Classification of Spirometer data using Feature Specific Outlier extraction and Multiple Classifiers

Vigneshwaran Santhalingam, *MIT*

***Abstract*— With all the air pollution, smoking and other lung infections, the respiratory ailments burgeoned drastically over the past decade, affecting more people every day. The difference between each ailment is so negligible that it is nearly impossible to correctly predict the actual reason for the cause of the disease. If there exists an efficient tool for analyzing this respiration data, a large proportion of the ailments can be cured with greater ease as compared to the past. In this paper we propose a novel ensemble machine learning approach, named Feature Segregator, for the specific purpose of classification of the Medical Dataset. Feature Segregator algorithm has proven to be superior to almost all the other micro classifiers and macro classifiers with accuracy as high 98% and an appreciable *f*-score. This Feature Segregator is based on feature specific outlier extraction.**

***Index Terms*—Dominant Subset, Feature Segregator, One Class-SVM, Outliers, Recessive Subset.**

I. Introduction

RESPIRATION is a complex mechanism and one of the most vital systems in the body. In the recent days, due to severe air pollution, smoking, and various other pathological conditions, external respiration has become difficult for many individuals and they contribute to a large percentage of mortality. Some of the most common pulmonary ailments include, asthma, bronchitis, Chronic Obstructive Lung Disorder (COLD) etc., pose a serious threat to the pollution at mass. Most of these acute and chronic illness can be cured when detected in the preliminary stages.

Almost all lung problems can be classified as being Obstructive and Restrictive [2]. While Obstructive Disorders occur in people whose airways have become narrow due to some infection or allergy or inflammation, leading to an increased time to exhale or empty the lungs, Restrictive Disorders frustrate people from inhaling fully as the lungs are *restricted* from fully expanding [3]. Spirometers measure the volume of air inhaled or exhaled as a function of time during the process of external respiration. Spirometer can be considered a diagnostic tool for analyzing Obstructive Disorders. The Spirometer data (N=100) used for this research has been collected from volunteers using commercially available devices and following the standard data acquisition protocol.

The collected data is classified into three classes of Obstructive Disorder, namely: Normal (class 0), Mild Obstructive (class 1), Severely Obstructive (class 2).

Classification is the most commonly used supervised machine learning method. A Classification algorithm learns from the training data about the labels and classifies and the labels of unseen data. Currently there are many micro classifiers for classifying data. But the problem with micro classifiers is that they may work efficiently on a certain portion of data and may not meet the expectations in certain other portions of data. That's when we bring in ensemble techniques like, *Bagging and Boosting* etc., where several basic models collaborate with each other for classification with the one goal to achieve more stable, reliable and accurate results. Ensemble techniques have proven to be more efficient than the basic classifiers both theoretically [8-9] and empirically [4-6].

We propose a novel ensemble method for biomedical data, Feature Segregator which classifies not only the biomedical data with absolute accuracy but also works perfect any other industrial data also for which it was never intended. Feature Segregator using multiple micro classifiers and then combine their output together to form a larger macro classifier. As with the already existing ensemble techniques, each micro classifier works on a subset of data, but in contrast to the other methods, the instances for each classifier are not chosen at random and none of the sets overlap each other. More about this will be seen in the subsequent sections.

References

[1] Pramila Vijayaraghavan, Mahesh Veezhinathan, Performance assessment of Optimized Extreme Learning Machine based on Evolutionary Computing for Spirometric Data Classification, WSEAS Transactions On Biology And Biomedicine, Volume 11, 2014, E-ISSN: 2224-2902

[2] Elman J. 1990. Finding Structure In Time, Cognitive Since. 14:179-211. DOI: 10.1016/0364- 0213(90)90002-E.

[3] Abboud, S., Barnea, O. Guber, A., Narkiss, N. And Bruderman,I. (1995). Maximum expiratory flow-volume curve: mathematical model and experimental results. Med. Eng. Phys. 17 (5), 332-336

[4] S.Hashem, Optimal linear combinations of neural networks , Neural Networks. 10(1997)599–614.

[5] R.Maclin, D.Opitz, Popular ensemble methods: An empirical study. arXiv:1106.0257.(2011).

[6] D.W.Opitz ,J.W.Shavlik, Generating accurate and diverse members of a neural-network ensemble, Adv.Neural Inf. Process.Syst.8(1996)535–541.

[7] SašoKarakatic, ViliPodgorelec, Improved classification with allocation method and multiple classifiers, Information Fusion 31 (2016) 26–42, Elseiver.

[8] L.K.Hansen, P.Salamon, Neural network ensembles, IEEE Trans. Pattern Anal. Mach. Intell.12(1990)993–1001.

[9] A.Krogh, J.Vedelsby, Neural network ensembles, crossvalidation, and active learning, Adv.Neural Inf Process.Syst.(1995)231–238.