid, "A Comparative Study of Ordinary Least Squares Regression and Theil-Sen Regression through Simulation in the Presence of Outliers," *Lasbela, U. J.Sci.Techl,* vol. V, pp. 137-142, 2016.

[37] Theil, "A rank-invariant method of linear and polynomial regression analysis. I, II, III," *Nederl. Akad. Wetensch,* vol. 53, pp. 386–392, 521–525, 1397–1412, 1950.

[38] P. K. Sen, "Estimates of the Regression Coefficient Based on Kendall's Tau," *Journal of the American Statistical Association,* vol. 63, no. 324, pp. 1379-1389 , 1968.

[39] R. Fernandes and S. G.leblanc, "Parametric (modified least squares) and non-parametric(Theil-sen linear regressuon for predicting biological parameters in the presence of measurment errors," *Remote Sensing of Environment,* vol. 95, pp. 303-316, 2005.

[40] S. Kar, S. Das and P. K. Ghosh, "Applications of neuro fuzzy systems: A brief review and future outline," *Applied Soft Computing,* vol. 15, p. 243–259, 2014.

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

[42] K. Crammer, O. Dekel and et al, "On-line passive-aggressive algorithms," *Journal of Machine Learning Research,* vol. 7, pp. 551-585, 2006.

[43] M. Blondel, Y. Kubo and N. Ueda , "Online Passive-Aggressive Algorithms for Non-Negative Matrix Factorization and Completion," *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics, PMLR,* vol. 33, pp. 96-104, 2014.

[44] N. Walia, H. Singh and A. Sharma, "ANFIS: Adaptive Neuro-Fuzzy Inference System- A Survey," *International Journal of Computer Applications,* vol. 123, pp. 32-38, 2015.

[45] P. D. CHRISTOPHER J.C.H. WATKINS, "Q,-Learning," *Machine Learning,* vol. 8, pp. 279-292, 1992.

[46] J. McCall, "Genetic algorithms for modelling and optimisation," *Journal of Computational and Applied Mathematics,* vol. 184, p. 205–222, 2004.

[47] M. Mitchell, "Genetic Algorithms: An Overview," *Complexity,* vol. 1, no. 1, pp. 31-39, 1995.

[48] C. Blum, "Ant colony optimization: Introduction and recent trends," *Physics of Life Reviews,* vol. 2, pp. 353-373, 2005.

[49] P. Sivakumar, "A Survey of Ant Colony Optimization," *International Journal of Advanced Research in Computer Science and Software Engineering,* vol. 6, no. 3, pp. 574-578, 2016.

[50] Q. Bai, "Analysis of Particle Swarm Optimization Algorithm," *Computer and Information Science,* vol. 3, pp. 180-184, 2010.

[51] S. Singh, "A Review on Particle Swarm Optimization Algorithm," *International Journal of Scientific & Engineering Research,* vol. 5, no. 4, pp. 551-553, 2014.

[52] W. B. Dolan, . P. T. Cummings and M. D. LeVan, "Process Optimization via Simulated," *AIChE Journal,* vol. 35, pp. 725-736, 1989.

[53] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, "Optimization by Simulated Annealing," *Science, New Series,* vol. 220, pp. 671-680, 1983.

[54] D. KARABOGA and . S. OKDEM, "A Simple and Global Optimization Algorithm for," *Turk J Elec Engin,* vol. 12, pp. 53-60, 2004.

[55] A. Musrrat, P. Millie and A. Ajith , "Simplex Differential Evolution," *Acta Polytechnica Hungarica,* vol. 6, pp. 95-115, 2009.

[56] D. Kalyanmoy, A. P. Associate and A. Sameer , "A Fast and Elitist Multiobjective Genetic Algorithm:," *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION,* vol. 6, pp. 182-197, 2002.

[57] Y. Yusoff, N. Mohd Salihin and A. Mohd Zain, "Overview of NSGA-II for Optimizing Machining Process Parameters," *Procedia Engineering,* vol. 15, p. 3978 – 3983, 2011.

[58] A. Rahmim, Y. Salimpour and et al, "Application of texture analysis to DAT SPECT imaging: Relationship to clinical assessments," *NeuroImage: Clinical,* vol. 12, pp. e1-e9, 2016.

[59] K. H. Leung, M. Salmanpour and et al, "Using deep-learning to predict outcome of patients with Parkinson's disease," in *IEEE Nucl. Sci. Symp. Conf. Record*, Sydney, 2018.

[60] M. P. Adams, B. Yang and et al, "Prediction of outcome in Parkinson's disease patients from DAT SPECT images using a convolutional neural network," in *IEEE Nucl. Sci. Symp. Conf. Record*, Sydeny, 2018.