Volume 2, Issue 3, March 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Feature Relevance Analysis and Classification of Parkinson Disease Tele-Monitoring Data Through Data Mining Techniques

Dr.R.Geetha Ramani
Professor & Head-Dept.of CSE
Rajalakshmi Engineering *College*
Thandalam, Chennai

G. Sivagami
Master of Engineering-Dept of CSE
Rajalakshmi Engineering College,
Thandalam, Chennai
shiv.haida@gmail.com

Shomona Gracia Jacob
Ph.D Research Scholar-Dept.of CSE
Rajalakshmi Engineering College,
Thandalam, Chennai

Abstract*— State-of-the-art research in the field of data mining is having great impact in the area of medical diagnosis and disease prediction. This paper places emphasis on classifying the severity of Parkinson's disease (idiopathic Parkinsonism). This disease is a neurodegenerative disorder of the central nervous system. The brain cells (neurons) in the human brain produce dopamine in a particular area of the brain called the substantia nigra. Symptomatic identification includes loss of these specific brain cells and decline in dopamine concentration. Unified Parkinson Disease Rating Scale (UPDRS), captures multiple aspects of Parkinson Disease that include Mentation, Behaviour and mood, Activities of Daily Life (ADL), Motor Examination and complications of therapy. In the Parkinson Disease Tele-monitoring Dataset, the data consists of 16 biomedical voice measures with test subject information, motor UPDRS and total UPDRS scores. The main goal of this work is to predict the motor and total UPDRS scores from the voice measures. Predicting the scores can be done by extracting useful knowledge and thereby providing the scientific decision-making classification rules necessary for the diagnosis of disease severity. This is done by precisely classifying the given dataset and relegating them into people with high scores and low scores. Moreover this paper highlights the impact of six feature relevance algorithms and thirteen classification algorithms on the Parkinson Tele-monitoring dataset. We report 100 percent accurate classification by the Random Tree classification algorithm with the features filtered by the ReliefF algorithm.

Keywords*— Data Mining, Parkinson Disease, Tele Monitoring Data, Feature Relevance Analysis, Classification, Clinical Data Mining.

I. INTRODUCTION

This Parkinson Disease is a degenerative disease of the brain that affects the nerve cells involved in movement [1, 2]. The most typical features of Parkinson's disease are a tremor or fine shake while the person is at rest, rigidity or increased tone in the body muscles, slowness of all movements (bradykinesia), and unsteady balance. Parkinson's disease is caused by the loss of brain cells that produce dopamine, an important neurotransmitter (a chemical that carries signals between the neurons in the brain), which enables us to perform smooth, coordinated movements.

Several previous studies [1]-[4] have shown that vocal impairment may be one of the earliest indicators of the disease, and deficiencies in speech affect approximately 75–90 % people with Parkinson's disease (PD). Non-invasive Tele-monitoring is an emerging option in general medical care, potentially affording reliable, cost effective screening of PWP (People with Parkinson's Disease) alleviating the burden of frequent and often inconvenient visits to the clinic. This also relieves national health systems from excessive additional workload, decreasing the cost and increasing the accuracy of clinical evaluation of the subject's condition.

Clinical data mining [5] is the application of data mining techniques to clinical data. In machine learning and pattern recognition, classification [6]-[7] refers to an algorithmic method or process for assigning a given portion of input data into one of a suggested number of categories. The term "classifier", refers to the mathematical operation, implemented by a classification algorithm, which maps the keyed-in data to a class or category. Feature Selection [5]-[7] is a phase in Dimensionality reduction that selects a subset of features from the original existing set in order to improve the classifier accuracy and reduce storage space and time required for computation. The selection of features is done based on the contribution of the attribute in predicting the target value.

In this paper we have adopted the Parkinson Disease Telemonitoring Dataset to evaluate the performance of six feature relevance algorithms and thirteen classification algorithms viz , Quinlan's C4.5 decision tree algorithm

(C4.5) ,Classification Tree(C-RT), Cost-Sensitive Classification Tree(CS-CRT), Partial Least Squares for Classification (C-PLS), Cost-sensitive Decision Tree algorithm(CS-MC4), SVM for classification(C-SVC), Iterative Dichomotiser(ID3), K-Nearest Neighbor(K-NN), Linear Discriminant Analysis (LDA), Partial Least Squares - Discriminant/Linear Discriminant Analysis(PLS-DA/LDA), Random Tree (Rnd Tree), Support Vector Machine(SVM) classification algorithms. We also investigate the effect of feature selection using Fisher Filtering (FF), ReliefF, Runs Filtering, Forward Logistic Regression (FLR), Backward Logistic Regression (BaLR) and Stepwise Discriminant (Step Disc) Analysis algorithms to enhance the classifier accuracy and reduce feature subset size.

A. Paper Organisation

Section II reviews the related work in the field of data mining on the Parkinson disease Tele-monitoring dataset. Section III portrays the system design and explains the different phases in the data mining framework while Section IV and Section V elaborate on the feature selection and classification algorithms respectively. Section VI reviews the experimental results while Section VII concludes the paper.

II. LITERATURE SURVEY

Previous research in the field of data mining and medical research is concisely presented in the following paragraphs.

Rusz et.al, [8] investigated the feasibility of automated acoustic measures for the identification of voice and speech disorders in Parkinson's disease (PD). The speech data were collected from 46 Czech native speakers, 24 with early PD, before receiving pharmacotherapy treatment. They have applied several traditional and non-standard measurements in combination with statistical decision-making strategy to assess the extent of vocal impairment of recruited speakers. Subsequently, they have made use of support vector machine to find the best combination of measurements to differentiate PD from healthy subjects. Their proposed method led to an overall classification performance of 85%. Moreover they have detected relationships between measures of phonation and articulation and bradykinesia and rigidity in PD. They have concluded by reporting that the acoustic analysis can ease the clinical assessment of voice and speech disorders, and serve as measures of clinical progression and also provide means for monitoring of treatment effects.

In another study Rusz et.al, [9] presented the potential of the simple Bayes rule to reveal changes in degradable speech performance in the course of PD-related dysarthria. The various speech data were recorded from 23 speakers with recently diagnosed PD and 23 healthy speakers. It was reported that 19 various acoustic measurements were able to differentiate PD significantly from healthy speakers. Subsequently, the Bayes theorem was applied to each of those measurements. As a result, the 21 PD patients and 21 healthy people were correctly classified according to their group. The Bayes theorem thus confirmed its feasibility to identify the features of the impaired voice.

Tsanas et.al,[10] have demonstrated rapid, remote replication of UPDRS assessment with clinically useful accuracy (5% prediction error), making use of simple, self-administered, and non-invasive speech tests. They characterize speech with signal processing algorithms, and statistically map these algorithms to UPDRS. Their findings have been verified on the largest database of PD speech in existence, that number close to 6,000 recordings from 42 PD patients, recruited to a six-month, multi-centre trial.

Later, Tsanas et.al, [11] demonstrated the potential of wavelets to reveal changes in fundamental frequency variations with PD progression. They have developed a set of new measures based on wavelets, energy, and entropy, which form robust indicators of the UPDRS. These results demonstrate that PD leads to dissimilar speech patterns in males and females, tentatively taken to indicate different patho-physiological mechanisms.

Tsanas et.al [12] has also made a study of disordered voices of people with Parkinson's disease (PD). They have demonstrated that a simple logarithmic transformation of these dysphonia measures can significantly enhance their potential for identifying subtle changes in PD symptoms. The superiority of the log-transformed measures is reflected in feature selection results using Bayesian Least Absolute Shrinkage and Selection Operator (LASSO) linear regression. Moreover they have demonstrated the effectiveness of this enhancement in the emerging application of automated characterization of PD symptom progression from voice signals, rated on the Unified Parkinson's Disease Rating Scale (UPDRS), the gold standard clinical metric for PD. Using least squares regression, they affirm the fact that UPDRS can be accurately predicted to within six points of the clinician's observations.

III. DATA MINING FRAMEWORK

The design framework for the classification of Parkinson disease severity comprises of the training phase which incorporates the process of training data selection, data pre-processing, feature relevance analysis and generation of severity prediction rules through classification algorithms. This is followed by an Evaluation phase wherein the classifiers produced by the feature filtering algorithms are ranked based on their misclassification rate. The Test phase verifies the chosen classifier's accuracy on classifying an unseen Parkinson test data.

A. Training Dataset

The training dataset [13] used in this research is the Parkinson Disease Tele-monitoring dataset downloaded from the UCI Irvine Machine Learning Repository. The dataset comprises of two target classes viz, Motor_UPDRS and Total_UPDRS. The predictor features are common to both target classes. There exist only two possible values for each of the target classes that classify the training dataset as High or Low. The predictor attributes are tabulated in Table I.

TABLE I
PARKINSON DISEASE TRAINING DATASET DESCRIPTION

S.No.	Attribute	Description
1	Subject	Integer that uniquely identifies each subject
2	Age	Subject age
3	Sex	Subject gender '0' - Male , '1' – Female
4	Test_Time	Time since recruitment into the trial. The integer part is the number of days since recruitment.
5	Jitter(per)	Measure of variation in fundamental frequency – Speech
6	Jitter(Abs)	Measure of variation in fundamental frequency – Speech
7	Jitter:RAP	Measure of variation in fundamental frequency – Speech
8	Jitter:PPQ5	Measure of variation in fundamental frequency – Speech
9	Jitter:DDP	Measure of variation in fundamental frequency – Speech
10	Shimmer	Measure of variation in Amplitude- Speech
11	Shimmer(dB)	Measure of variation in Amplitude- Speech
12	Shimmer:APQ3	Measure of variation in Amplitude- Speech
13	Shimmer:APQ5	Measure of variation in Amplitude- Speech
14	Shimmer:APQ11	Measure of variation in Amplitude- Speech
15	Shimmer:DDA	Measure of variation in Amplitude- Speech
16	NHR	Noise to Harmonics Ratio
17	HNR	Harmonics to Noise Ratio
18	RPDE	Recurrence Period Density Entropy - A non linear dynamical complexity measure
19	DFA	Detrended Fluctuation Analysis - A Signal Fractal scaling component
20	PPE	Pitch Period Entropy - A nonlinear measure of fundamental frequency variation

The training clinical dataset then needs to be pre-processed to ensure proper analysis of data through feature selection and classification. The pre-processing of data is done as explained in the following sub-section.

B. Data Pre-processing

Data pre-processing [12]-[14] is an important step in the Data Mining Process. Analysing data that has not been carefully screened for problems can produce misleading results. Any irrelevant or redundant data present, makes knowledge discovery during the training phase a tedious task. Data pre-processing includes cleaning, normalization, transformation, feature extraction and selection etc.

The phases involved in this analysis are clearly depicted in the data mining framework portrayed in Figure 1.

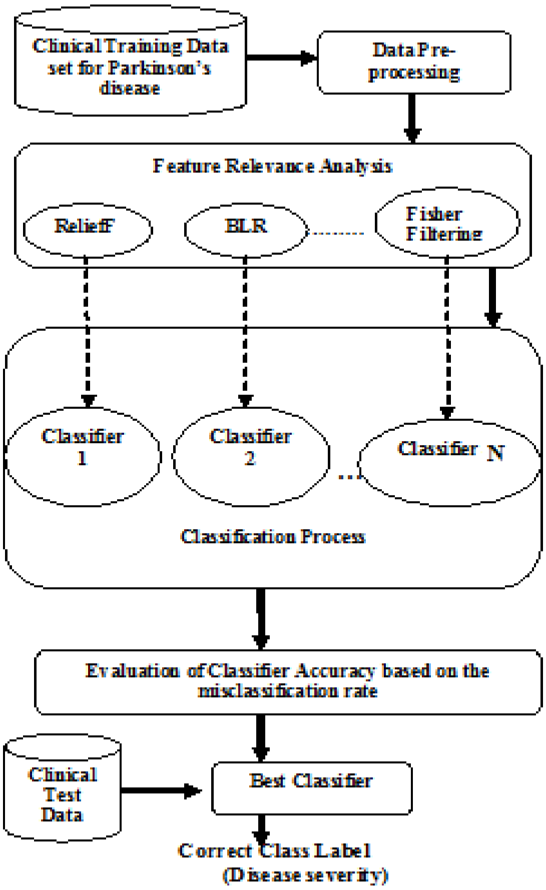

Fig 1. Data Mining Framework for Classification of Parkinson Disease Severity

The integer part of the test_time attribute gives the number of days since recruitment. The dataset is comprised of two class labels, Motor_UPDRS and Total_UPDRS. Two datasets are formed with one containing the Motor and other with Total as target class [15]. The class labels are continuous. Classifying continuous class labels leads to computation complexity. To avoid this, discretization of the class labels is done. In order to perform this activity, the dataset is split up into 3 parts with test_time as splitting criteria. The dataset is divided into baseline (first 3 months), after 3 months and after 6 months. The mean for that particular split is calculated and class label value falling above the split are considered as "High" and class label value falling below the splitting value are considered as "Low". Then finally the three fragmented data are combined to form a single dataset.

C. Feature Selection

Feature selection [12] [13] [16] has been a lively and effective research area in pattern recognition, statistics, and data mining communities. The main idea behind feature selection is to identify and choose a subset of the input variables by analysing and eliminating features with little or no predictive information. Feature selection can appreciably improve the comprehensibility and lucidity of the resulting classifier and often construct a model that generalizes better to unseen points. In this research work, we make use of six feature selection algorithms to identify the attributes that contribute significantly to the prediction of the target class value for both the datasets viz, Motor-UPDRS and Total_UPDRS. The algorithms used in Feature Selection and the experimental results are clearly brought out in Section 4 and Section 6 respectively.

D. Classification

Classification [13] [14] is the process of finding a set of models that describe and distinguish data classes. This is done to achieve the goal of being able to use the model to predict the class whose label is unknown. This work involves the design of an efficient classifier for binary classification on the Parkinson Disease Tele-Monitoring dataset by analysing the misclassification rate produced by the thirteen classification algorithms on the aforementioned dataset. The Random Tree classification algorithm produces 100 percent accuracy in classifying the datasets. The Random Tree algorithm and the detailed results of classification are brought out in Section 5 and Section 6 respectively.

E. Evaluation

The classifier results obtained with the filtered attributes of the various feature reduction algorithms are analysed and evaluated based on their accuracy. The Random Tree classification algorithm is reported to give 100 percent classifier accuracy for both the Motot_UPDRS and the Total_UPDRS datasets as clearly described in Section 6.

F. Test Phase

The designed classifier accuracy is verified by providing a test data for the two datasets namely Motor_UPDRS and the Total_UPDRS [17]. The test data predictor attributes are compared with the generated classifier rules and the test data class label is determined and checked for accuracy affirmation.

IV. FEATURE SELECTION ALGORITHMS

Feature Selection [12-14] is the process of extracting the relevant attributes necessary for identification of the target class. In this paper we make use of the wrapper approach to feature selection that aims at estimating the classifier accuracy to rank the credibility of the feature selection algorithms. A comparison of various feature selection algorithms are worked to find the most appropriate.

The ReliefF feature selection algorithm has given the best set of attributes for classification. The working principle of the ReleifF algorithm is given in sub-section A.

A. *RelieF Feature Selection Algorithm*

ReliefF is an extension of the popular Relief algorithm [7] [8]. A key idea of the ReliefF algorithm is to evaluate, estimate and assess the quality of features according to how well their values distinguish between sample points that are near to each other.

```
Step 1. Select_subset = { }
Step 2. Init. all feature weightage = 0
Step 3. for I = 1 to no_of_sample
        Step 3.1 Get one instance X from the training data
         set D.
        Step 3.2 Get nearhit     H = instance in D where
                     dist(X, H) is closest & X.class=H.class
         Step 3.3 Get nearmiss M = instance in D where
                     dist(X, M) is closest & X.class<>M.class
         Step 3.4 Update weightage for all features
Step 4. for j = 1 to no_of_feature
        Step 4.1 If weightage_j >= Threshold, add feature_j to
                     selected_subset
```

The feature selection phase is followed by the classification phase as described in the next section.

V. CLASSIFICATION ALGORITHMS

Classification [13] [19] is the process of finding a set of models that describe and distinguish data classes. This is done to achieve the goal of being able to use the model to predict the class whose label is unknown. The Random Tree Classification algorithm used in our experimental study is clearly explained with the pseudo code.

A. *Random Tree Classification Algorithm*

The Random tree [13] algorithm can be applied to both classification and regression problems. Random trees are a collection or assembly of tree predictors that is called forest [13]. The classification works as follows: the random trees classifier takes the input feature vector, classifies it with every tree in the forest, and outputs the class label that received the majority of “votes”. In the case of regression the classifier response is the average of the responses over all the trees in the forest. The algorithm for the Random tree algorithm is given below.

```
Start
 {
FF = {collection of all predictor features (forest)}
//forest is obtained from the feature selection algorithm
RF = {input data – feature vector}
Repeat {
 Compare the Attribute Values (av) of RF with FF.
If (RF.av == FF.av) then take the positive branch
else
  take the negative branch
   } for all RF until leaf node is reached.
End
```

VI. PERFORMANCE EVALUATION

A. *Performance Measures*

The measures and their exact meaning have been given as stated by Han and Kamber [19]-[20].

1) Confusion Matrix

Given m classes, a confusion matrix [16] is a table of at least size m*m. An entry, CMi, j in the first m rows and m columns indicates the number of tuples of class I that were named by the classifier as j. It is a valuable tool for analyzing how well your classifier can recognize tuples of different classes.

2) Precision and Recall

Precision [18] and Recall [18] are two basic measures for assessing the performance of text retrieval. In our study Precision refers to the data that is correctly classified by the classification algorithm. 1.000 precision indicates 100% accuracy. Recall is the percentage of information relevant to the class and is correctly classified.

3) Accuracy

The accuracy [15] of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier.

4) Error- Rate

The error rate [1, 16] is also called the misclassification rate. It is simply 1-Acc (M), where Acc (M) is the accuracy of M.

B. *Experimental Results*

The ReliefF feature selection algorithm provides the best set of attributes that generate 100 percent classifier accuracy. The results of the ReliefF feature filtering algorithm are tabulated in Table II.

TABLE II
FEATURE RELEVANCE

Feature Selection Algorithm	Filtered Feature Number
Backward Logit	1,2,3,5,6,9,14,15,17,19,20
Fisher Filtering	1,2,5,7,8,9,10,11,12,13,14,15,16,17,18,19,20
Forward Logit	1,2,3,13,14,17,19
ReliefF	1,2,4,19
Runs Filtering	1,2,3,4,16
Step wise Discriminant	1,2,3,6,9,13,14,17,19,20

The accuracy obtained by executing the thirteen classification algorithms with the feature selection algorithms on the Motor_UPDRS and the Total_UPDRS data sets is tabulated neatly in Table III and Table IV respectively.

It is to be noted here that the number of attributes for split in the Random Tree algorithm has to be specified according to the total number of attributes that is considered for predicting the value of the target class in order to obtain classifier accuracy with 0 percent error rate.

TABLE III
COMPARISON OF CLASSIFICATION ALGORITHM WITH VARIOUS FEATURE SELECTION FOR MOTOR_UPDRS DATASET

Classification Algorithm	Accuracy %					
	BaLR	Fisher	FLR	Relief F	Runs Filtering	Step Disc
C4.5	97.65	97.92	96.99	99.97	99.86	97.67
C-RT	95.76	95.76	95.76	99.88	99.81	95.76
ID3	95.83	95.83	95.83	95.83	96.05	95.83
KNN	96.05	95.8	96.27	99.63	99.98	96.2
CS-CRT	95.76	95.76	95.76	99.88	99.81	95.76
CS-MC4	96.15	96.15	96.15	99.97	99.83	96.24
Rnd Tree	100	100	99.42	100	100	100
C-PLS	71.23	71.23	69.24	73.23	73.38	70.91
C-SVC	75.08	74.81	75.23	75.68	75.8	75.08
LDA	69.96	70.79	69.02	72.83	70.18	69.91
PLS-LDA	71.42	71.88	70.09	72.94	74.96	71.17
PLS-DA	71.42	71.88	70.09	72.94	74.96	71.17
SVM	75.11	74.88	75.27	75.69	75.8	75.06

TABLE IV
COMPARISON OF CLASSIFICATION ALGORITHM WITH VARIOUS FEATURE SELECTION FOR TOTAL_UPDRS DATASET

Classification Algorithm	Accuracy %					
	BaLR	Fisher	FLR	Relief F	Runs Filtering	Step Disc
C4.5	97.65	97.92	96.99	99.97	99.86	97.67
C-RT	95.76	95.76	95.76	99.88	99.81	95.76
ID3	95.83	95.83	95.83	95.83	96.05	95.83
KNN	96.05	95.8	96.27	99.63	99.98	96.2
CS-CRT	95.76	95.76	95.76	99.88	99.81	95.76
CS-MC4	96.15	96.15	96.15	99.97	99.83	96.24
Rnd Tree	100	100	99.42	100	100	100
C-PLS	71.23	71.23	69.24	73.23	73.38	70.91
C-SVC	75.08	74.81	75.23	75.68	75.8	75.08
LDA	69.96	70.79	69.02	72.83	70.18	69.91
PLS-LDA	71.42	71.88	70.09	72.94	74.96	71.17
PLS-DA	71.42	71.88	70.09	72.94	74.96	71.17
SVM	75.11	74.88	75.27	75.69	75.8	75.06

The performance of the classifiers is graphically compared and displayed in Figure 3 and Figure 4 respectively.

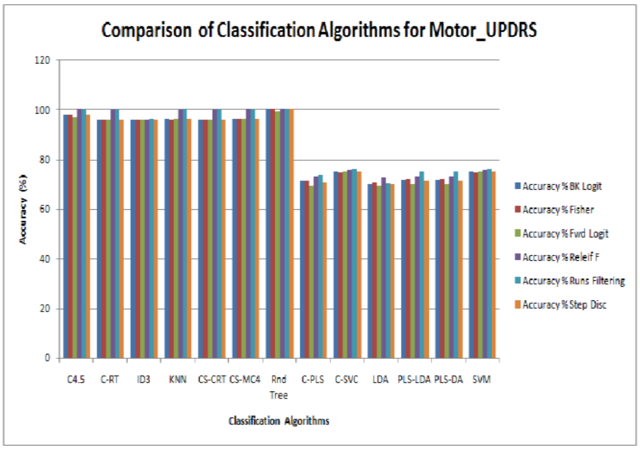

Fig 2. Performance Comparison of Classifier Accuracy on the Motor_UPDRS Dataset

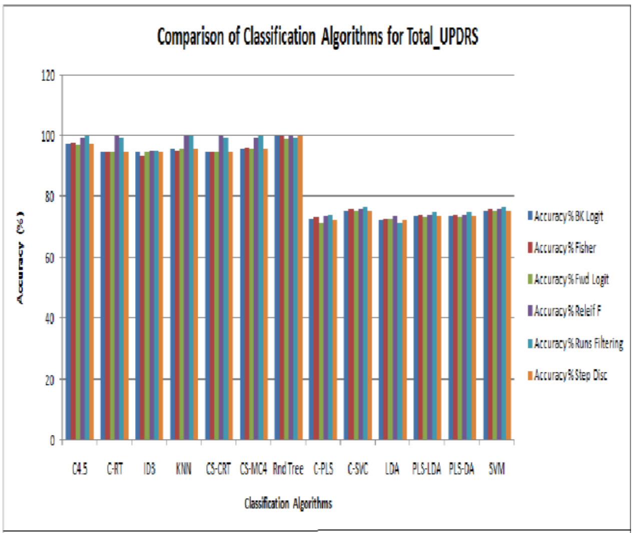

Fig 3. Performance Comparison of Classifier Accuracy on the Total_UPDRS Dataset

The Random Tree classification algorithm produces 100 percent classification accuracy with the attributes selected by the ReliefF algorithm on the two datasets of the Parkinson Telemonitoring dataset.

VII. CONCLUSION

Data mining is unearthing great potential for advancement in the field of medicine relating to disease prediction, diagnosis and therapy. Parkinson Disease is one of the dreaded ailments that target the elderly population who show very slow response to treatment at advanced stages of the disease. Early detection of this disorder will certainly aid in arresting the growth of the ailment thereby providing hope to many ailing minds. In this paper, we have investigated the performance of data mining techniques based on feature relevance analysis and classification in identifying the severity of the Parkinson disease that could provide clinicians a hope of diagnosing the disease at the earliest evidence of the malady. Random Tree Classification algorithm is reported to produce classification result with no errors on both the Motor_UPDRS and the Total_UPDRS datasets. Moreover the attributes selected by the ReliefF feature selection algorithm contribute most to the prediction of the correct class label. We believe the identification of classifier accuracy will definitely advance the current stage of medical decision-making and malady diagnosis.

ACKNOWLEDGMENT

This research work is a part of the All India Council for Technical Education(AICTE), India funded Research Promotion Scheme project titled “Efficient Classifier for clinical life data (Parkinson, Breast Cancer and P53 mutants) through feature relevance analysis and classification” with Reference No:8023/RID/RPS-56/2010-11, No:200-62/FIN/04/05/1624.

REFERENCES

[1] *Dr. R.Geetha Ramani and G.Sivagami. “Parkinson Disease Classification using Data Mining Algorithms”. International Journal of Computer Applications(IJCA), 32(7): 46-53, October 2011 DOI:10.5120/3920-5521. Published by Foundation of Computer Science, New York, USA.*

[2] *Movement Disorder Society Task Force on Rating Scales for Parkinson’s Disease. The Unified Parkinson’s Disease Rating Scale (UPDRS): Status and Recommendations.*

[3] *Goetz, C.G. et al. Testing objective measures of motor impairment in early Parkinson’s disease: Feasibility study of an at-home testing device.Movement Disorders (December 11, 2008) doi. 10.1002/mds.22379*

[4] Stebbins, G.T., Goetz, C.G., Lang, A.E., Cubo, E. Factor analysis of the motor section of the Unified Parkinson’s Disease Rating Scale during the off-state. *Movement Disorders* **14 (4),** 585-589 (1999)

[5] *Mrs.Shomona Gracia Jacob and Dr. R.Geetha Ramani,”Discovery of Knowledge Patterns in Clinical Data through Data Mining Algorithms: Multi-class Categorization of Breast Tissue Data”, International Journal of Computer Applications (IJCA), 32(7): 46-53, October 2011 DOI: 10.5120/3920-5521. Published by Foundation of Computer Science, New York, USA.*

[6] *P.Nancy and Dr.R.Geetha Ramani. Article: “A Comparison on Performance of Data Mining Algorithms in Classification of Social Network Data”. International Journal of Computer Applications 32(8):47-54, October 2011 DOI: 10.5120/3927-5555. Published by Foundation of Computer Science, New York, USA*

[7] *S.Shanthi, Dr.R.Geetha Ramani,” Classification of Vehicle Collision Patterns in Road Accidents using Data Mining Algorithms”, International Journal of Computer Applications, Vol.35, No.12, pp.30-37, December, 2011, DOI: 10.5120/4542-6455, Published by Foundation of Computer Science, New York, USA.*

[8] *J. Rusz, R. Cmejla, H. Ruzickova, J. Klempir, V. Majerova, J. Picmausova, J. Roth, E. Ruzicka. “Acoustic analysis of voice and speech characteristics in early untreated Parkinson’s disease” “*

[9] Jan Rusz1, , Roman Cmejla, Hana Ruzickova, Evzen Ruzicka,, Objectification of dysarthria in Parkinson’s disease using Bayes theorem, Recent Researches in Communications, Automation, Signal Processing, Nanotechnology, Astronomy and Nuclear Physics, ISBN: 978-960-474-276-9

[10] Athanasios Tsanas, Max A. Little, Patrick E. McSharry and Lorraine O. Ramig, Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson’s disease symptom severity, J. R. Soc. Interface (2011) 8, 842–855 doi:10.1098/rsif.2010.0456 published online 17 November 2010

[11] *Athanasios Tsanas, Max A. Little, Patrick E. McSharry, Lorraine O. Ramig. “Enhanced Classical Dysphonia measures and sparse Regression for Tele Monitoring of Parkinson Disease Progression”*

[12] Athanasios Tsanas*, Max A. Little, *Member, IEEE*, Patrick E. McSharry, *Senior Member, IEEE*, Lorraine O. Ramig, Accurate telemonitoring of Parkinson’s disease progression by non-invasive speech tests, TBME-00652-2009.R1

[13] *UCI Machine Learning Repository- Centre for Machine Learning and Intelligent System. http://mlr.cs.umass.edu/ml/datasets/Parkinsons+Telemonitoring*

[14] *Shomona Gracia Jacob, Dr.R.Geetha Ramani, P.Nancy, “Feature Selection and Classification in Breast Cancer Datasets through Data*

Mining Algorithms", Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research (ICCIC'2011), Kanyakumari, India, December 15-18, 2011, IEEE Catalog Number: CFP1120J-PRT, ISBN: 978-1-61284-766-5. Pp. 661-667

[15] *Skodda, S., Rinsche, H., Schlegel, U. Progression of dysprosody in Parkinson's disease over time – A longitudinal study. Movement Disorders (December 31, 2008). Doi: 10.1002/mds.22430*

[16] *Nancy.P, Dr.R.Geetha Ramani, Shomona Gracia Jacob, "Discovery of Gender Classification Rules for Social Network Data using Data Mining Algorithms", Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research(ICCIC'2011), Kanyakumari, India, December 15-18, 2011, IEEE Catalog Number: CFP1120J-PRT, ISBN:978-1-61284-766-5, pp 808-812*

[17] *Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J., Ramig, L.O. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease, IEEE Transactions Biomedical Engineering (September 11, 2008), doi: 10101/npre.2008.2298.1*

[18] Shomona Gracia Jacob, Dr.R. Geetha Ramani, Nancy .P, *"Efficient Classifier for Classification of Hepatitis C Virus Clinical Data through Data Mining Algorithms and Techniques", Proceedings of the International Conference on Computer Applications, Pondicherry, India, January 27-31, 2012,Techno Forum Group, India. ISBN: 978-81-920575-8-3: DOI: 10.73445/ISBN_0768, ACM#.dber.imera.10.73445.*

[19] *Jiawei Han, Micheline Kamber. Data Mining Concepts and Techniq.*Second Edition, 2000.

[20] *Knowledge Discovery in Databases. http://www2.cs.uregina.ca/~dbd/cs831/notes/ml/dtrees/c4.5/tutorial.html.*

AUTHOR'S PROFILE

Dr.R. Geetha Ramani is Professor & Head in Department of Computer Science and Engineering, Rajalakshmi Engineering College, India. She has more than 15 years of teaching and research experience. Her areas of specialization include Data mining, Evolutionary Algorithms and Network Security. She has over 50 publications in International Conferences and Journals to her credit. She has also published a couple of books in the field of Data Mining and Evolutionary Algorithms. She has completed an External Agency Project in the field of Robotic Soccer and is currently working on projects in the field of Data Mining. She has served as a Member in the Board of Studies of Pondicherry Central University. She is presently a member in the Editorial Board of various reputed International Journals.

Ms.Sivagami. G has completed her B.Tech in Information Technology at Tagore Engineering College affiliated to Anna University, Chennai, India. Currently she is pursuing her M.E. in Computer Science and Engineering at Rajalakshmi College of Engineering, affiliated to Anna University, Chennai, India. She has 5 years of industrial experience from JK Technosoft, a global software and solutions company from Bangalore. Starting from Trainee her career took a steadfast growth to system analyst. Her clients were Hindustan Unilever Limited (HUL); the India's Largest Fast Moving Consumer Goods Company with a heritage of over 75 years in India and touches the lives of two out of three Indians. Her areas of interest include ERP, Data Mining, and Databases.

Mrs.Shomona Gracia Jacob completed her M.E. in Computer Science and Engineering at Jerusalem College of Engineering, affiliated to Anna University, Chennai, India. She has more than 3 years of teaching experience. Presently she is pursuing her Ph.D in Computer Science and Engineering at Rajalakshmi Engineering College, affiliated to Anna University of Technology, Chennai. Her areas of interest include Data Mining, Artificial Intelligence and Software Engineering. She has attended and presented few papers at National and International Conferences.

