Improved MLP-NN based approach for Lung Diseases Classification

Ramandeep Kaur
M.Tech CSE Dept
CT Institute of Engg. Mgt. &Tech.
Punjab, India

Prince Verma
Assistant professor CSE Dept.
CT Institute of Engg. Mgt.& Tech
Punjab, India

ABSTRACT
The Data Mining is extracting or mining knowledge from large volume of data. Classification technique is used in different-2 application. In this paper proposes a new classifier utilizing MLP approach by grouping based on nearest neighbor i.e. improved MLP-NN. The MLP-NN approach can handle noisy data and reduce complexity. This technique has been applied for medical diagnosis. This paper analyzes the lung images (i.e. CT-scan images) for identifying and classifying them among the various lung diseases (i.e. bronchitis, emphysema, pleural effusion or normal) using 100 images data set and 80 images data set.

Keyword
Data Mining; Classification; Multilayer Perceptron.

1. INTRODUCTION
Classification is a data mining process that assigns items in a group to target classes. Data mining discovers designs and also relationships using data evaluation equipment and methods to construct models. There are 2 forms of models in data mining. One is predictive models i.e. based on probability of outcomes. Another is descriptive models, is a mathematical method that explains real-world functions and the relationships between factors responsible for them. The term Data Mining, also referred to as Knowledge Discovery in Databases. Classification represents a supervised learning method as well as a statistical method for classification. The MLP classification algorithm is used to classify because MLP do not make any assumption relating to the underlying probability density functions, Fast to run and high tolerance to noisy data.Avishkar Misra et al. [16] The automated lung segmentation able of segmenting the lung form all regions like top, Middle and Basal. The accurate and characterize segmentation of the respiratory organ is important for lung diseases. In this compare the two strategies. that's anatomy approach that is use for anatomical landmark detection this detection define the separation points between the regions the other methodology is machine learning approach that describe the lung size, shape and location properties and additionally classify a given respiratory organ into the appropriate region. Anita chaudhary et al. [7] in this paper discuss computed tomography (CT) in many cases are better than X-ray. And drawback arises as a result of period constraint in detecting the current of pulmonary cancer regarding on the varied diagnosis method used. Hence, a pulmonary detection system using image method can be utilized to classify the present of lung cancer in a CT- image. MATLAB are used through each procedures created. And also discuss more accurate result using segmentation and enhancement technique.

In the past few years, medical CT Images are widely-used in medical diagnosis. Computed Tomography images might be recognized for several body tissues based on their distinct gray levels. Computed tomography, more commonly called a CT scan. This diagnostic medical test done like classic X-rays produce multiple images inside of the body. The cross-sectional of CT scan images produced variety planes of body and this output can be generate in 3-D view. CT imaging had evolved far enough in terms of speed and resolution to make it a valuable tool in the imaging of the lungs these images can be viewed on a PC or printed on film. The lungs are an imperative a part of the whole body as thousands of times each day it performs responsibility for supplying oxygen to blood whereas exhaling Carbon dioxide. Many people suffer from lung disease which occurs due to smoking, infection and biological reasons.

This paper considers 4 categories of images i.e. bronchitis, emphysema, pleural effusion and normal. Bronchitis is a soreness as well as enlargement of the bronchial tubes, the air flow passages between the nose area as well as the lung tissue. Emphysema in the pleural cavity that will grow when germs infect the pleural space, generally in the context of pneumonia. Pleural effusion is excess liquid that accumulates in the pleural cavity, the fluid-filled space that surrounds the lung tissue. This excess may damage inhaling and exhaling by decreasing the growth of the lung tissue.

Classification is done with new classifier i.e. Improved MLP-NN approach that can handle complexity and increases the accuracy. The Proposed Approach would be executed in four phases i.e. in the very first phase, the preprocessing is done by a medium filter and morphological smoothening, then features are extracted from preprocessed image by MAD Technique, the features are selected by applying genetic algorithm to select the top ranked features. In the finalized phase, the classifiers Multi-Layer Perception Neural Networks (MLP-NN) and Upgrade MLP-NN are used to classification of the lung diseases

2. MATERIAL AND METHOD
The classification of CT scan images for different Lung diseases like Bronchitis, Pleural effusion, Emphysema and Normal are considered in this work. The age of patients whose lung images are considered, ranges from 30 to 70. The images are taken from radiopaedia.org website. Two datasets are considered. The first data set contains 80 images, 20 images of each disease; second data set contains 100 images, 25 images of each disease. Classification helps to classify the items. It classify according to the feature of images and texture of images. With the help of new algorithm easily classification is done and also reduces time. In lungs CT scan image identification of disease is difficult but these algorithms can easily identify images within minute.

3. PROPOSED APPROACH
3.1 Preprocessing
Preprocessing is a program that processes its input data to produce output that is used as input to another program. In the Proposed Approach preprocessing is performed by two filters i.e. Morphological Smoothening Filter and Median Filter. This filter is help to enhance the image. The median filter is utilized to remove the noise and salt while the morphological

smoothening filter enhances the image by erosion and dilation. The erosion is process of removing pixel from edge of image and dilation is the reverse process.

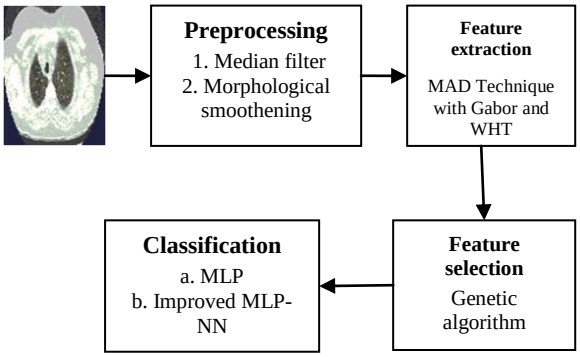

Figure 1: Block diagram of Proposed Approach

3.2 Feature Extraction

Feature extraction is method of extracting an exceptional feature by including a current feature. Feature extraction is extremely distinctive from Feature selection: the former in changing arbitrary data, for example text or images, into numerical features usable for machine learning. In our Proposed Approach Feature extraction is conducted by MAD with Gabor filter is combined using template level fusion. Gabor gets the salient visual feature corresponding to spatial localization, orientation selectivity and spatial frequency. Feature extraction is done by MAD with gabor and WHT. With this fusion it could be transformed into values that estimate the standard deviation

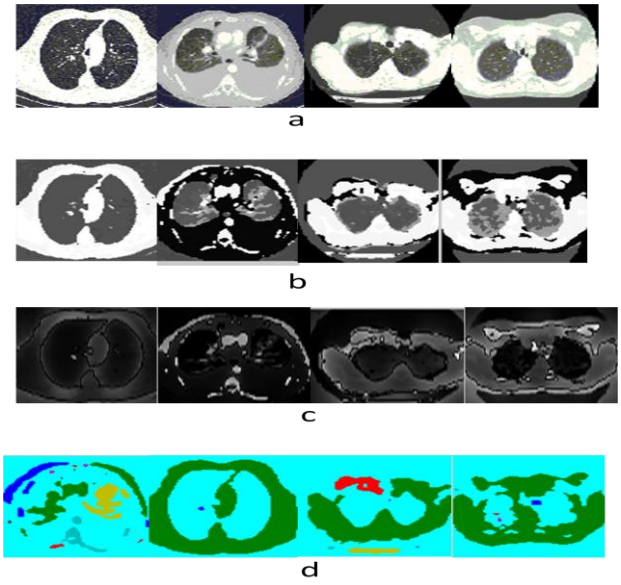

Figure 2: (a) Input image (b) Images after medium filter, (c) Images after application of Mad Technique with Gabor, (d) Images after application of Genetic Algorithm

3.3 Feature selection

Feature selection is the strategy of choosing a subset of the phrases occurring in the training set. It is the method of choosing a subset of appropriate attributes (variables, predictors) to be used in model development. This approach is entirely totally different from dimensionality reduction like it removes all the features whose variation doesn't meet certain threshold. The selection methods are ordinarily applied to extensions containing diverse features and also relatively variety of free samples. Feature selection is furthermore called variable selection or attributes selection. Genetic algorithm (GA) is an investigation heuristic that mimics the method of genuine selection. This heuristic (also sometimes called a meta heuristic) is regularly useful to generate useful methods to optimization and search problems .Genetic algorithms are part of the bigger class of evolutionary algorithms (EA), which produce solutions for optimization problems using techniques encouraged by natural evolution, such as inheritance, mutation, selection, and crossover.

3.4 Classification

Classification is a basic method associated with categorization, the method in which suggestions as well as items are identified, differentiated, and recognized. Classification is a data mining (machine learning) process accustomed to forecast cluster membership for data instances. The algorithms are going to discover relationships between the attributes to ensure you can also forecast the result. In this paper classification is done using MLP and the proposed Improved MLP-NN algorithm. These algorithms are used for classification and compared for the detailed analysis in this paper further.

3.4.1 MLP Algorithm

Multilayer perceptron classifier (MLPC) is a classifier depending on the feedforward artificial neural network. MLP involves numerous layers of nodes. Each and every layer is completely linked to the upcoming layer in the network. Nodes in the input layer represent the input data. All other nodes maps inputs to the outputs by performing linear composition of the inputs with the node's weights w and bias b and applying activation function. Multilayer perceptron (MLP) is an information processing technique based on biological nervous systems process information, such as in brain.

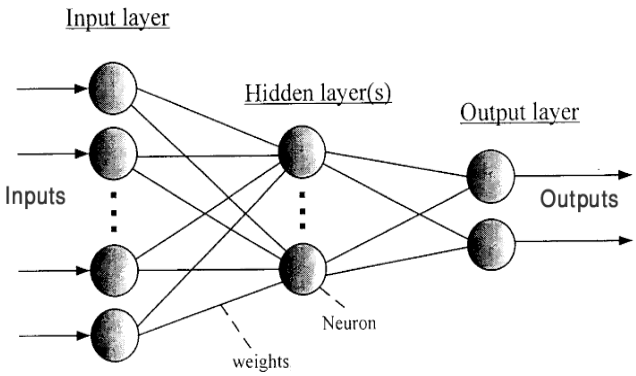

Figure 3: Block diagram of MLP

3.4.2 Improved MLP-NN

The MLP-NN makes initial group on the randomized approach which needs to eliminate unused information. So, the Improved MLP-NN will firstly create the group based on the nearest neighbor. The nearest neighbor is finding according to distance formula. When grouping is done then initialize all weights w= x1, x2, x3…xn according to input. After initialize weights then calculate the weights with their activation function. Then find activation of all hidden layer. The processing is done by all neuron of the network step by step. With the groups initialization based nearest neighbor the complexity of MLP-NN can be reduced. When unwanted information is removed then time complexity is automatically reduced. Improved Multilayer Perception (IMLP-NN) has a Hidden layer between the input

layer and output layer. Improved MLP-NN is implementing by following step:

Step1: Make a grouping based on nearest neighbor which is selected by GA.

Step2: Initialize input of all weights in the network.

Step3: Calculate the weight sum of the each and every input.

Step4: Calculate activation function of all layers.

Step5: Output of all layers.

4. IMPLEMENTATION & RESULTS

Our proposed method is implemented on lung diseases CT dataset. There are two dataset with 180 images. The first data set consider 80 images, 20 images of each disease. The second dataset consist of 100 images, 25 of each disease. The 25 images are normal lung image and each diseases of bronchitis, emphysema, pleural effusion images. The Improved MLP-NN is used for classify diseases. In Improved MLP-NN optimized value of threshold variable and training time. To evaluate the results of the MLP and Improved MLP-NN for the lung diseases image parameters used are i.e. accuracy, F-measure, precision, recall and correctly classified diseases.

4.1 Result of MLP an Improved MLP-NN with 80 Images Dataset

The output of two classification algorithms with 80 images dataset the classification accuracy is show in Fig. 4. The classification accuracy is tabulated form in Table 1.

Table 1: Classification Accuracy with 80 images dataset

Algorithm	Accuracy (%)
MLP	74.6
Improved MLP-NN	78.5

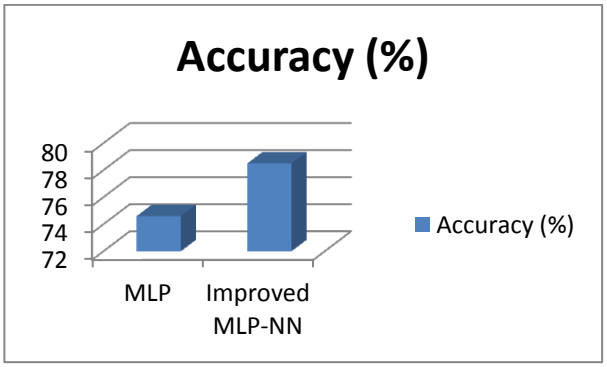

Figure 4: Classification accuracy (in percentage) with 80 images dataset

The performance measures such as precision, recall, F-Measure are presented in the Table 2 and graphical representation in the Fig. 5.

Table 2: Performance measure of two classifiers with 80 images dataset

Algorithm	Precision	Recall	F-measure
MLP	.89	.85	.83
Improved MLP-NN	.93	.92	.91

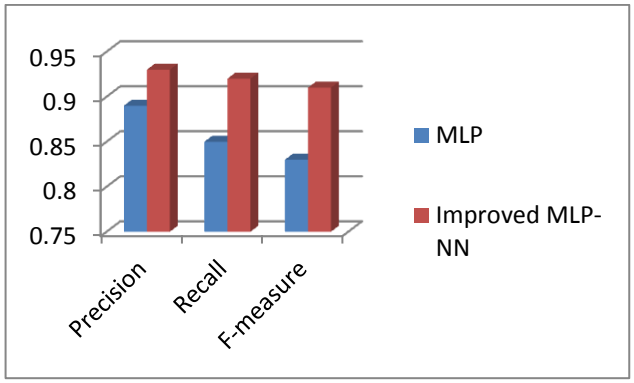

Figure 5: Performance measurements with 80 images dataset

4.2 4.2 Result of MLP and Improved MLP-NN with 100 images dataset

The output of two classification algorithms with 100 images dataset the classification accuracy is show in Fig. 6. The classification accuracy is tabulated form in Table 3.

The performance measures of 100 images dataset such as precision, recall, F-Measure are presented in the Table 5 and graphical representation in the Fig. 7.

Table 3 Classification Accuracy with 100 images dataset

Algorithm	Accuracy (%)
MLP	73.8
Improved MLP-NN	78.7

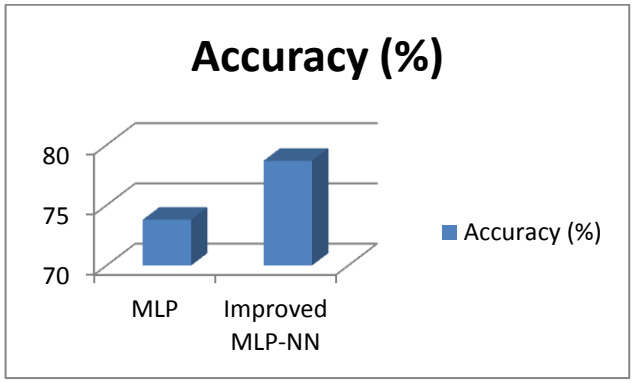

Figure 6: Classification accuracy (in percentage) with 100 images dataset

Table 5: Performance measure of two classifiers with 100 images dataset

Algorithm	Precision	Recall	F-measure
MLP	.88	.83	.82
Improved MLP-NN	.94	.92	.92

Table 6 presents the classification of 100 images datasets by their respective diseases for a given set of data.

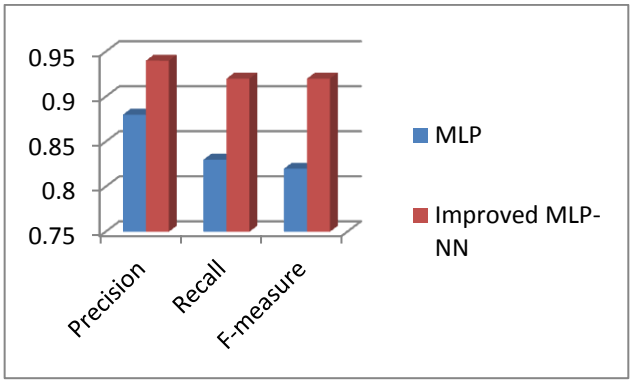

Figure 7: Performance measurements with 100 images dataset

Table 6: Classification details disease wise 100 images dataset

Algorithm	Correctly classified all diseases out of 80	Correctly classified all diseases out of 100
MLP	70	88
Improved MLP-NN	73	93

Correctly classification of all diseases using MLP and Improved MLP-NN on two data set i.e. 80 images and 100 images data set.

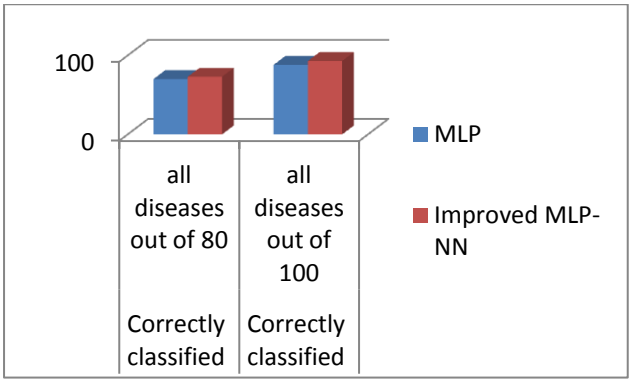

Figure 8: Correctly classification of all disease (in %age)

5. CONCLUSION

In this work feature extraction is performed by MAD with Gabor and feature selection is performed by genetic algorithm that chooses the best solutions. For classification, MLP and New proposed Improved MLP-NN are used. The proposed approach work for the detection of lung diseases with high performance as shown in the results. With large number of images(i.e.100 Images in database), both approaches performs better for detection the correct disease i.e. old MLP gives 88% and new MLP-NN gives 93% in 100 images dataset. The proposed approach performs better training as compare to traditional MLP approach. The work can further be extended by including more feature extraction or/and feature selection methods for classifying more lung disease categories like asthma, lung cancer, pleurisy and influenza.

6. REFERENCES

[1] J. Han and M. Kamber, (2000) "Data Mining: Concepts and Techniques," Morgan Kaufmann.

[2] Delveen Luqman Abd AL-Nabi, Shereen Shukri Ahmed "Survey on Classification Algorithms for Data Mining:(Comparison and Evaluation)" Computer Engineering and Intelligent Systems Vol.4, No.8, 2013.

[3] V.Vaithiyanathan, K. Rajeswari, Kapil Tajane, Rahul Pitale "comparison of different classification technique using different datasets" International Journal of Advances in Engineering & Technology, May 2013. ©IJAET.

[4] Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques, Elsevier 2006, ISBN1558609016.

[5] M. Kantardzic, Data Mining - Concepts, Models, Methods, and Algorithms, IEEE Press, Wiley-Interscience, 2003, ISBN 0-471-22852-4.

[6] Lin-Yu Tseng and Li-Chin Huang. "An adaptive thresholding method for automatic lung segmentation in CT images." In AFRICON, 2009. AFRICON'09. pp. 1-5. IEEE, 2009.

[7] Anita Chaudhary and Sonit Sukhraj Singh. "Lung cancer detection on CT images by using image processing." In Computing Sciences (ICCS), 2012 International Conference on, pp. 142-146. IEEE, 2012.

[8] Shojaii, Rushin, Javad Alirezaie, and Paul Babyn. "Automatic lung segmentation in CT images using watershed transform." In Image Processing, 2005. ICIP 2005. IEEE International Conference on, vol. 2, pp. II-1270. IEEE, 2005.

[9] K.Devaki and V. MuraliBhaskaran. "Study of computed tomography images of the lungs: A survey." In Recent Trends in Information Technology (ICRTIT), 2011 International Conference On, pp. 837-842. IEEE, 2011.

[10] Nihad Mesanovic, Svjetlana Mujagic, Haris Huseinagic, and Samir Kamenjakovic. "Application of lung segmentation algorithm to disease quantification from CT images." In System Engineering and Technology (ICSET), 2012 International Conference on, pp. 1-7. IEEE, 2012.

[11] Yiming Qian, and Weng Guirong. "Lung nodule segmentation using EM algorithm." In Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2014 Sixth International Conference on, vol. 1, pp. 20-23. IEEE, 2014.

[12] Qixin Gao, , ShengJun Wang, Dazhe Zhao, and Jiren Liu. "Accurate lung segmentation for X-ray CT images." In Natural Computation, 2007. ICNC 2007. Third International Conference on, vol. 2, pp. 275-279. IEEE, 2007.

[13] Yoshinori Itai, Hyoungseop Kim, Seiji Ishikawa, Shigehiko Katsuragawa, Takayuki Ishida, Katsumi Nakamura, and Akiyoshi Yamamoto. "Automatic segmentation of lung areas based on SNAKES and extraction of abnormal areas." In Tools with Artificial Intelligence, 2005. ICTAI 05. 17th IEEE International Conference on, pp. 5-pp. IEEE, 2005.

[14] XindongWu· Vipin Kumar · J. Ross Quinlan · Joydeep Ghosh · Qiang Yang ·Hiroshi Motoda · Geoffrey J. McLachlan · Angus Ng · Bing Liu · Philip S. Yu ·Zhi-Hua Zhou · Michael Steinbach · David J. Hand · Dan Steinberg "Top 10 algorithms in data mining" Springer-Verlag London Limited 2007, 4 December 2007.

[15] Aaditya Desai And Dr. Sunil Rai "Analysis of Machine Learning Algorithms using WEKA" International Conference & Workshop on Recent Trends in Technology, (TCET) 2012 Proceedings published in International Journal of Computer Applications® (IJCA).

[16] Misra, Avishkar, Mamatha Rudrapatna, and Arcot Sowmya. "Automatic lung segmentation: a comparison of anatomical and machine learning approaches." Ins *Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004. Proceedings of the 2004*, pp. 451-456. IEEE, 2004

[17] Dharminder Kumar and Suman "Performance Analysis of Various Data Mining Algorithms: A Review" International Journal of Computer Applications (0975 – 8887) Volume 32– No.6, October 2011.

[18] Niranjan J. Chatap and Ashish Kr. Shrivastava "A Survey on Various Classification Techniques for Medical Image Data" International Journal of Computer Applications, Volume 97– No.15, July 2014.

[19] G.Kesavaraj And Dr.S.Sukumaran "A Comparison Study on Performance Analysis of Data Mining Algorithms in Classification of Local Area News Dataset using WEKA Tool" International Journal Of Engineering Sciences &Research Technology 2(10),October 2013.

[20] M. S. Chen, J. Han, and P. Yu, 1996. Data mining: an overview from a database perspective. IEEE Transactions on Knowledge and Data Engineering, vol. 8, no. 6, pp. 866-883.

[21] Sang Jun Lee, Keng Siau "A review of data mining techniques" Industrial Management and Data Systems, University of Nebraska-Lincoln Press, USA, pp 41-46, 2001.

[22] Hye Suk Kim, Hyo-sun Yoon, Kien Nguyen Trung, and Guee Sang Lee. "Automatic lung segmentation in Ct images using anisotropic diffusion and morphology operation." In Computer and Information Technology, 2007. CIT 2007. 7th IEEE International Conference on, pp.557-561.IEEE, 2007.

[23] S. W. Purnami, A. Embong, J. M. Zain and S. P. Rahayu, "A New Smooth Support Vector Machine and Its Applications in Diabetes Disease Diagnosis," *Journal of Computer Science*, Vol. 5, No. 12, pp. 1006-1011.

[24] Ştefan, Raluca-Mariana. "A Comparison of Data Classification Methods." *Procedia Economics and Finance* 3 (2012): 420-425.

[25] Salama, G. I., Abdelhalim, M., & Zeid, M. A. E. (2012). Breast cancer diagnosis on three different datasets using multi-classifiers. *Breast Cancer (WDBC)*, *32*(569), 2.

[26] Roy, A., Dutta, D., & Choudhury, K. (2013). Training artificial neural network using particle swarm optimization algorithm. International Journal of Advanced Research in Computer Science and Software Engineering, 3(3), 430-434.