Journal of Pharmaceutical Research International

32(11): 63-73, 2020; Article no.JPRI.58718
ISSN: 2456-9119
(Past name: British Journal of Pharmaceutical Research, Past ISSN: 2231-2919,
NLM ID: 101631759)

Classification of Control and Neurodegenerative Disease Subjects Using Tree Based Classifiers

Syed Ahsin Ali Shah[1], Nazneen Habib[2], Wajid Aziz[1,3*], Ehsan Ullah Khan[1] and Malik Sajjad Ahmed Nadeem[1]

[1]Department of Computer Sciences and Information Technology, University of Azad Jammu and Kashmir, Muzaffarabad, 13100, Pakistan.
[2]Department of Sociology and Rural Development, University of Azad Jammu and Kashmir, Muzaffarabad, 13100, Pakistan.
[3]College of Computer Sciences and Engineering, University of Jeddah, Jeddah 21959, Saudi Arabia.

Authors' contributions

This work was carried out in collaboration among all authors. Authors SAAS, WA and EUK designed the study, performed the analysis, wrote the protocol and wrote the first draft of the manuscript. Author MSAN managed the analyses of the study. Author NH managed the literature searches. Authors NH, WA and MSAN reviewed the manuscript. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JPRI/2020/v32i1130546
Editor(s):
(1) Dr. Barkat Ali Khan, Gomal University, Pakistan.
Reviewers:
(1) Pratiyush Guleria Nielit, India.
(2) Florina Scarlatache, Technical University Gheorghe Asachi Iasi, Romania.
Complete Peer review History: http://www.sdiarticle4.com/review-history/58718

Original Research Article

Received 24 April 2020
Accepted 30 June 2020
Published 16 July 2020

ABSTRACT

Background: The medical researchers are developing different non-invasive methods for early detection of Neurodegenerative Diseases (NDDs) when pharmacological interventions are still possible to further prevent the disease progression. The NDDs are associated with the degradation in the complex gait dynamics and motor activity. The classification of gait data using machine learning techniques can assist the physicians for early diagnosis of the neural disorder when clinical manifestation of the diseases is not yet apparent.
Aims: The present study was undertaken to classify the control and NDD subjects using decision trees based classifiers (Random Forest (RF), J48 and REPTree).
Methodology: The data used in the study comprises of 16 control, 20 Huntington's Disease (HD), 15 Parkinson's Disease (PD), and 13 Amyotrophic Lateral Sclerosis (ALS) subjects, which were taken from publicly available database from Physionet. The age range of control subjects was 20-74, HD subjects was 36-70, PD subjects was 44-80, and ALS subjects was 29-71. There were 13

*Corresponding author: E-mail: kh_wajid@yahoo.com;

attributes associated with the data. Important features/attributes of the data were selected using correlation feature selection - subset evaluation (cfs) method. Three tree based machine learning algorithms (RF, J48 and REPTree) were used to classify the control and NDD subjects. The performance of classifiers were evaluated using Precision, Recall, F-Measure, MAE and RMSE.

Results: In order to evaluate the performance of tree based classifiers, two different settings of data i.e. complete features and selected features were used. In classifying control vs HD subjects, RF provides the robust separation with classification accuracy of 84.79% using complete features and 83.94% using selected features. While in classifying control vs PD subjects, and control vs ALS subjects, RF also provides the best separation with classification accuracy of 86.51% and 94.95% respectively using complete features and 85.19% and 93.64% respectively using selected features.

Conclusion: The variability analysis of physiological signals provides a valuable non-invasive tool for quantifying the system of dynamics of healthy subjects and to examine the alternations in the controlling mechanism of these systems with aging and disease. It is concluded that selected features encode adequate information about neural control of the gait. Moreover, the selected features along with tree based machine learning algorithms can play a vital for early detection of NDDs, when pharmacological interventions are still possible.

Keywords: Decision trees classifiers; machine leaning; neurodegenerative diseases; stride interval.

1. INTRODUCTION

Neurodegenerative disease (NDD) is an umbrella term used to describe neurological disorders due to the failure or malfunctioning of neurons in motor, sensory and cognitive system [1]. Neurodegenerative disorders include Amyotrophic lateral sclerosis (ALS), Alzheimer's disease (AD), Huntington's disease (HD) and Parkinson's disease (PD). At present, around 5 million Americans are suffering from AD, 1 million from PD, 30,000 from ALS, and 3000 from HD [2]. The NDD patients experience progressive loss of cognitive control and symptoms include, problems in speech, gait problems and dementia [3,4]. Due to appearance of NDD symptoms at advanced stage of disease, early diagnosis of NDDs turns out to be impossible using traditional manual methods. The early detection of the NDD onset is vital for an early treatment that may be helpful to prevent further disease progression. Among current diagnostic methods, neuropathology is considered as the gold standard [5] that is based on an autopsy and is performed after the patient's death. Agrawal and Biswas [2] used molecular diagnostic techniques for early detection and diagnosis of NDDs. However, this approach requires robust collaboration between neurologists, psychologists, biologist, and biomaterials scientists and other trained personnel. Thus, researchers from medical domain are in search of effective non-invasive diagnostic tools for an early detection of neurological disorders to take timely pharmacological interventions.

The dysfunction of the cognitive system is directly connected with gait abnormality and is a major symptom of neurological disorder. Understanding the relationship between abnormalities in gait dynamics and malfunctioning or loss of motor neurons can be helpful for assessing NDD progression and devising potential pharmacological interventions [6]. The human gait is the cyclic movement of feet alternatively striking the ground and patterns of change obtained due to repeated stride to stride movement [7]. Gait cycle duration or stride interval is the time between consecutive heel strikes of the same, which fluctuates in a complex manner [8]. The researchers used this parameter in numerous studies to study the complex dynamics of human gait [8-12]. In healthy individuals the neural control remains intact due to which fluctuation magnitude of strides in control subjects remain small, whereas in NDD subjects fluctuations of strides become high due to loss of neural control [8]. Since human gait patterns have direct link to neurological system of brain, thus stride to stride variability analysis can provide a vital role in diagnosing neurodegenerative disorders.

In a study by Zheng, et al. [13], classification of healthy and three NDDs (ALS, HD and PD) subjects using machine learning approaches (support vector machine (SVM), KStar, and RF) have been carried out. The results of study demonstrated that high classification accuracy can be achieved using the 10 extracted features from gait cycles. In order to distinguish normal walking and simulated gait (leg length difference and leg weight asymmetry), Barton and Lees [14] used neural networks. They achieved the correct assignment ratio of 83.3% for unknown gait pattern. In another study Xia, et al. [15], employed four machine learning algorithms (SVM, RF, *k*-nearest neighbor (*k*NN) and multilayer perceptron MLP) for the classification of healthy control and NDDs subjects based on features extracted using nine statistical measures. They used hill climbing method for selecting optimal feature subset and achieved accuracy of 96.83% in classifying healthy control and NDDs subjects. In a study by [16] classification of control and NDDs objects was performed using gait signals. For each right and left foot signals, 13 statistical features were computed. Performance of five different machine learning algorithms (MLP, A2DE, RF, DECORATE and K*) by incorporating 10-fold cross validation method were compared. They found that performance of K* classifier is better in comparison to other classifiers.

The advent of modern computing technologies has enabled the researchers and healthcare professionals to analyze clinical data and study recurring patterns within data that was previously not possible. Medical data mining has great potential for exploring the hidden patterns in the data sets in the medical domain [17]. These patterns have link with various diseases according to the characteristics of the subjects with respect to the predefined set of categories (classes). Decision trees are most commonly used classifiers which use a decision tree as a predictive model that maps observations about an item to conclude about the item's target value (class) [18]. In such tree structures, leaves represent class labels and branches represent conjunctions of features that lead to target class labels [18]. Data mining and machine learning literature shows that tree based classifiers have been extensively used for classification/prediction purposes and such methods show their efficacy in case of data with small numbers of attributes.

The present study is an attempt to classify the controlled and NDD subjects (HD, PD, and ALS) using three traditional decision tree based classifiers (RF, REPTree and J48). For classification purpose, data sets used in this study includes normal and NDD subjects having 13 features and this data was taken from publically available Physionet database. We have used controlled and each of the HD, PD, and ALS subjects separately in binary classification settings. All three decision trees based classifiers have shown better classification accuracies for complete features and selected features, in classifying controlled and ALS diseased subjects as compared to other PD and HD. Among all three classifiers RF performed slightly better.

The rest of the paper is organized as follows: First, we describe in detail the datasets used in this study along with decision trees based classifiers, algorithms and block diagram of classification process. Then the results of the study are presented and discussed followed by the conclusion section.

2. MATERIALS AND METHODS

2.1 Dataset

Datasets used in this study were taken from publicly available Physionet database [19] i.e. “Gait Dynamics in Neuro Degenerative Disease Data Base”. The detail of records in this database is presented in Table 1. The basic attributes associated with gait data of NDD database are listed in Table 2.

In machine learning, classification/prediction consists of three main phases namely preprocessing, processing and post processing (may or may not use). Preprocessing is very important step because it helps to refine the datasets used for classification. Feature selection is one of the preprocessing methods which helps to reduce the dimension of datasets.

2.2 Feature Selection

Not all the features recorded in datasets are relevant to the problem under discussion. Machine learning offers different methods for the selection of relevant features. Such methods play an important role in classification and improves

Table 1. Complete detail of controlled and NDD database

Subject name	No. of instance	Mean ages	Age range	Height (m) (mean ± std)	Weight (kg) (mean ± std)
Control	16	39	20–74	1.83 ± 0.02	66.8 ± 2.8
PD	15	67	44–80	1.87 ± 0.04	75.1 ± 4.4
HD	20	55	36–70	1.83 ± 0.02	72.1 ± 3.8
ALS	13	47	29–71	1.73 ± 0.03	73.3 ± 6.5

Table 2. List of attributes along with the measuring unit associated with the dataset

Attribute Name	Unit	Attribute Name	Unit
Elapsed Time	(sec)	Left Stance Interval	(sec)
Left Stride Interval	(sec)	Right Stance Interval	(sec)
Right Stride Interval	(sec)	Left Stance Interval	(% of stride)
Left Swing Interval	(sec)	Right Stance Interval	(% of stride)
Right Swing Interval	(sec)	Double Support Interval	(% of stride)
Left Swing Interval	(% of stride)	Double Support Interval	(sec)
Right Swing Interval	(% of stride)		

the classification ability of learning algorithms. In this work we use correlation feature selection - subset evaluation (cfs) method for the selection of relevant features using greedy step wise search. It is dimensionally reduction method and plays an important role in classification [20]. On the basis of obtained association and already set threshold values relevant features may be obtained. Using this feature selection method, we selected 07 features. After feature selection we applied decision trees methods.

2.3 Random Forest (RF)

Random Forest (RF) [21] is the combination of different decision trees, used to classify the data samples into classes. It is commonly used statistical technique used for the classification. The worth of each distinct tree in not essential, the purpose of random tree is to reduce the error rate of the whole forest [22]. The error rate depends upon two factors i.e. correlation between two trees and the strength of the tree. The algorithm to construct each tree in RF is as follows [21].

- Each tree is grownup by sampling N arbitrarily, if the training set includes N number of cases but these cases are used with replacement from the original data. For constructing the tree, these N samples are training set.
- The variable m is selected for input variables of M number, such that $m << M$ at each node, at random out of M, m variables are chosen and for splitting the node the best split is used at these m. The m value remains constant during the growing of forest.
- Each tree is grownup until the largest possible extent is met. No pruning is used.

2.4 Reduced Error Pruning Tree (REPtree)

REPTree method is proposed by Quinlan [23]. The REPTree algorithm generate a decision tree, by calculating the information gain using entropy. It helps to decrease the decision tree model complexity by reduced error pruning method and also reduces the error which arises from variance [24]. The information gain is a criteria that uses entropy as measure, and select the attributes having maximum information gain. Let T be a set of examples containing m elements belong to class X and n elements belong to class Y. The information required for deciding whether a random example from T belongs to X or Y is defined as

$$I(t,f) = -\frac{t}{t+f}\log_2\frac{t}{t+f} - \frac{f}{t+f}\log_2\frac{f}{t+f} \quad (1)$$

According to Rokach and Maimon [25], if T_i comprises of m_i examples belonging to X and n_i examples belonging to Y, then the expected information required (entropy) to classify examples in all sub trees T_i is.

$$E(A) = \sum_{1}^{v}\frac{ti+fi}{t+f}I(ti,fi) \quad (2)$$

The pruning method in decision trees can be done as post or pre pruning. Pre-pruning generates trees more rapidly, whereas post-pruning generates more effective trees [26].

2.5 J48

ID3 is the prominent decision tree algorithm proposed by Quinlan [27]. An extended version of ID3 is C4.5, proposed by Quinlan [28]. In Weka (a data mining tool), J48 is an open source Java implementation of the C4.5 algorithm used for the creation of decision tree based on a set of labeled input data [28]. Missing values are ignored when J48 algorithm is building a tree i.e. the value for that item can be predicted based on what is known about the attribute values for the other records. The basic idea of J48 algorithm is the division of data into different ranges. Each range is based on the attribute values for that item that are found in the training sample. J48 algorithm classifies in two ways, either by generating decision trees or by the generation of rules from them. The J48 algorithm works by taking three parameters as input, i.e. training data set along with their class labels, list of attributes that describe the training data set and selection method for attribute. A heuristics approach is used for attributes selection, that can best differentiate data tuples according to class. Usually gini index is used as attribute selection method for binary tree and information gain is used for multi-way splits.

3. RESULTS AND DISCUSSION

Classification of controlled and each NDD subject's separately was performed using three decision trees based algorithm RF, REPTree and J48. The classification parameters were optimized, and 10 fold cross validation is used to avoid over fitting and to explore the robust classifier. The degree of separation was quantified using precision, recall and F-measure. The classification was performed for two scenario of data set i.e. complete feature space and reduced feature space. RF approach is the combination of different decision trees, in which the worth of each distinct tree in not essential. The purpose of random tree is to reduce the error rate of the whole forest. The error rate highly depends upon strength of the tree (total number of trees). In order to minimize the overall error rate the parameter (i.e. total number of trees) should be optimized. In this study, the classification results of RF have been computed by setting the size of tree as 50, 100, 150 and 200. Although results have been computed at different tree size however, results are shown against only at optimal tree size for the classification of various groups.

In Table 3, the results of RF, REPTree and J48 in term of accuracy and error for the classification of controlled and NDD subjects separately using complete and selected feature space are presented. It is clear from the table that RF approach provided very good classification between controlled and ALS subjects as compared to REPTree and J48. The RF also provides better classification between controlled vs HD and controlled vs PD as compared to other two classifiers, however classification accuracy is minimal. Feature selection method cfs selects seven features (Left Stride Interval, Right Stride Interval, Right Swing Interval, Left Stance Interval, Right Stance Interval, Double Support Interval and Double Support Interval). The finding also indicates that RF approach gives optimal classification when complete feature space is compared with selected features. For controlled vs HD and Controlled vs PD the classification accuracy of REPTree and J48 classifiers is higher in case of selected feature space as compared to complete feature space whereas for controlled vs ALS the classification accuracy of REPTree and J48 classifiers is smaller in case of selected feature as compared to complete feature space but the difference between there accuracies is very minimal. This shows that selected features are more appropriate features for the analysis of controlled and NDD subjects.

In Table 4, the results of three decision trees classifiers in term of f-measure, recall and precision for the classification of controlled and HD subjects using complete feature and selected feature space are presented. All three performance measures (f-measure, recall and precision) shows better classification rate at higher values (values approache to 1). It is clear from the table that RF provides higher values of f-measure, recall and precision as compared to REPTree and J48. For RF the average values of all three performance measures in case of complete features are (0.850, 0.848, 0.848) respectively whereas for selected features the average values of all three performance measures are (0.841, 0.839, 0.840) respectively. For REPTree the average values of all three performance measures in case of complete features are (0.821, 0.819, 0.819) respectively, whereas for selected features the average values of all three performance measure are (0.821, 0.818, 0.818) respectively. In case of J48 the average values of all three performance measure for complete features are (0.810, 0.810, 0.810) respectively, whereas for selected features the average values of all three performance measure are (0.845, 0.844, 0.844) respectively.

In Table 5 the results RF REPTree and J48 in term of f-measure, recall and precision for the classification of controlled and PD subjects using complete feature and selected feature space are presented. It is clear from the table that RF provides higher values of f-measure, recall and precision compared to REPTree and J48. For RF, average values of all the three performance measures in case of complete features are (0.865, 0.865, 0.865)respectively whereas for selected features the average values of all three performance measures are (0.852, 0.852, 0.852) respectively. For REPTree the average values of all three performance measures in case of complete features are (0.843, 0.843, 0.843) respectively, whereas for selected features the average values of all three performance measure are (0.846, 0.846, 0.846) respectively. In case of J48 the average values of all three performance measure for complete features are (0.836, 0.836, 0.835) respectively, whereas for selected features the average values of all three performance measure are (0.845, 0.844, 0.844) respectively.

In Table 6 the results of three decision trees classifiers: RF REPTree and J48 in term of f-measure, recall and precision for the classification of controlled and ALS subjects using complete feature and selected feature space are presented. It is clear from the table that RF provides higher values of f-measure, recall and precision as compared to REPTree and J48. For RF the average values of all three performance measures in case of complete features are (0.950, 0.950, 0.950) respectively whereas for selected features the average values of all three performance measures are (0.937, 0.936, 0.937) respectively. For REPTree the average values of all three performance measures in case of complete features are (0.938, 0.938, 0.938) respectively, whereas for selected features the average values of all three performance measure are (0.929, 0.928, 0.928) respectively. In case of J48 the average values of all three performance measure for complete features are (0.936, 0.936, 0.936) respectively, whereas for selected features the average values of all three performance measure are (0.934, 0.933, 0.933) respectively.

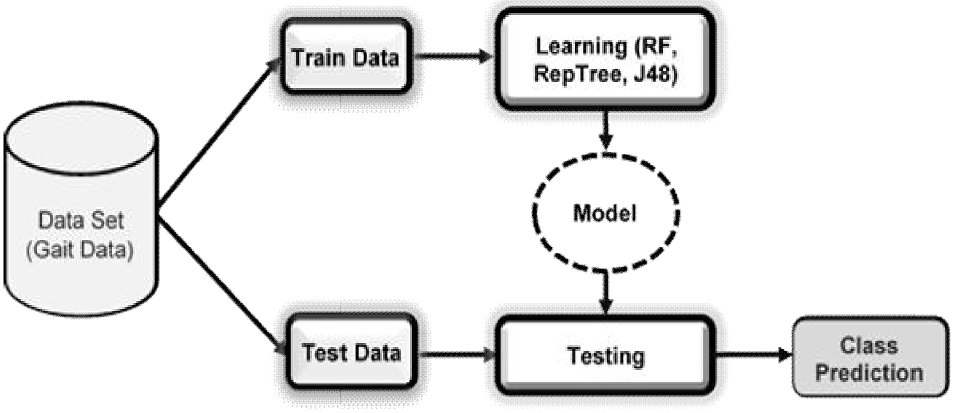

Fig. 1 (a). Block diagram of the decision trees based classification using complete features

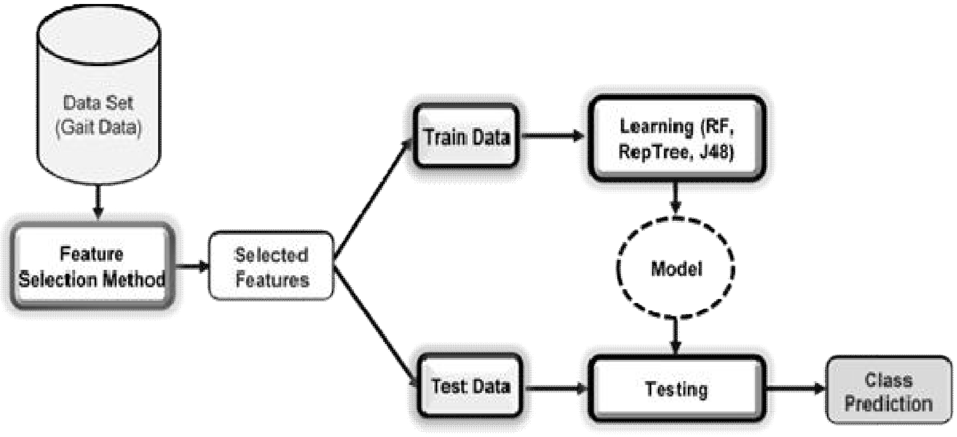

Fig 1 (b): Block diagram of the decision trees based classification using selected features

Fig. 1(a) and 1(b). Illustrates the methodology for both scenarios (complete features and selected features) of data sets

Table 3. Comparison of control and NDD subjects using decision trees based classifiers in terms of accuracy

Classifier	Classification	Complete features	Selected features
Control vs HD			
RF	Accuracy	84.79%	83.94%
	Error	15.21%	16.06%
RepTree	Accuracy	81.89%	81.79%
	Error	18.11%	18.21%
J48	Accuracy	81.01%	82.13%
	Error	18.99%	17.87%
Control vs PD			
RF	Accuracy	86.51%	85.19%
	Error	13.49%	14.81%
RepTree	Accuracy	84.29%	84.61%
	Error	15.71%	15.39%
J48	Accuracy	83.56%	84.43%
	Error	16.44%	15.57%
Control vs ALS			
RF	Accuracy	94.95%	93.64%
	Error	5.05%	6.36%
RepTree	Accuracy	93.75%	92.83%
	Error	6.25%	7.17%
J48	Accuracy	93.56%	93.28%
	Error	6.44%	6.71%

Table 4. Comparison of three decision trees based classifiers for the classification of control and HD subjects in terms of performance measures

Classifier	Performance measures	Complete features		Selected features	
		Control	HD	Control	HD
RF	Precision	0.813	0.881	0.809	0.868
	Recall	0.867	0.832	0.850	0.831
	F-Measure	0.839	0.856	0.829	0.849
REPTree	Precision	0.782	0.855	0.780	0.855
	Recall	0.782	0.855	0.838	0.801
	F-Measure	0.782	0.855	0.808	0.827
J48	Precision	0.793	0.825	0.790	0.850
	Recall	0.793	0.825	0.829	0.814
	F-Measure	0.793	0.825	0.809	0.832

Table 5. Comparison of three decision trees based classifiers for the classification of control and PD subjects in terms of performance measures

Classifier	Performance measures	Complete features		Selected features	
		Control	PD	Control	PD
RF	Precision	0.861	0.870	0.848	0.857
	Recall	0.886	0.842	0.876	0.826
	F-Measure	0.873	0.856	0.861	0.841
REPTree	Precision	0.839	0.848	0.840	0.854
	Recall	0.839	0.848	0.873	0.816
	F-Measure	0.839	0.848	0.856	0.834
J48	Precision	0.829	0.843	0.839	0.851
	Recall	0.865	0.803	0.871	0.815
	F-Measure	0.847	0.823	0.855	0.833

The comparison between three trees based classifiers for the classification of controlled and NDD subjects using complete feature and selected feature space were also assessed using mean absolute error (MAE) and root mean squared error (RMSE). MAE is an indication of the average deviation of the predicted values from the corresponding observed values. MAE present information on long term performance of the models. Lower values of MAE shows better long term prediction of model. RMSE presents information on the short term efficiency. RMSE with lower values represent more accurate evaluation.

Table 6. Comparison of three decision trees based classifiers for the classification of control and ALS subjects in terms of performance measures

Classifier	Performance measures	Complete Features		Selected Features	
		Control	ALS	Control	ALS
RF	Precision	0.965	0.926	0.954	0.909
	Recall	0.953	0.945	0.942	0.928
	F-Measure	0.959	0.935	0.948	0.918
REPTree	Precision	0.958	0.907	0.950	0.896
	Recall	0.958	0.907	0.933	0.921
	F-Measure	0.958	0.907	0.941	0.908
J48	Precision	0.957	0.904	0.955	0.899
	Recall	0.938	0.932	0.935	0.929
	F-Measure	0.947	0.918	0.945	0.914

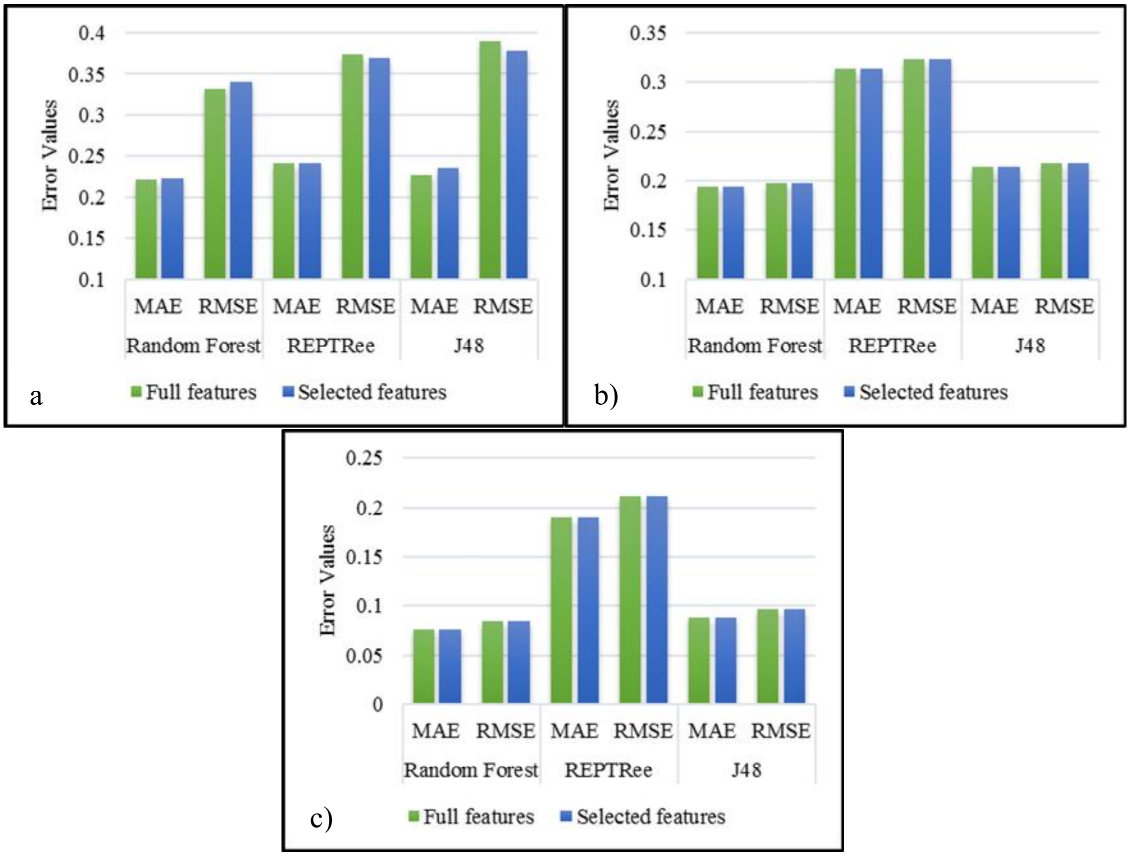

Fig. 2. Comparison of three decision trees based classifiers for the classification of a) controlled vs HD subjects b) controlled vs PD subjects and c) controlled vs ALS subjects in terms of error measures

In Fig. 2, MAE and RMSE values of three decision trees based classifiers for the classification of controlled vs PD, controlled vs HD and controlled vs ALS subjects are presented. Lower values of both error measures show better prediction of models. It is clear from the figure that for both of complete and selected features, MAE and RMSE values against RF are lower as compared to REPTree and J48.

3.1 Discussion

The extraction of information related to the physiological behavior of system by analyzing the biological signals is an interesting and imperative research field. More robust techniques with better classification ability are needed to quantify the dynamics of biological signals in normal and abnormal conditions. Human gait is a highly integrated and complex set of coordinated activities with multiple inputs and numerous outputs [8-12]. Variability in quantitative gait data arises from many potential sources including natural temporal dynamics of neuro-motor control, pathologies of neurological or musculoskeletal systems, the effect of aging and variations in external environment, assistive devices, instrumentation or data collection methodologies. Various linear and nonlinear measures have been used to study variability in human gait time series data [8-12].

The present study is aimed at classifying the human gait data using three traditional decision trees based learning algorithms. The data used for analysis was taken from publicly available Physionet database: “Gait in NDDs”. RF is the combination of different decision trees, used to classify the data samples into classes. The worth of each distinct tree in not essential, the purpose of random trees is to reduce the error rate of the whole forest. REPTree is classifier that built a decision tree by computing the information gain using entropy. It reduces the model complexity of decision tree by “reduced error pruning method”. J48 is an open source Java implementation of the C4.5 algorithm. The C4.5 picks those attributes of the data that most adequately split the set of samples into subsets enhanced in one class or the other. Usually the information gain is used as splitting criterion. The attribute with the highest information gain is selected for making decision. For the classification, data set with two settings i.e. complete feature space and selected feature space have been used. The findings indicated that RF provides better classification between control and NDD subjects as compared to REPTree and J48, especially the classification accuracy of controlled vs ALS is very good. Also, for all three comparison categories of controlled and NDD subjects, selected features provide almost same accuracy as of complete feature space which depicts that the selected features are the most suitable features for the classification.

4. CONCLUSION

The variability analysis of physiological signals provides a valuable non-invasive tool for quantifying the system of dynamics of healthy subjects and to examine the alternations in the control mechanism of these systems with aging and disease. During last three decades mounted research has been carried out to understand the process of human locomotion and to evaluate performance of various measures to quantify internal and external stress conditions. In this study, the performance of three decision trees based algorithms, namely RF, REPTree and J48 were evaluated for the classification of controlled and NDD subjects. The comparison of controlled and NDD subjects have been carried out as controlled vs PD, controlled vs HD and controlled vs ALS. The data set have been used in two scenarios i.e. data set with complete features and data set with selected features. The classification accuracies of above mentioned classifiers have been almost similar, but the RF performed marginally better. Among all the three different comparison categories, the classifiers show better classification between controlled and ALS subjects. The outcomes of the study can be used for assessing the efficacy of neuropharmacological interventions before and after using the drugs by exploring the effects of medicine on the behavior (Neuropsycho-pharmacology) and to understand neurochemical interactions (molecular neuropharmacology).

CONSENT

It is not applicable.

ETHICAL APPROVAL

It is not applicable.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Meriggi P, Castiglioni P, Rizzo F, Gower V, Andrich R, Rabuffetti M, Ferrarin M, Di Rienzo M. Potential role of wearable, ambulatory and home monitoring systems for patients with neurodegenerative diseases and their caregivers. In Pervasive Computing Technologies for Healthcare (Pervasive Health), 2011 5th International Conference. IEEE. 2011;316-319.
2. Agrawal M, Biswas A. Molecular diagnostics of neurodegenerative disorders. Frontiers in molecular biosciences. 2015;2:54.
3. Harter A, Hopper A, Steggles P, Ward A, Webster P. The anatomy of a context-aware application. Wireless Networks. 2002;8(2/3):187-97.
4. Iram S. Early detection of neurodegenerative diseases from bio-signals: A machine learning approach (Doctoral dissertation, Liverpool John Moores University).
5. Chung S, Sonntag KC, Andersson T, Bjorklund LM, Park JJ, Kim DW, Kang UJ, Isacson O, Kim KS. Genetic engineering of mouse embryonic stem cells by Nurr1 enhances differentiation and maturation into dopaminergic neurons. European Journal of Neuroscience. 2002;16(10): 1829-38.
6. Hausdorff JM, Lertratanakul A, Cudkowicz ME, Peterson AL, Kaliton D, Goldberger AL. Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. Journal of applied physiology. 2000;88(6): 2045-53.
7. Han J, Jeon HS, Jeon BS, Park KS. Gait detection from three dimensional acceleration signals of ankles for the patients with Parkinson's disease. In Proceedings of the IEEE. The International Special Topic Conference on Information Technology in Biomedicine, Ioannina, Epirus, Greece 2006;2628.
8. Hausdorff JM, Peng CK, Ladin ZV, Wei JY, Goldberger AL. Is walking a random walk? Evidence for long-range correlations in stride interval of human gait. Journal of Applied Physiology. 1995;78(1):349-58.
9. Aziz W, Arif M. Complexity analysis of stride interval time series by threshold dependent symbolic entropy. European Journal of Applied Physiology. 2006;98(1): 30-40.
10. Aziz W, Arif M. Genetically optimized hybrid gait dynamics classifier. In 2006 International Conference on Emerging Technologies. IEEE. 2006;765-770.
11. Abbasi AQ, Loun WA. Symbolic time series analysis of temporal gait dynamics. Journal of Signal Processing Systems. 2014;74(3):417-422.
12. Qumar A, Aziz W, Saeed S, Ahmed I, Hussain L. Comparative study of multiscale entropy analysis and symbolic time series analysis when applied to human gait dynamics. In 2013 International Conference on Open Source Systems and Technologies. IEEE. 2013;126-132.
13. Zheng H, Yang M, Wang H, McClean S. Machine learning and statistical approaches to support the discrimination of neuro-degenerative diseases based on gait analysis. In Intelligent patient Management. Springer, Berlin, Heidelberg. 2009;57-70.
14. Barton JG, Lees A. An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams. Gait & Posture. 1997; 5(1):28-33.
15. Xia Y, Gao Q, Ye Q. Classification of gait rhythm signals between patients with neuro-degenerative diseases and normal subjects: Experiments with statistical features and different classification models. Biomedical Signal Processing and Control. 2015;18:254-62.
16. Aydin F, Aslan Z. Classification of neurodegenerative diseases using machine learning methods. International Journal of Intelligent Systems and Applications in Engineering. 2017;1(5):1-9.
17. Wasan SK, Bhatnagar V, Kaur H. The impact of data mining techniques on medical diagnostics. Data Science Journal. 2006;5:119-26.
18. Xu G, Zong Y, Yang Z. Applied data mining. CRC Press; 2013.
19. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. Physio Bank, Physio Toolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation. 2000;101(23):e215-20.
20. Hall MA, Smith LA. Feature subset selection: A correlation based filter approach. In Proceedings of International

Conference on Neural Information Processing and Intelligent Information Systems, Berlin: Springer. 1997;855-858.

21. Breiman L. Random forests. Machine Learning. 2001;45(1):5-32.
22. Tomas P, Krohova J, Dohnalek P, Gajdos P. Classification of cardiotocography records by random forest. In Telecommunications and Signal Processing (TSP) 2013. 36th International Conference. IEEE. 2013;620-923.
23. Quinlan JR. Simplifying decision trees. International Journal of Man-machine Studies. 1987;27(3):221-34.
24. Witten IH, Frank E, Hall MA, Pal CJ. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann; 2016.
25. Rokach L, Maimon O. Top-down induction of decision trees classifiers-a survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2005;35(4):476-87.
26. Alpaydin E. Introduction to machine learning. MIT Press; 2009.
27. Quinlan JR. Induction of decision trees. Machine Learning. 1986;1(1):81-106.
28. Quinlan JR. C4. 5: Programs for machine learning. Elsevier; 2014.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/58718