Original Article

Prognosis Of Idiopathic Parkinsonism Using Support Vector Machine And Random Forest Classifiers

Raghavendra M Devadas[1*], Vani Ashok Hiremani[2]

[1*,2]Assistant Professor, Dept of CSE, School of Engineering, Presidency University, Bengaluru

***Corresponding Author: -** Raghavendra M Devadas

*Assistant Professor, Dept of CSE, School of Engineering, Presidency University, Bengaluru
DOI: 10.47750/pnr.2022.13.S08.571

Abstract

Many of individuals all over the world suffer with Idiopathic Parkinsonism (IP) which is more common in people over 50. Even today, despite numerous technological developments and breakthroughs, early disease identification is still difficult. This calls for the development of machine learning-based automatic methods that assist doctors in accurately identifying this disease in its early stages. This research paper's main goal is to give a thorough analysis of and comparison of the current machine learning methods used for IP detection. In order to compare and determine which of the two classifiers is the most effective and accurate for classifying IP, this paper discusses Support Vector Machine and Random Forest on a dataset. Accuracy and Kappa scores for support vector machine is 85.6% and 0.814. Accuracy 86.45% and kappa score of 0.81 was found in random forest.

1. INTRODUCTION

Our body motions, including speech, are affected by Idiopathic Parkinsonism (IP) or Parkinson's diseases (PD), a neurological condition that worsens with time. In 1817, Dr. James Parkin-Son identified this illness and gave a description of it under the name "Shaking Palsy" [1]. IP is regarded as the second most prevalent neurodegenerative disease among many others, including Alzheimer's disease, degenerative nerve diseases brain cancer, and Epilepsy [2]. Tremor at rest, Rigidity, Akinesia (or bradykinesia), and Postural instability are the four essential traits of PD that the acronym TRAP can be used to group. Additionally, flexed posture and freezing (motor blocks) have been added to the list of traditional characteristics of parkinsonism [3]. IP tremor develops while at rest but subsides with conscious movement, therefore it normally does not interfere with daily activities. Rigidity is the term for increased resistance (stiffness) to passive movement of a patient's limbs. A number of symptoms, including Bradykinesia (slowness of movement), Hypokinesia (decrease in movement amplitude), and Akinesia (lack of typical unconscious movements, such as an arm swing when walking), can be present [4]. Advanced IP can be reliably and easily identified, but successful treatment is difficult to achieve. Additionally, if treatment is started at an advanced stage, it might have a lower chance of stopping IP's progression. In this situation, an early and correct diagnosis of IP is necessary, which aids patients in keeping a high quality of life. There is currently laboratory test that can be used to diagnose IP and track its development. However, for the early identification of IP, rating scales as the Hoehn and Yahr scale (1967), the Unified Parkinson Disease Rating Scale (UPDRS), and its modified version MDS-UPDRS are sometimes employed [5].

Recently, Machine Learning (ML) has become widely employed for diagnosing medical conditions because it's simple to use and highly accurate [6]. In the literature, ML has also been applied to the control of IP. The paper is structured as follows: The thorough literature evaluation of several machine learning methods used for IP detection is described in Section 2, in section 3 paper describes learning techniques used, Section 4 covers research methodology followed in this paper, section 5 will present findings and observations, and section 6 will wrap up the entire body of work.

2. LITERATURE REVIEW

The authors investigated various approaches and found that neural networks offered the greatest performance for the issue [7]. In the task of diagnosing PD, for feature weighting, researchers employed fuzzy C-means clustering, and for classification, they used k-NN. The best k value was chosen after a weighted PD dataset was submitted to a k-NN classifier for various k values [8]. Voice features were employed by the authors in [9]. They used feature augmentation to produce 177 features from the dataset's 44 features. After feature augmentation, ReliefF was used to filter out the most useful characteristics, and 66 features in total were employed to classify PD. For PD prediction and feature subset construction from the entire feature space, authors employed a fuzzy k-nearest neighbour technique using Principal Component Analysis. According to the authors, their suggested method performed better than the other ways in the literature [10]. For Feature Selection (FS) to be employed for ML in brain surgery, researchers analysed the publications. During PD brain surgery, the real area of the brain to operate on is determined using an ML-based method [11]. SVM was used for classification. In contrast to the others, PD was also treated using an unsupervised strategy [12]. Authors attempted to predict PD using motion data collected from human upper limbs. Both PD patients and healthy volunteers served as the experiment's experimental subjects. The researchers had participants do a variety of performance tests while wearing a gadget on their upper limbs [13]. Extreme Learning Machines were utilised by the authors to diagnose PD. A weighted

approach plus a non-linear kernel function mapping improved imbalanced data. For FS and parameter optimization, ABC method was utilised [14]. Researchers successfully diagnosed Parkinson's disease using PCA (Principal Component Analysis) for dimension reduction, FDR (Fisher Discriminant Ratio) for FS, and SVM for classification [15]. Machine learning techniques are also used for ranking of software requirements [16].

3 MACHINE LEARNING TECHNIQUES

3.1 Support Vector Machine (SVM)

One of the well-known machine learning methods is SVM. Boser, Guyon, and Vapnik initially presented SVM at COLT-92 in 1992. SVM is used in conjunction with a number of related supervised learning techniques for classification and regression [17]. These machines are a part of the generalized linear classifier family. The goal of a support vector machine's basic design is to maximise the space between items belonging to various classes. When the classes to which the elements belong are known at the outset, the problem is referred to as classification. The data set used to determine the class boundary limit is referred to as the training set, and the data set used to evaluate the effectiveness of the method is referred to as the validation set. Numerous other crucial fields, like image processing, pattern recognition, and medical diagnosis technology, also utilise this technique.

In a handwriting identification test, SVM's accuracy is equivalent to that of other well-known modelling techniques like neural networks with extended features, making it more crucial when using pixel maps as input [18]. Due to its numerous difficult characteristics and improved empirical performance, Vapnik's Support Vector Machines (SVM) are producing good results [19]. SVM fundamentally applies the Structural Risk Minimization (SRM) concept, which is superior to the common Empirical Risk Minimization (ERM) approach employed by conventional neural networks [20].

3.2 Random Forest

Random Forest is a decision tree-based classification technique that combines a variety of tree predictors. The distribution of a vector's values across all the trees in the forest is the same, and each tree depends on them separately. As the number of trees in the forest increases, error with generalisation converges. The strength of the individual trees in the forest and their correlation with one another are the two main factors that affect the inaccuracy associated with this classifier's model. Even with more noise present in the training set, Random Forest performs well. The underlying workings of this method produce improved internal estimates that monitor correlation, error, and strength. The effect to raising the number of characteristics employed in the data splitting is then demonstrated using these [21].

4. METHODOLOGY

The below Fig. 4.1 illustrates the methodology followed in this work.

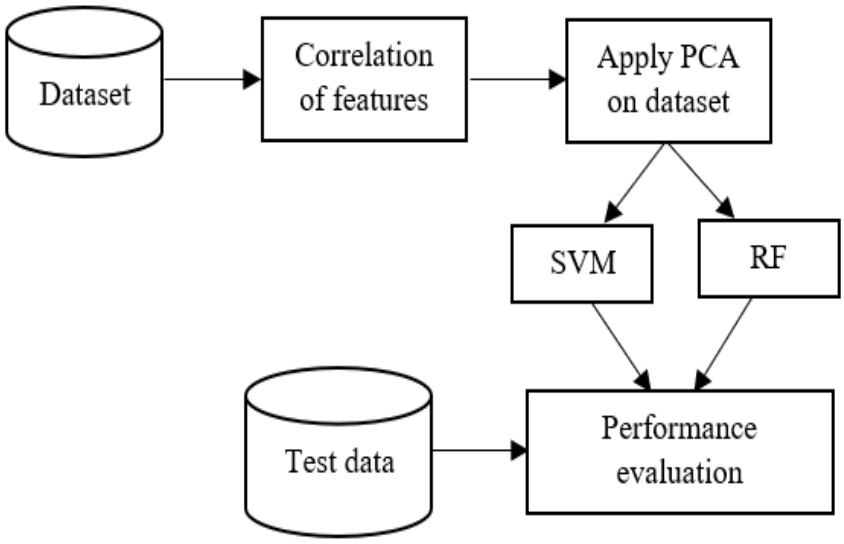


Figure 4.1 Research methodology followed

4.1 Dataset Description

The Parkinsons Data Set from the UCI Machine Learning Repository was used in this study. The table 1 below illustrates the characteristics of the dataset used in this work.

Table 1. Dataset Description

Data Set Characteristics	Multivariate
Number of Instances	197
Attribute Characteristics	Real
Number of Attributes	23
Task	Classification

Some of the attribute information is as shown below.

Matrix column entries (attributes):
name - ASCII subject name and recording number
MDVP:Fo(Hz) - Average vocal fundamental frequency
MDVP:Fhi(Hz) - Maximum vocal fundamental frequency
MDVP:Flo(Hz) - Minimum vocal fundamental frequency
MDVP:Jitter(%),MDVP:Jitter(Abs),MDVP:RAP,MDVP:PPQ,Jitter:DDP - Several measures of variation in fundamental frequency
MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,MDVP:APQ,Shimmer:DDA - Several measures of variation in amplitude
NHR,HNR - Two measures of ratio of noise to tonal components in the voice
status - Health status of the subject (one) - Parkinson's, (zero) - healthy
RPDE,D2 - Two nonlinear dynamical complexity measures
DFA - Signal fractal scaling exponent
spread1,spread2,PPE - Three nonlinear measures of fundamental frequency variation

At first glance, the Parkinson's Disease Dataset reveals,

- The Parkinson's Dataset has no null values.
- The dataset's record inputs are all distinct.
- There are 147 Parkinson's patients and 48 healthy people.

4.2 Correlation of features

Any statistical relationship between two random variables is known as correlation (or dependence) in statistics. The strength of the linear link between two variables can also be used to define it. The Spearman correlation coefficient is the sort of correlation coefficient employed here. The Spearman correlation coefficient is calculated using the Pearson's correlation coefficient between rank variables. Below is a formula for Pearson's correlation coefficient.

$$p = \frac{n(\Sigma\alpha\beta)-(\Sigma\alpha)(\Sigma\beta)}{\sqrt{[n\Sigma\alpha^2-(\Sigma\alpha)^2][n\Sigma\beta^2-(\Sigma\beta)^2]}} \qquad \text{Eqn (1)}$$

Where, n is the sample size, $'\alpha', '\beta'$ are the raw scores in the sample data Spearman Correlation Coefficient Formula is as follows,

$$p_s = \rho_{pg_\alpha, pg_\beta} = \frac{cov(pg_\alpha, pg_\beta)}{\sigma_{rg_\alpha}, \sigma_{pg_\beta}} \qquad \text{Eqn (2)}$$

Where 'ρ' represents Pearson Correlation Coefficient applied to rank variables, cov (p α, pg β) is the covariance of the rank variable σpgα and σpgβ are the standard deviations of the rank variables.

The correlation plot between attributes using the generated correlation matrix is shown in Fig 4.2. Also, Fig 4.3 shows the plot of correlation between attributes by including both correlation values and p-values in the correlation plot in order to gain more understanding.

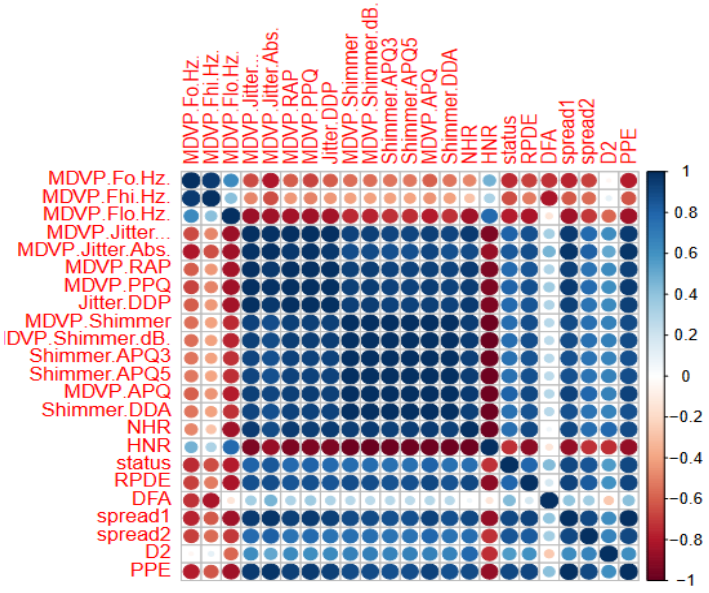


Figure 4.2 Correlation among attributes

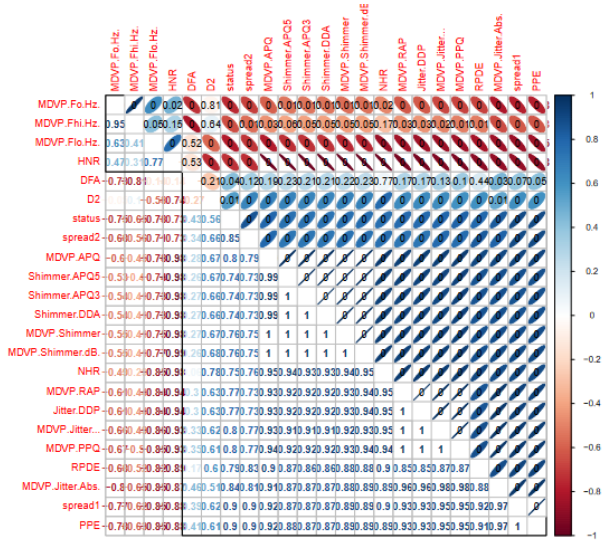


Figure 4.3 Plot of correlations with p-values and corr-values

4.3 Principal Component Analysis

A large number of (possibly linked) variables is reduced to a more manageable number of uncorrelated variables known as Principal Components through the mathematical process of Principle Component Analysis (PCA). It is an analytical technique that entails iteratively identifying the linear combination of a set of variables with the greatest variation, eliminating it, and repeating the process.

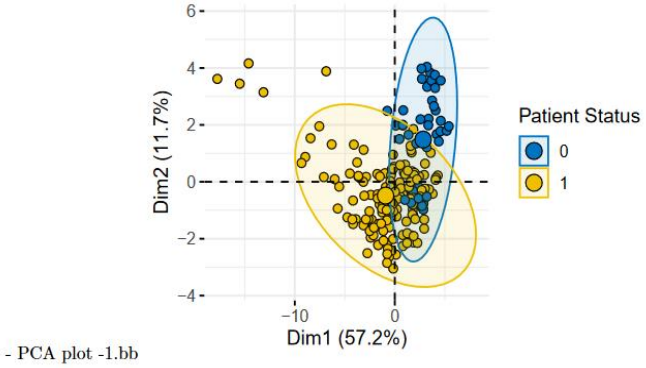
The process of separating principal components from a raw dataset can be stated in six steps:

- Ignore the labels and take the entire dataset with d+1 dimensions, making our new dataset d dimensional.
- Determine the mean for each dimension throughout the whole dataset.
- Create a matrix of covariance for the entire dataset.
- Create the corresponding eigenvalues and eigenvectors.
- To create a d × k dimensional matrix W, order the eigenvectors by decreasing eigenvalues and select the k eigenvectors with the highest eigenvalues.
- By use of this d × k eigenvector matrix, the samples should be projected into the new subspace.

Below is a snapshot of the Principal Component Analysis performed on the dataset.

```
## Importance of components:
##                           PC1     PC2     PC3     PC4     PC5     PC6     PC7
## Standard deviation     3.6256 1.6410 1.25590 1.21260 1.00533 0.85649 0.80032
## Proportion of Variance 0.5715 0.1171 0.06858 0.06393 0.04394 0.03189 0.02785
## Cumulative Proportion  0.5715 0.6886 0.75719 0.82113 0.86507 0.89696 0.92481
##                           PC8     PC9     PC10     PC11     PC12     PC13     PC14
## Standard deviation     0.66946 0.59816 0.53667 0.47149 0.37331 0.32377 0.26406
## Proportion of Variance 0.01949 0.01556 0.01252 0.00967 0.00606 0.00456 0.00303
## Cumulative Proportion  0.94430 0.95985 0.97238 0.98204 0.98810 0.99266 0.99569
##                           PC15     PC16     PC17     PC18     PC19     PC20     PC21
## Standard deviation     0.18947 0.14777 0.13253 0.11150 0.08288 0.05868 0.03288
## Proportion of Variance 0.00156 0.00095 0.00076 0.00054 0.00030 0.00015 0.00005
## Cumulative Proportion  0.99725 0.99820 0.99896 0.99950 0.99980 0.99995 1.00000
##                           PC22      PC23
## Standard deviation     0.0006015 0.000182
## Proportion of Variance 0.0000000 0.000000
## Cumulative Proportion  1.0000000 1.000000
```

Figure 4.4 shows a 2D-Plot for PCA using a 23-feature dataset.



- PCA plot -1.bb

Figure 4.5 2D-Plot for PCA

Calculating the eigenvalues, variance proportion, and cumulative variance proportion for various dimensions or principle components is shown below.

```
##          eigenvalue variance.percent cumulative.variance.percent
## Dim.1  1.314527e+01     5.715333e+01                     57.15333
## Dim.2  2.692943e+00     1.170845e+01                     68.86178
## Dim.3  1.577273e+00     6.857709e+00                     75.71949
## Dim.4  1.470409e+00     6.393083e+00                     82.11257
## Dim.5  1.010689e+00     4.394301e+00                     86.50687
## Dim.6  7.335692e-01     3.189431e+00                     89.69631
## Dim.7  6.405124e-01     2.784837e+00                     92.48114
## Dim.8  4.481805e-01     1.948611e+00                     94.42975
## Dim.9  3.577979e-01     1.555643e+00                     95.98540
## Dim.10 2.880117e-01     1.252225e+00                     97.23762
## Dim.11 2.223062e-01     9.665486e-01                     98.20417
## Dim.12 1.393597e-01     6.059116e-01                     98.81008
## Dim.13 1.048291e-01     4.557785e-01                     99.26586
## Dim.14 6.972919e-02     3.031704e-01                     99.56903
## Dim.15 3.589816e-02     1.560790e-01                     99.72511
## Dim.16 2.183532e-02     9.493616e-02                     99.82004
## Dim.17 1.756358e-02     7.636340e-02                     99.89641
## Dim.18 1.243327e-02     5.405769e-02                     99.95047
## Dim.19 6.868404e-03     2.986262e-02                     99.98033
## Dim.20 3.443165e-03     1.497028e-02                     99.99530
## Dim.21 1.080936e-03     4.699721e-03                    100.00000
## Dim.22 3.618178e-07     1.573121e-06                    100.00000
## Dim.23 3.312204e-08     1.440088e-07                    100.00000
```

5. RESULTS

The experiment uses the Parkinsons Data Set [22] from the UCI Machine Learning Repository. Random Forest and support vector machines are used to classify the data set. A comparison of prediction performances using both techniques is used to analyze the results. R language is used for the implementation as it provides plethora of packages for conducting various ML experiments. Few of the packges used for this study are as follows: dplyr, corrplot, mlbench, caret.

5.1 Evaluation of classifiers

The training and test data were divided into 80% and 20% each after 10 cross validations and 3 repeats. Accuracy and Kappa score measures are used for performance evaluation and Table 2 summarizes the findings using the Parkinson on both SVM and Random Forest approaches.

Table 2. Accuracy and Kappa scores of SVM, Random Forest classifiers

Classifier	Accuracy (%)	Kappa Score
SVM	85.6	0.814
Random Forest	86.45	0.81

6 CONCLUSION AND FUTURE WORK

According to the results, the classification accuracy of the random forest classifier can be compared to that of SVMs. The random forest classifier also has the advantage of only requiring two parameters to be configured, as opposed to many user-defined parameters for SVMs. In contrast to SVMs, the random forest classifier can handle categorical data, imbalanced data, and data with missing values. In future various feature selction techniques can be applied for selecting required features which may result in improvement in accuracy percentage.

REFERENCES

1. J. Parkinson, An Essay on Shaking Palsy. London: Whitting-ham and Rowland Printing, 1817.
2. D. B. Calne, “Is idiopathic parkinsonism the consequence of an event or a process,” Neurology, Vol. 44, no. 15, pp. 5–5, 1994.
3. Jankovic J, Parkinson’s disease: clinical features and diagnosis Journal of Neurology, Neurosurgery & Psychiatry 2008;79:368-376.
4. William Dauer, Serge Przedborski, Parkinson's Disease: Mechanisms and Models, Neuron,
5. Volume 39, Issue 6, 2003, Pages 889-909, ISSN 0896-6273, https://doi.org/10.1016/S0896-6273(03)00568-3.
6. C. G. Goetz, et al., “Movement disorder society-sponsored revision of the unified Parkinson’s disease rating scale (MDS-UPDRS): Scale presentation and clinimetric testing results,” Mov. Disord., Vol. 23, no. 15, pp. 2129–70, 2008.
7. Rayan Z, Alfonse M, Salem A-BM. Machine learning approaches in smart health. Procedia Comput Sci 2019;154:361–8. https://doi.org/10.1016/j.procs.2019.06.052.
8. Das R. A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Syst Appl 2010;37(2):1568–72.
9. Polat K. Classification of Parkinson’s disease using feature weighting method on the basis of fuzzy C-means clustering. Int J Syst Sci 2012;43(4):597–609.
10. Yaman O, Ertam F, Tuncer T. Automated Parkinson’s disease recognition based on statistical pooling method using acoustic features. Med Hypotheses 2019:109483.
11. H.-L. Chen, C-C. Huang, X-G. Yu, X. Xu, X. Sun, G. Wang, and S-J. Wang, “An efficient diagnosis system for detection of Parkinson’s disease using fuzzy k-nearest neighbor approach,” Expert Syst. Appl., Vol. 40, pp. 263–71, 2013.
12. Wan KR, Maszczyk T, See AAQ, Dauwels J, King NKK. A review on microelectrode recording selection of features for machine learning in deep brain stimulation surgery for Parkinson’s disease. Clin Neurophysiol 2019;130(1):145–54.
13. Nilashi M, Ibrahim O, Ahmadi H, Shahmoradi L, Farahmand M. A hybrid intelligent system for the prediction of Parkinson’s Disease progression using machine learning techniques. Biocybern Biomed Eng 2018;38(1):1–15.
14. Cavallo F, Moschetti A, Esposito D, Maremmani C, Rovini E. Upper limb motor preclinical assessment in Parkinson’s disease using machine learning. Park Relat Disord 2019;63:111–6.
15. Wang Y, Wang AN, Ai Q, Sun HJ. An adaptive kernel-based weighted extreme learning machine approach for effective detection of Parkinson’s disease. Biomed Signal Process Control 2017;38:400–10.

16. Singh G, Vadera M, Samavedham L, Lim ECH. Machine learning-based framework for multi-class diagnosis of neurodegenerative diseases: a study on Parkinson's Disease. IFAC-Papers OnLine 2016;49(7):990–5.
17. Raghavendra M Devdas, Nagaraj Cholli. Interdependency aware QUBIT and BROWNBOOST rank for large scale requirement prioritization. International Journal of Computing and Digital Systems. Vol 11, Issue 1, pp. 625-634.
18. Yao, Z.; Lei, L.; Yin, J., "R-C4.5 Decision tree model and its applications to health care dataset". Proceedings of International Conference on Services Systems and Services Management 2005, pp. 1099-1103.
19. Burke HB, Goodman PH, Rosen DB, et al. 1997. Artificial neural networks improve the accuracy of cancer survival prediction. Cancer, 79:857-62.
20. Leenhouts HP. 1999. Radon-induced lung cancer in smokers and nonsmokers: risk implications using a two-mutation carcinogenesis model. Radiat Environ Biophys, 1999 38:57-71.
21. E. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, pages 144{152. ACM Press, 1992.
22. Michie, D. J. Spiegelhalter, C. C. Taylor, and J. Campbell, editors. Machine learning, neural and statistical classi_cation. Ellis Horwood, Upper Saddle River, NJ, USA, 1994. ISBN 0-13-106360-X. Data available at http://archive.ics. uci.edu/ml/machine-learning-databases/statlog/.
23. 'Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection', Little MA, McSharry PE, Roberts SJ, Costello DAE, Moroz IM. BioMedical Engineering OnLine 2007, 6:23 (26 June 2007)