International Journal of Advances in Science Engineering and Technology, ISSN: 2321-9009, Vol-5, Iss-1, Spl. Issue-2 Feb.-2017
http://iraj.in

PERFORMANCE OF CLASSIFICATION TECHNIQUES ON PARKINSON'S DISEASE

[1]ISMAIL SARITAS, [2]MURAT KOKLU, [3]KEMAL TUTUNCU

[1,3]Department of Electrical – Electronics Engineering, Selcuk University, Konya, TURKEY
[2]Department of Computer Engineering, Selcuk University, Konya, TURKEY
E-mail: [1] isaritas@selcuk.edu.tr, [2]mkoklu@selcuk.edu.tr, [3] ktutuncu@selcuk.edu.tr

Abstract- Nowadays, many methods and algorithms have been developed that may influence the decision-making process and are used to extract meaningful information. One of the well know methods or approaches in information extraction is data mining. Data mining tries to establish the best model to support decision system, to extract information and to categorize, to summarize and etc. according to given data set. The Parkinson's disease-related data obtained from UCI Machine Learning Database is used to try several data mining techniques and methods to see the successes of techniques regarding to diagnosis accuracy ratio to support the expert. So far, Parkinson's disease can actually be diagnosed after medical examinations. However, diagnosis with computer has been the subject of many researches due to demand to help physician. In this study, a research is conducted using 16 different data mining techniques and methods to support the doctors in the decision-making process. The results of the applied methods for the study regarding to diagnosis accuracy ratesare as follows; IB1 (96.4103%), RotationForest (92.3077%) RandomForest (91.7949%), MultilayerPerceptron (90.7692%), ClassificationViaRegression (88.2051%), Bagging (87.6923%), JRip (87.6923%) SMO (87.1795%), OneR (86.1538%), NBTree (86.1538%), Dagging (85.6410%), DTNB (85.1282%), DecisionTable (81.0256%), J48 (80.5128%), BayesNet (80.0000%) and Naïve Bayes (69.2308%).

Keywords- Data Mining, Classification, Parkinson Disease.

I. INTRODUCTION

Parkinson's disease has been defined as "shaking palsy" with the alias name shaky stroke by British physician James Parkinson for the first time in 1817 [1, 2]. Today, Parkinson's disease, which is known by the name of the researcher and without its known paralysis characteristic in today's sense, is the most important and most common type of Parkinsonism that should be recognized. The word "Parkinsonism" evokes a state known for its symptoms emerging with different reasons rather than a specific disease [2, 3].

Parkinson's disease is a neurological disorder that causes the motor reflexes affecting the nervous system and behavioral and other vital functions such as speaking and thinking to partially or wholly lost. Nowadays, many studies about computer-aided diagnosis of Parkinson's disease using artificial intelligence techniques are carried out due to the increasing technological advances. It has recently been seen in studies that specifically "dysphonia" removed from the audio recordings, which means the deterioration in the vocal cords caused by the disease, plays an important role in the diagnosis and the monitoring of the disease [2].

Data mining is used as an information source to find unities, make classification, clustering and estimations by using information discovery systems which are the combination of data warehouses, artificial intelligence techniques and statistical methods [4].

Classification is a function or a model derived for determining the class of a data based on its features. To establish the model, a data set of states of which the results are known previously and values of factors related to these states are used as the training data. Also, a set of data outside the training data is defined as the testing data to be used to determine the accuracy of the model. A classification function or model is established by analyzing the relationship between the features and the classes of the training data. Such a classification function or model can be used to develop a better understanding the classes of the data in the database and to classify new data added to the cluster. For instance, a classification model can be established which can determine a patient's disease from the diagnosis data acting as the training data obtained from patients previously [5].

Numerous data mining techniques and methods are used on the Parkinson's disease data obtained from UCI Machine Learning Database. In this study, 16 different data mining techniques and methods have been tested on Parkinson's disease data.

II. LITERATURE SURVEY

Previous studies including upper mentioned on the related dataset are summarized with the accuracy ratios in Table 1.

Table 1: The Result Obtained In The Literature

Algorithm	Accuracy (%)	Referernces
hGA	96.32	AydoganKızılkaya E., et all, 2012, [10]
Parzen	94.74	Atiya A.F. and Al-Ani A., 2009, [7]
KNNw	94.23	Atiya A.F. and Al-Ani A., 2009, [7]
KNNds	93.21	Atiya A.F. and Al-Ani A., 2009, [7]

PLC	92.95	Atiya A.F. and Al-Ani A., 2009, [7]
Neural Network	92.90	Das R., 2010, [8]
RF	92.56	Chakraborty S., 2011, [9]
KNN	92.05	Atiya A.F. and Al-Ani A., 2009, [7]
PCA-RF	91.77	Zhang L. and Suganthan P.N., 2014, [10]
GPCo	91.67	Atiya A.F. and Al-Ani A., 2009, [7]
Semi-BSVM	91.45	Chakraborty S., 2011, [9]
SVMs	91.40	Little M.A. et all, 2009, [11]
RF-ensemble	91.23	Zhang L. and Suganthan P.N., 2014, [10]
RF	90.78	Zhang L. and Suganthan P.N., 2014, [10]
LDA-RF	90.11	Zhang L. and Suganthan P.N., 2014, [10]
μcAnt-Miner	90.00	Salama, K.M., 2013, [12]
Jrip	88.76	Salama, K.M., 2013, [12]
Regression	88.60	Das R., 2010, [8]
SVM	88.49	Chakraborty S., 2011, [9]
BSVM	88.33	Chakraborty S., 2011, [9]
PLCm	87.69	Atiya A.F. and Al-Ani A., 2009, [7]
SVM2	87.44	Atiya A.F. and Al-Ani A., 2009, [7]
cAnt-Miner	87.40	Salama, K.M., 2013, [12]
TSVM	86.92	Chakraborty S., 2011, [9]
GP-COACH	86.48	AydoganKızılkaya E., et all, 2012, [10]
FART	86.18	Salama, K.M., 2013, [12]
FRBCS_GP	85.75	AydoganKızılkaya E., et all, 2012, [10]
DMNeural	84.30	Das R., 2010, [8]
Decision tree	84.30	Das R., 2010, [8]
RBF SVM	83.98	Li, Y. and Leng, Q., 2015, [13]
GPC	83.97	Atiya A.F. and Al-Ani A., 2009, [7]
GMM	83.46	Atiya A.F. and Al-Ani A., 2009, [7]
SVM1	83.33	Atiya A.F. and Al-Ani A., 2009, [7]
GCCL-Ishibuchi	83.27	AydoganKızılkaya E., et all, 2012, [10]
SMA	82.76	Li, Y. and Leng, Q., 2015, [13]
NCA	82.69	Atiya A.F. and Al-Ani A., 2009, [7]
SAMA	81.84	Li, Y. and Leng, Q., 2015, [13]
2SLAVE	81.75	AydoganKızılkaya E., et all, 2012, [10]
ANNs	81.30	Ene M., 2008, [14]
Linear SVM	78.06	Li, Y. and Leng, Q., 2015, [13]
GP-PITT-Tsakonas	74.53	AydoganKızılkaya E., et all, 2012, [10]
OM	74.41	Li, C., et all, 2014, [15]
PCA-Ravi	73.63	AydoganKızılkaya E., et all, 2012, [10]
VDM	70.19	Li, C., et all, 2014, [15]
LVDM	62.76	Li, C., et all, 2014, [15]

III. MATERIAL AND METHOD

3.1. Dataset

The data for this study consist of 195 sustained vowel phonations from 31 male and female subjects, of which 23 were diagnosed with Parkinson's Disease (PD). The time since diagnoses ranged from 0 to 28 years, and the ages of the subjects ranged from 46 to 85 years (mean 65.8, standard deviation 9.8). Averages of six phonations were recorded from each subject, ranging from 1 to 36 s in length. See Table I for subject details. Fig. 1 [16] shows plots of two of these speech signals. The phonations were recorded in an Industrial Acoustics Company (IAC) sound-treated booth using a head-mounted microphone (AKG C420) positioned at 8 cm from the lips. The microphone was calibrated as described in [17] using a class 1 sound-level meter (B&K 2238) placed 30 cm from the speaker. The voice signals were recorded directly on computer using Computerized Speech Laboratory (CSL) 4300B hardware (Kay Elemetrics), sampled at 44.1 kHz, with 16-bit resolution. Although amplitude normalization affects the calibration of the samples, the study is focused on measures insensitive to changes in absolute speech pressure level. Thus, to ensure robustness of the algorithms, all samples were digitally normalized in amplitude prior to calculation of the measures. [16]

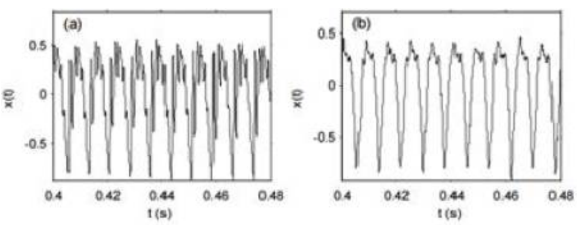

Fig. 1. Two selected examples of speech signals: (a) healthy, (b) subject with PD. The horizontal axis is time in seconds, the vertical axis is signal amplitude (no units)

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 - Several measures of variation in fundamental frequency
- MDVP:Shimmer - 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

The data is in ASCII CSV format. The rows of the CSV file contain an instance corresponding to one voice recording. There are around six recordings per patient, the name of the patient is identified in the first column [18, 19].

Table 2: List Of Subjects With Sex, Age, Parkinson's Stage And Number Of Years Since Diagnosis

Subject code	Sex	Age	Stage (H&Y)	Years since diagnosis
S01	M	78	3.0	0
S34	F	79	2.5	¼
S44	M	67	1.5	1
S20	M	70	3.0	1
S24	M	73	2.5	1
S26	F	53	2.0	1½
S08	F	48	2.0	2
S39	M	64	2.0	2
S33	M	68	2.0	3
S32	M	50	1.0	4
S02	M	60	2.0	4
S22	M	60	1.5	4½
S37	M	76	1.0	5
S21	F	81	1.5	5
S04	M	70	2.5	5½
S19	M	73	1.0	7
S35	F	85	4.0	7
S05	F	72	3.0	8
S18	M	61	2.5	11
S16	M	62	2.5	14
S27	M	72	2.5	15
S25	M	74	3.0	23
S06	F	63	2.5	28
S10 (healthy)	F	46	n/a	n/a
S07 (healthy)	F	48	n/a	n/a
S13 (healthy)	M	61	n/a	n/a
S43 (healthy)	M	62	n/a	n/a
S17 (healthy)	F	64	n/a	n/a
S42 (healthy)	F	66	n/a	n/a
S50 (healthy)	F	66	n/a	n/a
S49 (healthy)	M	69	n/a	n/a

3.2. Software – WEKA

Weka (Waikato Environment for Knowledge Analysis) written in Java was developed at the University of Waikato, New Zealand [20]. Weka supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection. All techniques of Weka's software are predicated on the assumption that the data is available as a single flat file or relation, where each data point is described by a fixed number of attributes (normally, numeric or nominal attributes, but some other attribute types are also supported) [21].

3.3. Methods

IB1: Uses normalized Euclidean distance to find the training instance closest to the given test instance, and predicts the same class as this training instance. If multiple instances have the same (smallest) distance to the test instance, the first one found is used [22, 23].

RotationForest: Generates an ensemble classifier by training a base learner on a randomly selected subspace of the input data that has been rotated using principal component analysis. [24].

Random Forest (RF): RF is an algorithm based on ensemble techniques for classification and regression analysis [25]. The main principle in RF is the combination of various decision trees using several bootstrap samples and selecting a subset of explanatory variables at every node. [26].

MultilayerPerceptron: A multilayer perceptron (MLP) is a feedforward artificial neural network model that maps sets of input data onto a set of appropriate outputs. A MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. [27].

ClassificationViaRegression: It is possible to use regression methods to solve classification tasks. In order to apply the continuous prediction technique of regression models to discrete classification problems, an approximation of the conditional class probability function can be considered. [28].

Bagging: The Bagging (Bootstrap Aggregating) algorithm uses bootstrapping (equiprobable selection with replacement) on the training set to create many varied but overlapping new sets. The base algorithm is used to create a different base model instance for each bootstrap sample, and the ensemble output is the average of all base model outputs for a given input [29, 30].

Java Repeated Incremental Pruning (JRip): JRIP is a prepositional rule learner, i.e. Repeated Incremental Pruning to Produce Error Reduction (RIPPER). Initial rule set for each class is generated using IREP [31].

SMO: Sequential Minimal Optimization (SMO) is a new algorithm for training Support Vector Machines (SVMs). Training a support vector machine requires the solution of a very large quadratic programming (QP) optimization problem. [27].

OneR: Class for building and using a 1R classifier; in other words, uses the minimum-error attribute for prediction, discretization numeric attributes [32].

Naive Bayes Tree (NBTree): The NBTree algorithm is a hybrid between decision-tree classifiers and Naive Bayes classifiers. It represents the learned knowledge in the form of a tree which is constructed recursively [33].

Dagging: This meta classifier creates a number of disjoint, stratified folds out of the data and feeds each chunk of data to a copy of the supplied base classifier. Predictions are made via majority vote. [34]

Decision Table/Naive Bayes (DTNB): Class for building and using a decision table/naive Bayes hybrid classifier. At each point in the search, the algorithm evaluates the merit of dividing the attributes into two disjoint subsets: one for the decision table, the other for naive Bayes. [35].

DecisionTable: Decision Table is an accurate method for numeric prediction from decision trees and it is an ordered set of If-Then rules that have the potential to be more compact and therefore more understandable than the decision trees [36].

J48: J48 algorithm is a modified version of c4.5 and ID3 algorithm which is used to construct the decision trees. The decision tree uses tree like graph and acts as decision support system. [37].

BayesNet: It is probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph. For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. [27].

NaïveBayes: Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. [27].

IV. EXPERIMENTAL STUDY

Following classifier techniques of WEKA have been applied to Parkinson's Disease Datasets: IB1, RotationForest, RandomForest, Multilayer Perceptron, ClassificationViaRegression, Bagging, JRip, SMO, OneR, NBTree, Dagging, DTNB, DecisionTable, J48, BayesNet, NaïveBayes. The results obtained from related classification techniques were presented in Table 3 according to each dataset.

Table 3: Ratio OfEach Classification Technique On Each Parkinson's Disease Dataset

No	Algorithm	Accuracy (%)
1	IB1	96,4103
2	Rotation Forest	92,3077
3	Random Forest	91,7949
4	Multilayer Perceptron	90,7692
5	Classification Via Regression	88,2051
6	Bagging	87,6923
7	JRip	87,6923
8	SMO	87,1795
9	OneR	86,1538
10	NBTree	86,1538
11	Dagging	85,6410
12	DTNB	85,1282
13	Decision Table	81,0256
14	J48	80,5128
15	BayesNet	80,0000
16	Naive Bayes	69,2308

V. RESULTS AND DISCUSSION

In this study the performances of 16 different classification methods were evaluated in terms of classification accuracy on Parkinson's Disease dataset. When comparing the performances of algorithms it’s been found that IB1 (96,4103%) have highest accuracy whereas NaiveBayes (69,2308%) had the worst accuracy. For the future work, more classification algorithms should be applied to more datasets to see impacts of the different performance of algorithms on different datasets. Additionally, one can also study one classification technique to adapt it to have best results on different datasets. This can be done by studying hard on the parameters to related classification technique to adapt them according to the structures of different datasets.

ACKNOWLEDGMENTS

This study has been supported by Scientific Research Project of Selcuk University.

REFERENCES

[1] Adams, R. D., Victor, M., Ropper, A. H., and Daroff, R. B. (1997). Principles of Neurology. Cognitive and Behavioral Neurology, 10(3), 220.
[2] İsenkul M. E. (2001). A Data Collection and Pattern Recognition System for Diagnosis of Parkinson’s Disease, Istanbul University, Graduate School of Natural Sciences, Department of Computer Engineering, Master Thesis, Istanbul, TURKEY.

[3] Fahn, S. and Przedborski, S. (2000) Parkinsonism. In: Rowland LP ed. Merritt's Textbook of neurology. 10 th ed., Lippincott Williams and Wilkins, Philadelphia, p. 679-693.
[4] Gray, P. and Watson, H. J. (1998). Present and future directions in data warehousing. ACM SIGMIS Database, 29(3), 83-90.
[5] Yongjian, F. (1997). Data mining: tasks, techniques and applications. IEEE Potentials, 16(4), 18-20.
[6] AydoganKızılkayaE. ,Karaoglan, I. and Pardalos, P. M. (2012). hGA: Hybrid genetic algorithm in fuzzy rule-based classification systems for high-dimensional problems. Applied Soft Computing, 12(2), 800-806.
[7] Atiya, A. F., and Al-Ani, A. (2009). A penalized likelihood based pattern classification algorithm.Pattern Recognition, 42(11), 2684-2694.
[8] Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Systems with Applications, 37(2), 1568-1572.
[9] Chakraborty, S. (2011). Bayesian semi-supervised learning with support vector machine. Statistical Methodology, 8(1), 68-82.
[10] Zhang, L., and Suganthan, P. N. (2014). Random forests with ensemble of feature spaces. Pattern Recognition,47(10), 3429-3437.
[11] Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Transactions on Biomedical Engineering 2009.
[12] Salama, K. M., Abdelbar, A. M., Otero, F. E., and Freitas, A. A. (2013). Utilizing multiple pheromones in an ant-based algorithm for continuous-attribute classification rule discovery. Applied Soft Computing, 13(1), 667-675.
[13] Li, Y., and Leng, Q. (2015). Alternating multiconlitron: A novel framework for piecewise linear classification. Pattern Recognition, 48(3), 968-975.
[14] Ene, M. (2008). Neural network-based approach to discriminate healthy people from those with Parkinson's disease. Annals of the University of Craiova-Mathematics and Computer Science Series, 35, 112-116.
[15] Li, C., Jiang, L., and Li, H. (2014). Local value difference metric. Pattern Recognition Letters,49, 62-68.
[16] Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., and Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE transactions on biomedical engineering, 56 (4), 1015-1022.
[17] J. Svec, P. Popolo, and I. Titze, "Measurement of vocal doses in speech: Experimental procedure and signal processing," Logoped. Phoniatr. Vocol., vol. 28, pp. 181–192, 2003.
[18] Blake A.C.L. and Merz C.J. (1998). University of California at Irvine Repository of Machine Learning Databases, http://www.ics.uci.edu/~mlearn/MLRepository.html. http://archive.ics.uci.edu/ml/datasets/Parkinsons Last Access: 20.10.2016
[19] Vadavalli A.K. (2016). Comparative Analysıs of Parkinson's Disease, International Journal of Pharmacy and Technology, ISSN: 0975-766X, Vol. 8, Issue No.2, pp.12934-12939.
[20] Weka, http://www.cs.waikato.ac.nz/~ml/weka/, Last access: 10.09.2016.
[21] Arora, R. (2012). Comparative analysis of classification algorithms on different datasets using WEKA. International Journal of Computer Applications, 54(13).
[22] Rodriguez, J. J., Kuncheva, L. I. and Alonso, C. J. (2006). Rotation forest: A new classifier ensemble method. IEEE transactions on pattern analysis and machine intelligence, 28(10), 1619-1630.
[23] WEKA Classifiers, http://weka.sourceforge.net/doc.stable/weka/classifiers/lazy/IB1.html Last access: 10.09.2016.
[24] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 11(1), 10-18.
[25] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
[26] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2(3), 18-22.
[27] Weka, http://www.cs.waikato.ac.nz/~ml/weka/ Last access: 10.04.2015.
[28] Frank, E., Wang, Y., Inglis, S., Holmes, G., & Witten, I. H. (1998). Using model trees for classification. Machine Learning, 32(1), 63-76.
[29] Breiman, L. (1996). Bagging predictors. Machine learning, 24(2), 123-140.
[30] Barutcuoglu, Z. and Alpaydin, E. (2003). A comparison of model aggregation methods for regression. In Artificial Neural Networks and Neural Information Processing—ICANN/ICONIP 2003 (pp. 76-83). Springer Berlin Heidelberg.
[31] Cohen, W. W. (1995, July). Fast effective rule induction. In Proceedings of the twelfth international conference on machine learning (pp. 115-123).
[32] Holte, R. C. (1993). Very simple classification rules perform well on most commonly used datasets. Machine learning, 11(1), 63-90.
[33] Kohavi, R. (1996, August). Scaling Up the Accuracy of Naive-Bayes Classifiers: A Decision-Tree Hybrid. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, (Vol. 96, pp. 202-207).
[34] Bhoomi Trivedi, Neha Kapadia, (2012). INDUS institute of Eng. and Tech,TCET, Kandivali(E),Ahmedabad Modified Stacked generalization with Sequential Learning. TCET 2012 on IJCA.
[35] Hall, M. A., & Frank, E. (2008, May). Combining Naive Bayes and Decision Tables. In Proceedings of the 21st Florida Artificial Intelligence Society Conference (FLAIRS) Vol. 2118, pp. 318-319.
[36] Kohavi, R. (1995, April). The power of decision tables. In European conference on machine learning (pp. 174-189). Springer Berlin Heidelberg.
[37] Quinlan, J. R. (1996). Improved use of continuous attributes in C4. 5. Journal of artificial intelligence research, 4, 77-90.

★★★