Prediction of Parkinson's disease using Ensemble Machine Learning classification from acoustic analysis

Journal of Physics: Conference Series **1372** (2019) 012041 doi:10.1088/1742-6596/1372/1/012041

Prediction of Parkinson's disease using Ensemble Machine Learning classification from acoustic analysis

Amit Kumar Patra[1], Ratula Ray[2], Azian Azamimi Abdullah[3], Satya Ranjan Dash[4]*
[1]Tata Consultancy Services Limited, USA
[2]School of Biotechnology, KIIT University, Bubaneswar, India
[3]School of Mechatronic Engineering, Universiti Malaysia Perlis, Malaysia
[4]School of Computer Applications, KIIT University, Bubaneswar, India

*sdashfca@kiit.ac.in

Abstract. Parkinson's disease (PD) is a neurodegenerative disorder, which upon progression affects the movements. Tremors associated with Parkinson's disease are the major symptoms to look out for in such cases. This generally results in breakdown of the neurons producing dopamine. Qualitative speech starts to decline as the disease progresses and the variability in the vocal cord vibration (also known as fundamental frequency) starts to occur. People with PD have shown to produce greater variability in frequency as compared to normal people. In our paper, we are focused on comparison of the voice measurement features of patient dataset to understand whether a patient is suffering from PD or not using Machine Learning classifiers. We have implemented Decision Trees, Logistic Regression and K-nearest neighbors as base classifiers and have compared their performance with Ensemble learning classifiers Bagging, Random Forest and Boosting. We have compared the accuracy (%) of the classifiers and discussed which one of them is more accurate at predicting the outcome of the disease. We also found out the most relevant features associated with the classification and ranked them based on feature importance. Our main aim here is the classification of healthy individuals from people suffering from PD by detection of dysphonia (difficulty in speaking due to declining health conditions).

1. Introduction

With increase in the aging population, the people affected with neurological disorders are also increasing rapidly and thus, the need for quick and effective means for monitoring them becomes more and more important. It has been noticed that in almost 90% of the cases, the people suffering from PD experience impairment of the vocal cords, which result in dysphonia. At much more severe stage, articulatory deficit can also manifest along with vocal impairment [1].

Understanding the measurements related to voice detection thus can play an important role in predicting the progression of the disease. According to the traditional sound measurement standards, jitter, shimmer, pitch, noise to harmonics ratio and pressure level of absolute sound have been used effectively. Inspired by nonlinear dynamical systems theory, very recently, few other measures have also been included in the system which can be used for the purpose of sound measurement. These tools used for voice disorders detection include Recurrence Period Density Entropy(RPDE) and Detrended Fluctuation Analysis(DFA)[2].

In our paper, we are aiming to apply the Machine learning classifiers in order to predict the status of health of the patients (that is, if the person is suffering from PD or not) fusing acoustic analysis . We have taken under consideration the various traditional as well as newer non-linear tools for voice measurements as the attributes for classification of the health status of the individuals. So along with classification, we will be also looking into the most important features that are responsible for making the decision and check if the newer attributes are better than the traditional ones.

2. Literature review

Previous studies have shown the implementation of Neural networks in assessment of the risk factors associated with Parkinson's disease. Factors such as genes, age, stroke and diabetes were found out to have higher correlation with the outcome of the prediction [3].

Feature relevance analysis has also been carried out previously using Parkinson's voice dataset. Tensor flow deep learning library has also been used to detect the increase in disease severity in case of Parkinson's disease [4].

Subtype identification of PD has also been done before using both supervised and unsupervised machine learning methods. The progression rate of the disease can be predicted using these learning systems and can be classified as: highly progressive, moderately progressive and mild progressive state of Parkinson's Disease. This has been utilized effectively in order to improve the quality of the counseling of patients, improve clinical trial design and using resources available in a better way [5].

Work has also been carried out on MRI dataset of the brain for identification of medical image related biomarkers. Supervised Machine Learning classifiers such as Support Vector Machine has been utilized in order to understand the progression of the disease and role of the biomarkers in identifying the progression behavior [6].

Gait classification in case of patients suffering from PD has also been studied using Artificial Neural Networking (ANN) and Support Vector Machine classifiers. Kinematic, kinetic and spatiotemporal gait are the 3 basic parameters, which are to be taken under consideration before performing the classification task [7].

3. Methods

3.1. Dataset

We have used the Parkinson's dataset of UCI Repository which comprises of measurements of voices of 31 people, out of which 23 of them suffer from PD and the rest are healthy individuals[8]. The non-linear measurements are also added to the dataset, which are used for classification purpose along with the traditional measurements[9].

3.2. Dataset Attributes

The different acoustic signal measurements chosen as attributes for the classification of the presence and absence of Parkinson's disease are as follows:

- MDVP:Fo(Hz) - Average vocal fundamental frequency
- MDVP:Fhi(Hz) - Maximum vocal fundamental frequency
- MDVP:Flo(Hz) - Minimum vocal fundamental frequency
- MDVP:Jitter(%),MDVP:Jitter(Abs),MDVP:RAP,MDVP:PPQ,Jitter:DDP - Several measures of variation in fundamental frequency

- MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,MDVP:APQ,Shimmer:DDA - Several measures of variation in amplitude
- NHR,HNR - Two measures of ratio of noise to tonal components in the voice
- status - Health status of the subject (one) - Parkinson's, (zero) - healthy
- RPDE,D2 - Two nonlinear dynamical complexity measures
- DFA - Signal fractal scaling exponent
- spread1,spread2,PPE - Three nonlinear measures of fundamental frequency variation

3.3. Classifiers

3.3.1. Base Classifiers

- Decision Tree: In case of Decision Tree classifier, the input is split into sub-spaces based upon certain functions. It helps in reaching a conclusion based upon conditional control statements[11].
- Logistic Regression: This classification occurs on the basis of a sigmoid function, known as the logistic function which takes in a real input and gives out a value in between 0 and 1[12].
- k-Nearest Neighbors: It is a "k" biased classification technique where majority voting among the neighborhood data points are taken under consideration for classification, which may or may not rely upon the distance based weighting parameters[13]. We have considered the value of n=5 in our classification problem.

3.3.2. Ensemble Classifiers

- Random Forest Classifier: This is an ensemble learning classifier which randomly generates decision trees and averages the result, thereby reducing over-fitting of the model[14]. It selects only a subset of features randomly and the best feature is used to split at the node.
- Bagging Classifier: This classification technique is almost like Random Forest but differs from it in the sense that it does not consider any subset of features for splitting. Rather, it takes into consideration all the features available for splitting at the node[15].
- AdaBoost Classifier: Also known as Adaptive Boosting classification technique, this ensemble learner works upon fitting the weak classifiers and improve the estimate upon each iteration[16].

3.4. Validation procedure

In the paper, we have used 10- fold cross validation over the entire dataset. In 10-fold cross validation system, the entire dataset is divided into 10 parts with equal number of data in each part. We have to then consider (10-1) to be the training model and the rest is the testing model. This is repeated for 10 times with (10-1) as training set and the rest as validation(testing) set [10].

3.5. The metrics used for this purpose are as follows:

$$\text{Recall}=\frac{\boldsymbol{True\ Positive}}{\boldsymbol{True\ Positive+False\ Negative}}$$

$$\text{Precision}= \frac{\boldsymbol{True\ Positive}}{\boldsymbol{True\ Positive+False\ Positive}}$$

$$\text{F1-score}= \frac{\boldsymbol{2*(Recall*Precision)}}{\boldsymbol{(Recall+Precision)}}$$

$$\text{Accuracy}= \frac{\boldsymbol{True\ Positive}}{\boldsymbol{True\ Positive+False\ Positive}}$$

4. Results and Discussion

We have used three base classifiers whose performance has been compared in Table 1, based on the statistical metrics being used.

Table 1. Performance analysis of the Base classifiers

Base Classifiers	**Precision**	**Recall**	**F1-score**	**Accuracy%**
Decision Tree	83%	82%	82%	79.95%
Logistic Regression	81%	82%	81%	82.03%
K-nearest neighbors(n=5)	75%	76%	75%	76.31%

The result obtained suggests that the performance has increased substantially for the ensemble classifiers. For the base classification system, Logistic Regression showed good results in terms of accuracy, though Decision Tree shows better stability as a classifier in terms of f1 score.
Ensemble classifiers are the set of classifiers which takes under consideration the performance of individual base classification systems and give the final result. We have used both averaging method (Bagging and Forests of randomized trees) and sequential method (Boosting) here to build upon the result of the base decision tree classifier. Table 2 shows our result for the same.

Table 2. Performance analysis of the Ensemble classifiers

Ensemble Classifiers	**Precision**	**Recall**	**F1-score**	**Accuracy%**
Random Forest	83%	82%	82%	84.59%
Bagging	80%	81%	80%	81.11%
Boosting(Adaboost classifier)	83%	83%	83%	82.51%

Inappropriate management of PD is very common and can be often misdiagnosed as depression. Acoustic analysis provides a more distinct approach for differentiating between the patients suffering from both of these diseases by considering the temporal measure of the speech. Understanding the most relevant features associated with the status of the patients in case of Parkinson's Disease is also very important and has been depicted in Fig. 1. We have selected the best performing classifier (in this case, Random Forest), and have evaluated the most important attributes associated with the prediction of PD in a set of individuals. For a forest of decision trees, the measure of how much each feature decreases the weighted impurity in a tree is important to note. The mean of the impurity decrease can be used to rank the features accordingly.

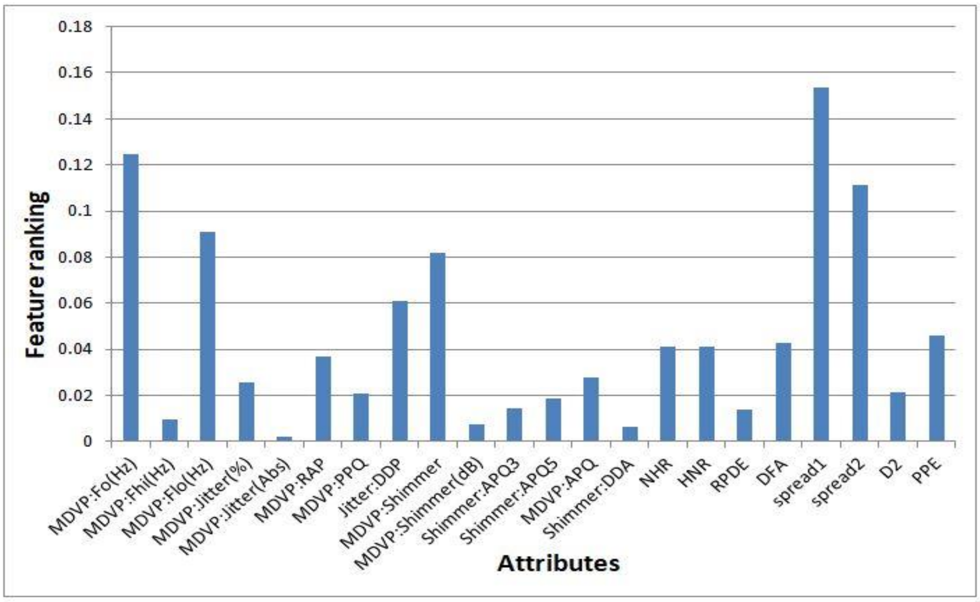

Figure 1. Feature ranking based on the importance of the features associated with the classification

5. Conclusion

Our observations have shown that Ensemble Classifiers show much better accuracy as compared to the base classifiers in most of the cases. Out of them all, Random Forest Classifier gives the best result with an accuracy of 84.59%, though if we consider the f1-scoring system as the standard metrics for classification, Boosting shows a slightly better result (83%). Both of these ensemble methods can thus be used to classify the dataset effectively. Among the base classifiers, Logistic Regression performs quite well with an accuracy of 82%.

Our observations also highlight the most important features that are to be considered for prediction of dysphonia in patients suffering from PD. Non-linear measure of fundamental frequency variations (spread 1 and spread 2) have shown to have much higher importance and is thus thought to be most related in the prediction outcome. Among the traditional methods used, Average vocal fundamental frequency and Minimum vocal fundamental frequency have showing higher ranking, thus showing more correlation to the prediction outcome of the classification.

The main highlight of this paper is to showcase the power of Ensemble classifiers over base classification systems and how combining the power of multiple heterogeneous classifiers can give a more reliable accuracy, free from bias. Using f1 scoring system as the standard metric for assessing the performance of the classifier gives us an opportunity for taking into account both Recall and Precision. Unlike accuracy, it gives a more balanced result by considering the False Negative while computing the final score.

References

[1] Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1999). Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural neurology*, *11*(3), 131-137.

[2] Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A., & Moroz, I. M. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Biomedical engineering online*, *6*(1), 23.

[3] Joshi, S., Shenoy, D., Rrashmi, P. L., Venugopal, K. R., & Patnaik, L. M. (2010, February). Classification of Alzheimer's disease and Parkinson's disease by using machine learning and neural network methods. In *2010 Second International Conference on Machine Learning and Computing* (pp. 218-222). IEEE.

[4] Grover, S., Bhartia, S., Yadav, A., & Seeja, K. R. (2018). Predicting Severity of Parkinson's Disease Using Deep Learning. *Procedia computer science*, *132*, 1788-1794.
[5] Faghri, F., Hashemi, S. H., Leonard, H., Scholz, S. W., Campbell, R. H., Nalls, M. A., & Singleton, A. B. (2018). Predicting onset, progression, and clinical subtypes of Parkinson disease using machine learning. *bioRxiv*, 338913.

[6] Salvatore, C., Cerasa, A., Castiglioni, I., Gallivanone, F., Augimeri, A., Lopez, M., ... & Quattrone, A. (2014). Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy. *Journal of Neuroscience Methods*, *222*, 230-237

[7] Tahir, N. M., & Manap, H. H. (2012). Parkinson Disease Gait Classification based on Machine Learning
Approach. *Journal of Applied Sciences*, *12*(2), 180-185.

[8] Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE transactions on bio-medical engineering*, *56*(4), 1015.

[9] Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A., & Moroz, I. M. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Biomedical engineering online*, *6*(1), 23.

[10] Sharma, R. C., Hara, K., & Hirayama, H. (2017). A machine learning and cross-validation approach for the discrimination of vegetation physiognomic types using satellite based multispectral and multitemporal data. *Scientifica*, *2017*.

[11] Mashat, A. F., Fouad, M. M., Philip, S. Y., & Gharib, T. F. (2012). A decision tree classification model for university admission system. *Editorial Preface*, *3*(10).

[12] Stylianou, N., Akbarov, A., Kontopantelis, E., Buchan, I., & Dunn, K. W. (2015). Mortality risk prediction in burn injury: Comparison of logistic regression with machine learning approaches. *Burns*, *41*(5), 925-934.

[13] Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003, November). KNN model-based approach in classification. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"* (pp. 986-996). Springer, Berlin, Heidelberg.

[14] Kumari, G. P. (2012). A Study Of Bagging And Boosting Approaches To Develop Meta-Classifier.

[15] Freund, Y., Schapire, R., & Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, *14*(771-780), 1612.