*International Journal of Luminescence and Applications (ISSN: 2277-6362)*
*Vol. 7, No. 3-4, October - December 2017. Article ID: 252. pp.475-479.*

# Soft Computing Models for Parkinson's Disease Diagnosis using Gait Features

**Chandra Prakash Rathore[1], Rekh Ram Janghel[2], Kshitiz Verma[3], Swati Rathore[4*]** and **Saroj Kumar Pandey[2]**

[1]*Department of Information Technology, Dr. C. V. Raman University, Kota, Bilaspur, Chhattisgarh, India.*
[2]*Department of Information Technology, National Institute of Technology, Raipur, Chhattisgarh, India.*
[3]*University Teaching Department, Chhattisgarh Swami Vivekananda Technical University, Bhilai, Chhattisgarh, India.*
[4]*Department of Computer Application, Chhattisgarh Swami Vivekananda Technical University, Bhilai, Chhattisgarh, India.*

***Abstract—*** *Parkinson's disease is a very common disease among elder population effecting approx 6.3 million people worldwide across all genders, races and cultures worldwide. It reduces quality of life of the patients due to motor and non-motor complications. In current paper four soft computing models i.e. Classification Tree, Cubist, Naïve Bayes and Linear Regression models are implemented for Parkinson's disease diagnosis using gait features. Performances of these soft computing models are then evaluated based on performance measures viz. true positive, false positive, false negative, true negative, accuracy, sensitivity, specificity and RMSE to identify the best performing soft computing model to diagnose the disease. Comparison of average accuracy, sensitivity, specificity and RMSE of the soft computing models over 5-rounds shown that Cubist model outperformed among others with average accuracy, sensitivity, specificity and RMSE being 86.336%, 89.480%, 81.934%, 0.318 respectively on training datasets, and 69.720%, 73.632%, 64.242%, 0.504 respectively on testing datasets.*

***Keywords—*** *Soft Computing; Cubist; Classification Tree; Naïve Bayes; Linear Regression*

## 1. INTRODUCTION

Parkinson's disease is a very common disease among elder population effecting approx 6.3 million people worldwide across all genders, races and cultures worldwide [1]. It reduces quality of life of the patients and increases economic burden on the patients and healthcare system of the country. Currently there is no cure available for the disease, but, its symptoms can be controlled and quality of life can be improved by taking various available treatments. Although, there's no precise test available today to diagnose Parkinson's disease, but, doctors attempt to diagnose it based on the disease symptoms and medical history. Hoehn and Yahr (H&Y) Scale, Schwab and England Activities of Daily Living (ADL) Scale, and Unified Parkinson's Disease Rating Scale (UPDRS) are scales to measure the disease severity. Soft computing techniques have played a vital role in diagnosing several diseases and have been helpful in diagnosing Parkinson's Disease as well using various symptoms of the disease like voice characteristics, gait and hand writing patterns of Parkinson's patients [2].

## 2. LITERATURE REVIEW

Study shows that gait features are very useful in diagnosing Parkinson's disease. Salarian et al. found that spatio-temporal parameters of gait such as gait cycle time, double support, stance, stride velocity, and stride length have high correlation with UPDRS subscores [3]. Cho et al. proposed a vision-based diagnostic system which employed combination of principal component analysis and linear discriminant analysis to differentiate different categories of gait simultaneously. Their experiment showed that linear discriminant analysis can recognize Parkinsonian gait [4]. Li. et al. developed a recognition system based on local linear embedding algorithm to extract and recognize the gait features from the information provided by 16-node body sensor network. The system was able to recognize gait patterns of Parkinson's disease successfully with higher recognition rate than principal component analysis [5]. Chen et al. implemented a computer vision-based gait analysis approach with kernel-based principal component analysis to support clinical assessments of Parkinson's disease by quantitatively determining gait cycle, stride length, stride velocity and cadence [6]. Klucken et al. proposed a mobile biosensor based Embedded Gait Analysis using Intelligent Technology (eGaIT) system to classify specific stages and motor symptoms in Parkinson's disease automatically and objectively [7].

## 3. SOFT COMPUTING MODELS

Four soft computing models have been used in the current research work which is briefly explained in following subsections.

* *Corresponding Author: rathore_1st@yahoo.com*

### 3.1 Classification Tree (CT) Model

It uses decision tree based machine learning algorithm to build a prediction model from training data when labels of class are known in advance. The decision tree has three types of nodes i) root node, which has no incoming edges, but, can have 0 or more outgoing edges ii) internal node, which has exactly one incoming edge, but, can have two or more outgoing edges, iii) terminal nodes, each of which has exactly one incoming edge with no outgoing edges and class label assigned to these nodes. CT model is built following splitting rule and in order to get the best split the data is divided into two parts with maximum homogeneity [8].

### 3.2 Cubist Model

It is a rule based model where tree is converted to a set of rules which initially are paths from the top of the tree to the bottom. It yields better response than those generated by simple techniques like multivariate linear regressions. A tree is grown having intermediate terminal nodes and terminal leaves each of which contains a regression model. The linear regression model at terminal node of the tree does prediction, but smoothed by taking into account prediction from the linear model in previous node of the tree [9].

### 3.3 Naïve Bayes (NB) Model

It is a very efficient classifier model based on Bayes probability theorem with assumption that attributes in a dataset is mutually independent. It can deal with real, discrete, streaming data and is useful in field of automated disease diagnosis. According to [10], assuming that Y is any discrete-valued response variable and attributes $X_1 \dots X_n$ are any discrete or real-valued attributes with the objective of training a classifier that will provide the probability distribution over possible values of Y for each new instance X. The probability of kth value of Y according to Bayes rules is given by equation (1).

$$P(Y = y_k | X_1 \dots X_n) = \frac{P(Y=y_k)P(X_1 \dots X_n | Y=y_k)}{\sum_j P(Y=y_j)P(X_1 \dots X_n | Y=y_j)} \quad (1)$$

### 3.4 Linear Regression (LR) Model

It models relationship between a scalar dependent variable y and one or more independent variables X. There are several types of LR model exist out of which Binomial logistic regression model is used in current research work. In this model likelihood $l(\beta)$ of the predictions is given by equation (2) [11].

$$l(\beta) = \sum_{i=1}^{n}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \quad (2)$$

Where, $\beta$ is a regression parameter, n is a number of responses, $y_i$ is a response, $p_i$ is a probability of $y_i$

## 4. EXPERIMENTAL RESULTS

The experiment starts with gait dataset collection of PD patients and healthy controls. The dataset is obtained from PhysioNet [12] which is created by Hausdorff et al. [13]. The dataset consists of 1355 instances of gait data taken from 29 PD patients and 25 healthy controls during walking. Each instance of the dataset consists of 19 gait features (specified in Table 1) about the subject and a classifier label which denotes status of the subject i.e. PD patient or the healthy control. The dataset is divided into two subsets i.e. training dataset consisting of 1104 instances and testing dataset consisting of 251 instances. The training and testing datasets are used for training and testing of the soft computing models respectively. First of all specified four soft computing models are trained, then, evaluation is done on the training and testing datasets based on performance measures viz. True Positive (TP), False Positive (FP), False Negative (FN), True Negative (TN), accuracy, sensitivity, specificity and Root Mean Square Error (RMSE). This evaluation is done over five folds where in each fold training and testing datasets are recreated taking different instances from full gait dataset. Finally, performances of the soft computing models are compared based on average accuracy, sensitivity, specificity and RMSE yielded over 5 folds. The comparison exhibited that Cubist model outperformed among others with average accuracy, sensitivity, specificity, RMSE being 86.336%, 89.480%, 81.934%, 0.318 respectively on training dataset and 69.720%, 73.632%, 64.242%, 0.504 respectively on testing dataset.

**Table 1: Description of gait features collected from the subjects**

| *Column No.* | *Feature* |
|---|---|
| 1 | Sampling Time |
| 2-9 | Vertical Ground Reaction Force (in Newton) on each of 8 sensors located under the left foot |
| 10-17 | Vertical Ground Reaction Force (in Newton) on each of 8 sensors located under the right foot |
| 18 | Total force under the left foot |
| 19 | Total force under the right foot |

**Table 2: Performance matrix of the soft computing models on training datasets**

| *Round* | *Method* | *TP* | *FP* | *FN* | *TN* | *Accuracy(%)* | *Sensitivity (%)* | *Specificity(%)* | *RMSE* |
|---|---|---|---|---|---|---|---|---|---|
| R-1 | CT | 499 | 103 | 87 | 315 | 81.08 | 85.15 | 75.36 | 0.38 |
| | Cubist | 527 | 59 | 59 | 359 | 88.25 | 89.93 | 85.89 | 0.31 |
| | LR | 449 | 200 | 137 | 218 | 66.43 | 76.62 | 52.15 | 0.45 |
| | NB | 351 | 139 | 235 | 279 | 62.75 | 59.90 | 66.75 | 0.61 |
| R-2 | CT | 495 | 102 | 91 | 316 | 80.78 | 84.47 | 75.60 | 0.38 |
| | Cubist | 508 | 81 | 78 | 337 | 84.16 | 86.69 | 80.62 | 0.34 |
| | LR | 445 | 212 | 141 | 206 | 64.84 | 75.94 | 49.28 | 0.46 |
| | NB | 376 | 157 | 210 | 261 | 63.45 | 64.16 | 62.44 | 0.60 |
| R-3 | CT | 506 | 152 | 80 | 266 | 76.89 | 86.35 | 63.64 | 0.41 |
| | Cubist | 546 | 74 | 40 | 344 | 88.65 | 93.17 | 82.30 | 0.29 |
| | LR | 450 | 205 | 136 | 213 | 66.04 | 76.79 | 50.96 | 0.45 |
| | NB | 378 | 158 | 208 | 260 | 63.55 | 64.51 | 62.20 | 0.60 |
| R-4 | CT | 505 | 140 | 80 | 279 | 78.09 | 86.32 | 66.59 | 0.40 |
| | Cubist | 540 | 72 | 45 | 347 | 88.35 | 92.31 | 82.82 | 0.30 |
| | LR | 454 | 210 | 131 | 209 | 66.04 | 77.61 | 49.88 | 0.46 |
| | NB | 384 | 163 | 201 | 256 | 63.75 | 65.64 | 61.10 | 0.60 |
| R-5 | CT | 509 | 165 | 76 | 254 | 76.00 | 87.01 | 60.62 | 0.42 |
| | Cubist | 499 | 92 | 86 | 327 | 82.27 | 85.30 | 78.04 | 0.35 |
| | LR | 442 | 209 | 143 | 210 | 64.94 | 75.56 | 50.12 | 0.46 |
| | NB | 364 | 157 | 221 | 262 | 62.35 | 62.22 | 62.53 | 0.61 |
| Average | CT | 502.8 | 132.4 | 82.8 | 286.0 | 78.568 | 85.860 | 68.362 | 0.398 |
| | Cubist | 524.0 | 75.6 | 61.6 | 342.8 | 86.336 | 89.480 | 81.934 | 0.318 |
| | LR | 448.0 | 207.2 | 137.6 | 211.2 | 65.658 | 76.504 | 50.478 | 0.456 |
| | NB | 370.6 | 154.8 | 215.0 | 263.6 | 63.170 | 63.286 | 63.004 | 0.604 |

**Table 3: Performance matrix of the soft computing models on testing datasets**

| *Round* | *Method* | *TP* | *FP* | *FN* | *TN* | *Accuracy(%)* | *Sensitivity (%)* | *Specificity(%)* | *RMSE* |
|---|---|---|---|---|---|---|---|---|---|
| R-1 | CT | 109 | 41 | 37 | 64 | 68.92 | 74.66 | 60.95 | 0.48 |
| | Cubist | 103 | 32 | 43 | 73 | 70.12 | 70.55 | 69.52 | 0.51 |
| | LR | 87 | 28 | 59 | 77 | 65.34 | 59.59 | 73.33 | 0.94 |
| | NB | 84 | 44 | 62 | 61 | 57.77 | 57.53 | 58.10 | 0.65 |
| R-2 | CT | 113 | 40 | 33 | 65 | 70.92 | 77.40 | 61.90 | 0.46 |
| | Cubist | 107 | 39 | 39 | 66 | 68.92 | 73.29 | 62.86 | 0.51 |
| | LR | 84 | 27 | 62 | 78 | 64.54 | 57.53 | 74.29 | 0.88 |
| | NB | 99 | 43 | 47 | 62 | 64.14 | 67.81 | 59.05 | 0.60 |
| R-3 | CT | 108 | 42 | 38 | 63 | 68.13 | 73.97 | 60.00 | 0.47 |
| | Cubist | 109 | 41 | 37 | 64 | 68.92 | 74.66 | 60.95 | 0.50 |
| | LR | 74 | 27 | 72 | 78 | 60.56 | 50.68 | 74.29 | 0.95 |
| | NB | 89 | 36 | 57 | 69 | 62.95 | 60.96 | 65.71 | 0.61 |
| R-4 | CT | 102 | 49 | 45 | 55 | 62.55 | 69.39 | 52.88 | 0.50 |
| | Cubist | 110 | 39 | 37 | 65 | 69.72 | 74.83 | 62.50 | 0.51 |
| | LR | 82 | 27 | 65 | 77 | 63.35 | 55.78 | 74.04 | 0.93 |
| | NB | 101 | 43 | 46 | 61 | 64.54 | 68.71 | 58.65 | 0.60 |
| R-5 | CT | 122 | 48 | 25 | 56 | 70.92 | 82.99 | 53.85 | 0.44 |
| | Cubist | 110 | 36 | 37 | 68 | 70.92 | 74.83 | 65.38 | 0.49 |
| | LR | 83 | 29 | 64 | 75 | 62.95 | 56.46 | 72.12 | 0.88 |
| | NB | 81 | 38 | 66 | 66 | 58.57 | 55.10 | 63.46 | 0.64 |
| Average | CT | 110.8 | 44.0 | 35.6 | 60.6 | 68.288 | 75.682 | 57.916 | 0.47 |
| | Cubist | 107.8 | 37.4 | 38.6 | 67.2 | 69.720 | 73.632 | 64.242 | 0.504 |
| | LR | 82.0 | 27.6 | 64.4 | 77.0 | 63.348 | 56.008 | 73.614 | 0.916 |
| | NB | 90.8 | 40.8 | 55.6 | 63.8 | 61.594 | 62.022 | 60.994 | 0.620 |

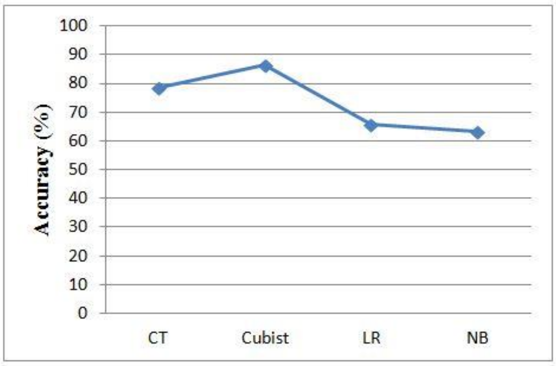


**Fig. 1: Comparison of average accuracy of the coft computing models on training datasets**

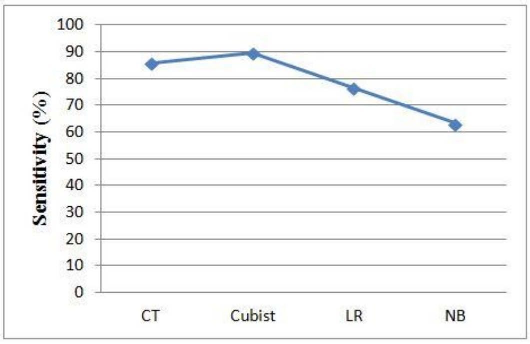


**Fig. 2: Comparison of average sensitivity of the soft computing models on training dataset**

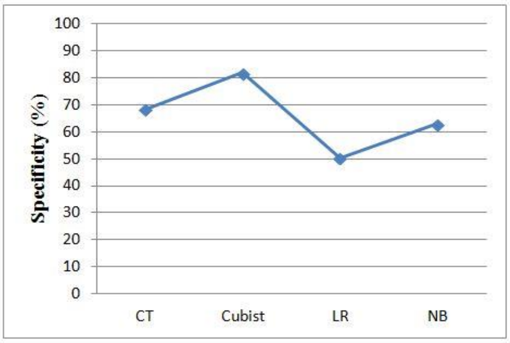


**Fig. 3: Comparison of average specificity of the soft computing models on training datasets**

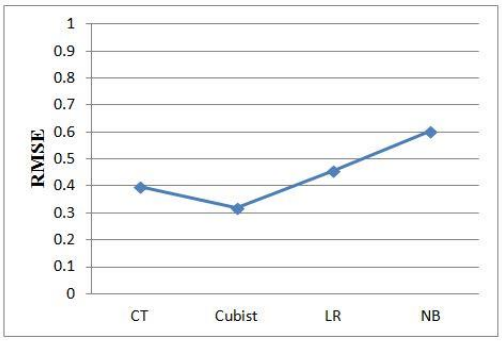


**Fig. 4: Comparison of average RMSE of the soft computing models on training datasets**

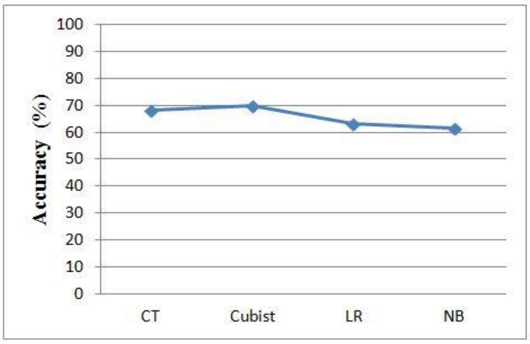


**Fig. 5: Comparison of average accuracy of the soft computing models on testing datasets**

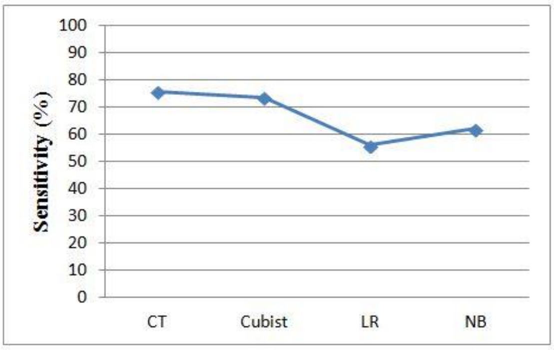


**Fig. 6: Comparison of average sensitivity of the soft computing models on testing datasets**

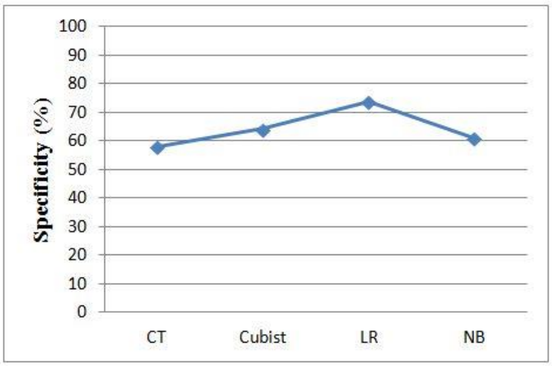


**Fig. 7: Comparison of average specificity of the soft computing models on testing datasets**

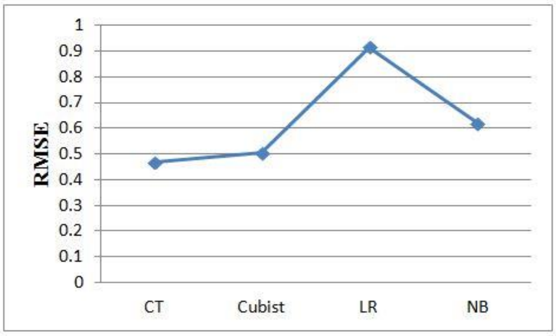


**Fig. 8: Comparison of average RMSE of the soft computing models on testing datasets**

## 5. CONCLUSION

In current paper four soft computing models i.e. Classification Tree, Cubist, Naïve Bayes and Linear Regression model are implemented which are then trained and tested using gait datasets collected from PD patients and the healthy controls. Then, the soft computing models are evaluated based on performance parameters viz. TP, FP, FN, TN, accuracy, sensitivity, specificity and RMSE. Finally, average accuracy, sensitivity, specificity and RMSE of the soft computing models calculated over 5-folds and compared. The comparison exhibited that Cubist model outperformed among others. In future new soft computing models or hybrid ensemble models can be developed to further improve accuracy and reduce RMSE in the disease diagnosis.

## REFERENCES

[1] Medtronic Inc. Parkinson's Disease. Retrieved on 22-Sep-2010 from http://www.medtronic.eu/ yourhealth/parkinsons-disease/

[2] Y. S. Morsi, A. Shukla and C. P. Rathore, Optimizing Assistive Technologies for Aging Populations, IGI Global, PA, USA, 2015.

[3] A. Salarian et al., Gait assessment in Parkinson's disease: Toward an ambulatory system for long-term monitoring. IEEE Transactions on Bio-Medical Engineering, 51(8), 1434–1443, 2004.

[4] C. W. Cho et al., A vision-based analysis system for gait recognition in patients with Parkinson's disease. Expert Systems with Applications, 36(3), 7033–7039, 2009.

[5] S. Li, J. Wang, and X. Wang, A novel gait recognition analysis system based on body sensor networks for patients with Parkinson's disease. Proceedings of IEEE Globecom 2010 workshop on Advanced Sensor Integration Technology, 256 – 260, Miami, FL. IEEE.

[6] S. Chen et al., Quantification and recognition of Parkinsonian gait from monocular video imaging using kernel-based principal component analysis. Biomedical Engineering Online, 10(1), 99, 2011

[7] J. Klucken et al., Unbiased and mobile gait analysis detects motor impairment in Parkinson's disease. PLoS ONE, 8(2), e56956, 2013.

[8] R. Timofeev, Classification and regression trees (CART) theory and applications. Center of Applied Statistics and Economics, pp. 7-39, 2004.

[9] M. Kuhn et al., Cubist models for regression. The Comprehensive R Archive Network, 2012.

[10] T. M. Mitchell, Machine Learning. McGraw-Hill, NY, USA, 1997.

[11] C. J. Geyer, Generalized linear models in R. The Comprehensive R Archive Network, 2003.

[12] A. L. Goldberger et al., Components of a new research resource for complex physiologic signals. Circulation, vol. 101(23), pp. e215-e220, 2013.

[13] J. M. Hausdorff et al., Rhythmic auditory stimulation modulates gait variability in Parkinson's disease. The European Journal of Neuroscience, vol. 26(8), pp. 2369–2375, 2007.