Feature Selection Based on Robustness for Classifying Lung Disease in Computer Tomography using Structural and Texture Feature

Anju Abraham[1] Shyma Kareem[2]
[1]M.Tech Student [2]Guide
[1,2]Department of Computer Science & Engineering
[1,2]Musaliar College of Engineering & Technology, Kerala, 689653, India

Abstract— Computer aided system is not able to generalize the lung diseases . The main reason behind this is that in computer tomography, quantitative algorithm are not robust enough during image acquisition and reconstruction. It do not consider the technical factor. Thus algorithm is used that considers the robustness of the feature based on the technical parameter of the image. We evaluated various topographic images to classify various lung diseases. These images were separated into three datasets based on there texture and structural features. The classes of diseases that is considered includes idiopathic pulmonary fibrosis, interstitial pneumonia, cystic fibrosis, etiology, airways. Two classifier was created one with robustness feature and another without any robustness feature. Performance is compared on classifiers by using the mean, precision, recall, kappa and standard deviation

Key words: Computed Tomography, Pattern Reorganization, Interstitial Lung Diseases, Idiopathic Pulmonary Fibrosis

I. INTRODUCTION

There are numerous pros of computer aided system as compared to the traditional approaches like prefabricated implant based on scaling and reducing inter and intra reader variability. The major cons behind this is its lack of generalization. Failure of generalization is due to its quantitative algorithm are not robust while considering the technical parameter during image acquisition and reconstruction. As per the example, in idiopathic pulmonary fibrosis, it has been illustrated that infected index measures based on computer aided system measurements can change significantly on computation based on different kernels or slice thicknesses [1]–[3]. Thus to produce generalized computer aided system, it is mandatory that the underlying algorithms must be robust enough to variations in image acquisition and reconstruction factors.

Computer topographic based classification in order to classify fibrotic disorders becomes challenging due to differences in the imaging equipments and due to technical factors like thickness of the image, kernel, vessels which influences the texture and structure of the image[4]. This can be overcome by maintaining the uniformity in the computer tomography protocols. Another solution is considering various factor and for each of it developing a separate classifier by training it accordingly. But there are various technical and logical shortcomings in collecting large amount of samples thus computer topographic system is implemented in the limited environment. We proposed the system that classify the structural and textural features and parameters by considering robust feature against the threshold value.

The previous work in the area of computer aided system found out that dataset with three different slice thickness of the different image gave the same performance irrespective of different images[10]. This provided the framework for evaluating the robustness of various other algorithms and comparing it with the system.

This paper presents the algorithm for selecting feature that crosses the threshold value and the benchmark that is set upon the robustness of the image on slice thickness and kernel. It prioritizes the robust feature. By applying robustness based feature selection yields a reduced robust feature that resulted into a consistent classifier. It thus maintains the accuracy of the classifier. Performance is also compared on the system with robustness feature and without robustness feature.

II. FEATURE SELECTION BASED ON ROBUSTNESS

We proposed a feature selection algorithm that improves the robustness of the computer aided systems classifier without any trained dataset. It maintains the stable result even with different inputs. It is assumed that robust feature is directly related to the robust classifier.

A. Procedure

We firstly calculate the robustness of the feature in the dataset and each of the feature is marked with the index value. We then discard the feature with larger than the threshold value. This resultant feature is further subjected to the other existing feature selection methods. This is summarized in the Fig 1.

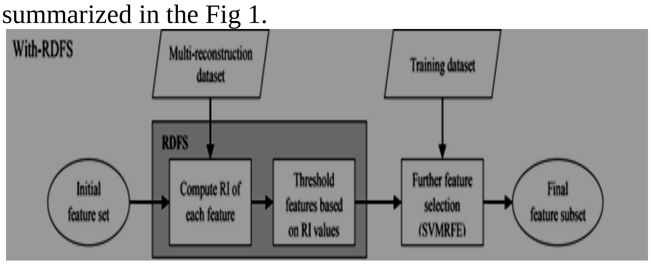

Fig. 1: Feature selection for classifier

Robustness of the Feature is classified by considering thecomputer topographic raw dataset , the image is constructby considering different parameters. Raw data is again reconstructed with numerous subjects. One of the setis considered as the reference dataset and all other imagesare calculated against this.ndex Value Calculation

Index value is calculated by using multi reconstructed dataset. Robustness value is first obtained by volume of interest and is constructed again. One of the value is made as the benchmark value. Standard deviation is computed with respect to the benchmark value for the given dataset. Then the values are considered as paired value in which benchmark value is the first value and second is taken as the reconstructed value. Paired differences are calculated. Finally standard deviation is obtained with paired values.

Standard deviation is calculated for the technical factor and for the biological factor. Thus robustness index is the ration of the standard deviations of both technical factor and biological factor. Greater index value indicates that technical parameter is more than biological factor. Less value of index infers that the feature is robust. Next the threshold index is found out in order to remove feature that is not robust.

III. METHODS

Dataset is obtained from the computer tomography system. Obtained dataset is classified into three sets training, testing and reconstructed sets. High resolution images are selected for the training and testing dataset. For reconstructed set raw computer aided system data is considered. The diseases of interest are pulmonary lung disease, airways, honeycombing and ground glass. In our experiment initially the classifier is trained with the collected data. Classifier used here is support vector classifier. Feature is standardized with. Feature selection is done by using robustness procedure. Feature that is not robust are eliminated. After this one more feature selection is done by using support vector machine to obtain a better output multiple support vector feature elimination machines can be used. Two different classifier model is constructed, one with robustness and without robustness model. Two classifier models are compared. One of the model is to evaluate reconstructed multiple dataset. The aim was to and enhance robustness of the image by using robustness procedure. We took images with larger technical factors for this purpose. Mean and kappa formula was used to evaluate robustness of the feature.

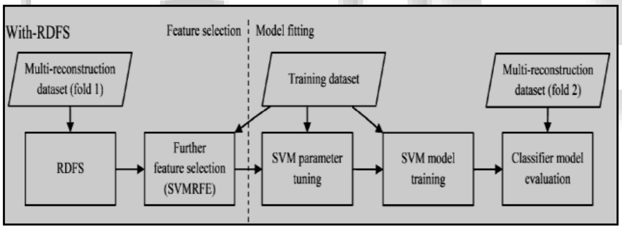

Fig. 2: Feature evaluation on reconstructed set

The second procedure is to evaluate classifier models against the standalone testing dataset. The main aim of this set up was to make sure that the robust feature that has been extracted meets both robustness feature as well as the improved performance. Recall, kappa, precision and mean is used to evaluate the performance. In this case robustness is calculated by using robustness procedure and tuning of parameter, training of image is done by using trained dataset.

IV. OUTCOME

Here first we characterized robust feature and non robust feature. Then we evaluated against the multiple reconstructed dataset and then against the standalone dataset. We give first the input image, obtain the interpolated image and Gaussian blurred image and classify the image as pulmonary lung disease, ground glass, airways etc. based upon the structural and textural features of the image. Input image given identifies the diseases against the trained dataset. The diseases can be identified as any of the following diseases like pulmonary fibrosis, ground glass, honeycombing, normal lung parenchyma and airways.

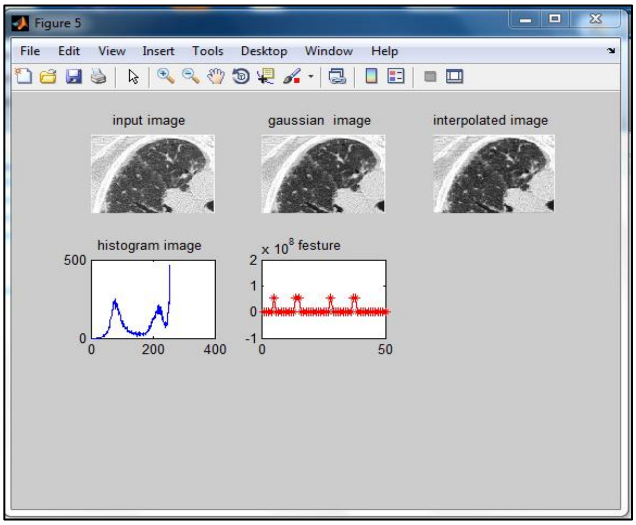

Fig. 3: Gaussian, interpolated image of airways

Gaussian blurring image and sub image size based on the window size like 0.5,1.0,2.0 and 4.0. It is calculated for each level of scale. The robustness index is calculated on each of these windows, smaller value represents very good robustness and larger value indicates poor robustness. A threshold RI is found out and based on that the corresponding parameter is classified as good or poor robust value.

Interpolated image is obtained by resizing and distorting the image from one pixel grid to another. Image resizing is done to increase and decrease the total number of pixels. The above input is filtered to obtain the Gaussian image and then the interpolated image it gives the better clarification to the resultant image.

Fig. 4: Dialog Box representing airways

V. CONCLUSION

We were able to develop a robust system for features election. This improved the robustness of the classifier to identify the type of the lung diseases. It could maintain the performance of the system on standalone testing dataset. Improving the generalization of the system on multiple parameter. This approach helps in the clinical diagnosis of diseases.

REFERENCES

[1] K.L Boedeker., "Effect of reconstruction algorithm on CT imaging measures" Radiology, no. 1: pp.295-301. 2004.
[2] R.Akbani., "Applying support vector machine to imbalanced datasets" Springer, no. 1, 2004.
[3] Y.Xu., "Computer aided classification of interstitial lung diseases" no. 22, 2009.

[4] S.T Bartel,et.al., ”equating quantitative emphysema measurements on different CT image reconstruction.” Med.phy., 19th Annual, pp. 4894-4902, 2010.
[5] H.J. Kim et al. ”Classification of parenchymel abnorma. In lung using novel approach” Acad.Radiol, no. 8 ,vol. 28. 2010.
[6] J. A. Cohen, ”A Coefficient of agreement for nominal scales” Psychology., vol. 20, 2011
[7] M.Hall and L. Breiman., ”The WEKA data mining software: An update,” Radiology, no. 1., 2013.
[8] C. W. Hsu and C.J. Lin., ”A comparison of methods for multiclass support vector machine,” IEEE Transaction, no. 2., 2013.
[9] R.M. Haralick, Shanmugan, and I.Dinstein., ”Texture feature for image classification,” IEEE Transaction, no. 6, 2014.
[10] R.M. Haralick., ”Statistical and structural approaches to texture ,” IEEE Transaction, no. 5, 2015.
[11] M. Galloway., ”Texture analysis using run length matrices,” IEEE Transaction, no. 2, 2015.