

# Lung Cancer Detection using Local Energy-Based Shape Histogram (LESH) Feature Extraction Using Adaboost Machine Learning Techniques



Sharvani, Hemant K

**Abstract:** It is difficult to find the exact symptoms of lung cancer due to the formation of the majority of cancer tissues in which the large tissue structure intersects differently. With digital images, this question can be evaluated. Images with the basic operation of the LESH Algorithm will be examined in this strategy. GLCM approach is used in this paper to pre-process the snap shots and feature extraction system and to check a patient's disease rate at its it's premature or unnatural to know it. The cancer stage will be assessed with the aid of the results . Using the data set and the cancer patient's survival rate can be calculated. The conclusion is based entirely on the accurate and incorrect arrangement of tissue patterns.

**Keywords:** Echo State Network (ESN), Clinical Decision Support Systems (CDSSs), Local Energy based Shape Histogram (LESH), Extreme Learning Machine (ELM), Echo State Network (ESN), AdaBoost, Support Vector Machine (SVM).

## I. INTRODUCTION

Lung cancer causes more than one million deaths worldwide every year. As per American Lung Cancer Society, lung cancer is the 2<sup>nd</sup> most common cancer in both men and women without counting skin cancer and accounts for nearly 14 percent of all recent cancer cases. The American Cancer Society reports about 224,390 new cases of lung cancer involving about 117,920 in men and 106,470 in women and about 158,080 deaths from lung cancer in the United States in 2016, including 85,920 in men and 72,160 in women. [34]. Lung cancer fatality rate may reduce early diagnosis and therapy. Together with chest radiology, Clinical decision support systems (CDSSs) can help doctors get an early detection of lung cancer. The field of science in emerging healthcare systems is constantly changing, and so the world of computing is shifting in parallel. Modern medicine or research in medical field this goes hand in hand with the art computational paradigms. Our proposed CDSS framework has significant potential for leading to enhanced patient treatment by diligent monitoring of cases of lung cancer. The chest radiograph dataset were selected for experimental purposes, which is part of a primary medical imaging procedure applied for pulmonary disease evaluation.

Gomathi et al. in [1] specific techniques of image processing such as media filter, erosion and dilation, and outlining for detecting the lung area which was further segmented using technique of Fuzzy Possibilistic C Means (FPCM). Extreme Learning Machine (ELM) was subsequently used to detect false positive nodules from segmented nodules from CT scans. Tong et al. in [2] proposed Computer-assisted detection scheme (CAD) based on segmentation of adaptive threshold, morphological mathematics, Gaussian filter and Hessian matrix algorithms for true positive CT nodules detection. Bhuvanewari and Therese in [3] Proposed Genetic Algorithm Method in combination with K-Nearest Neighbors (K-NN) algorithm for identifying lung CT scanning images of cancer (with cancer/without cancer) and achieved classification accuracy of 90%. Bhuvanewari et al. in [4] performed lung cancer classification upon computed tomography (CT) images using three steps process. At the first step, images were processed using median filtering methods and morphological filtering techniques. Next, Gabor filter and Walsh Hadamard transform features were extracted and fused using the technique of median absolute deviation (MAD) and fed into classifiers such as decision tree, K-nearest neighbour (KNN) and Perceptron multilayer neural networks (MLP-NN). This method achieved a classification accuracy of 90%. Bush in [5] applied a model of a convolutional neural network (CNN) to identify chest X-rays as non-nodule, benign nodule, or malignant nodule. Nodule / non-nodule lung classification was achieved with a sensitivity of 92 percent and a specificity of 86 percent. Jaffar et al. in [6] developed a lung nodule detection system that uses computed tomography (CT) scanned images to segment nodules. For each ROI, shape features were extracted and fed to the vector support machine (SVM) to classify them as a region of interest (ROI) nodule / non-nodule. Sammouda et al [15]. Applied Hopfield neural networks (HANN) for 3D computed tomography (CT) chest segmentation for cancer diagnosis and achieved 90 percent sensitivity with 0.05 false positives per slice..

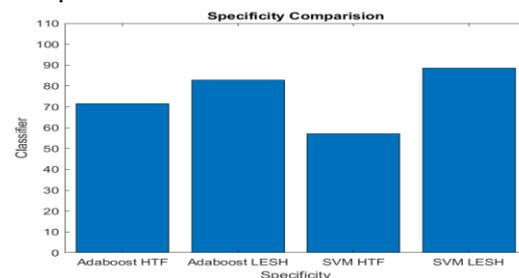


Figure 1.1: Specificity comparison between different Classifiers.

Revised Manuscript Received on January 30, 2020.

\* Correspondence Author

Sharvani\*, Student, P.G. Department of Electronics and Telecommunication, T.C.E.T, Mumbai, India.

Hemant K, Associate Professor, Department of Electronics and Telecommunication, T.C.E.T, Mumbai, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

# Lung Cancer Detection using Local Energy-Based Shape Histogram (LESH) Feature Extraction Using Adaboost Machine Learning Techniques

## II. METHODOLOGY

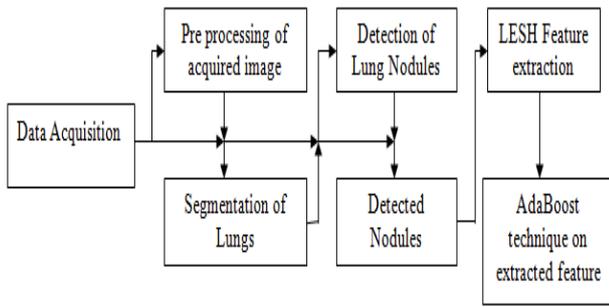


Figure 2.1: Basic methodology for feature extraction

The axial CT images are taken as a data set. The images are pre processed as per Gabor filters to remove any noise present in the image and thus only lung section is segmented out from the extracted CT image. As the lungs is composed of nodules which are the prime source for cancerous cells, the nodules are found by techniques of extraction separate from the segmented picture of lungs. The next step is to calculate the LESH features with the algorithm mentioned in the below section. Finally we use the AdaBoost machine learning technique to compare the results of combinations of LESH with AdaBoost and other feature extraction techniques.

## III. ALGORITHM FOR LESH FEATURE EXTRACTION

Algorithm:  $I$  is the CT image of the chest at location  $z = (x, y)$ .

Begin:

1. First convolute image  $I$  with 2D log-Gabor filter bank  $G(\omega)$  with different orientations  $\theta$  and scales  $s$ .

$$G(\omega) = \exp\left\{-\frac{(\log(\frac{\omega}{\omega_0}))^2}{2(\log(\frac{k}{\omega_0}))^2}\right\}$$

The resulting response vector for the convolution is determined using equation 1.

2. Using equation 2 to determine the response amplitude.
3. Use equation 3 to calculate the measure of sensitive phase deviation.
4. Calculate local energy with the help of equation 4.
5. Next, using equation 5 calculate 2-D phase congruence for the image.
6. Use equations 7 and 8, measure the LESH feature vector.

END;

### 3.1 LESH Base Feature Extraction Technique

The technique of LESH function extraction is primarily principle of determining the pattern of histogram for local energy of object of interest. Morrone et al measure local energies according to different orientations using a phase congruence process[13]. The picture is transformed with a 2D log-Gabor filter bank with different orientations and

scale and then phase congruency (PC) is calculated. The log-Gabor transfer function is stated as:

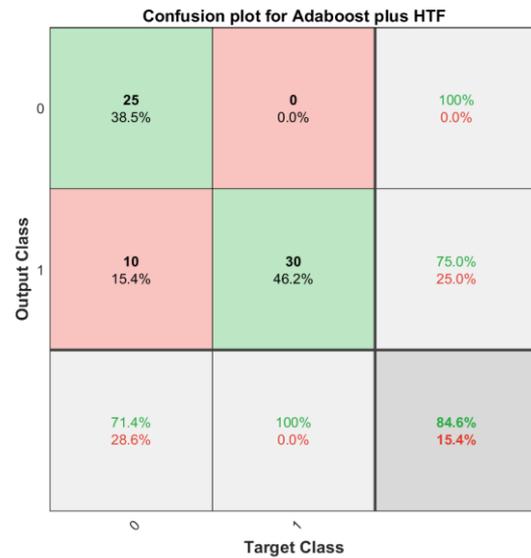


Figure 3.1: Confusion plot for HTF feature extraction using AdaBoost Classifier

Let  $G_{even}$  and  $G_{odd}$  be the even-symmetric and odd-symmetric filters at scale  $s$  and orientation  $\theta$ . Then convolution with image  $I$  results in the response vector given as:

$$[e_{so}(z), o_{so}(z)] = [I(z) * G_{even}, I(z) * G_{odd}] \quad (1)$$

where  $z = (x, y)$  represents a position of an image. As a result, the amplitude and orientation of the response can be determined as:

$$A_{so} = \sqrt{(e_{so}(z))^2 + (o_{so}(z))^2} \quad (2)$$

And measurement of sensitive phase deviation is provided by equation:

$$\Delta\phi_m(z) = \cos(\phi_m(z) - \overline{\phi_m(z)}) - |\sin(\phi_m(z) - \overline{\phi_m(z)})| \quad (3)$$

Finally, local energy is calculated using equation given as:

$$E(z) = \sqrt{(\sum_m e_{so}(z))^2 + (\sum_m o_{so}(z))^2} \quad (4)$$

At last, 2-D image phase congruency sum of Fourier amplitude components is calculated as local energy standardized as:

$$PC(z) = \frac{E(z)}{\sum_m A_{so} + \epsilon} \quad (5)$$

Further detail for LESH feature extraction can be found in our paper. Such filters are designed to detect features in all directions. For each orientation, the number of energies is determined by the total sum and amplitude scales. The resulting vector of the LESH function is determined as follows:

$$h_{r,b} = \sum W_r \times PC(z) \times \delta_{r-b} \quad (6)$$

$$W_r = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{[(x-r_{xo})^2+(y-r_{yo})^2]}{\sigma^2}} \quad (7)$$

Here in a region  $r$  of an image,  $W_r$  represents the Gaussian weighting function,  $\delta_{r-b}$  with the orientation label map  $L$  is the Kronecker's delta having current bin  $b$ .  $PC$  is the local energy by the equation.

#### IV. RESULTS & DISCUSSIONS

Using LESH with AdaBoost, we get better results compared to current HTF using SVM or LESH using SVM. You can see the sensitivity, specificity and accuracy plot as well. In the confusion plot we get 89.2% correct which means cancerous cells are correctly detected and only 10.8% false results are obtained. The ROC Plot of AdaBoost plus LESH shows that it is near to 1, but most importantly since it is covering more area under the curve, it indicates that compared to others it has better results. The outcomes were analysed using accuracy, sensitivity and specificity of the classification. MATLAB tool was used to conduct the experiment. We used the version developed by Herbert Jaeger et al to apply ESN [15]. We used the original G-B template for ELM. Huang [20].

Once the classifier is trained using 10-fold cross validation, analysis and testing of results using the performance measures listed in the section described below. Various classifiers and their quality are compared with respect to different subsets.

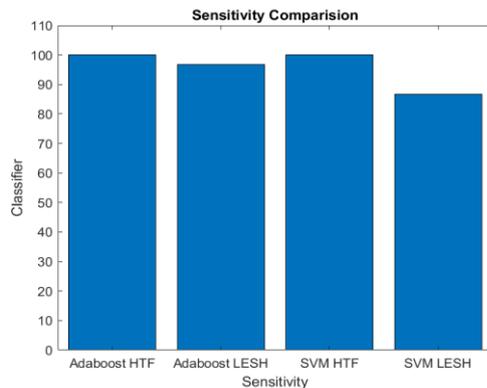


Figure 4.1: Sensitivity comparison between different Classifiers

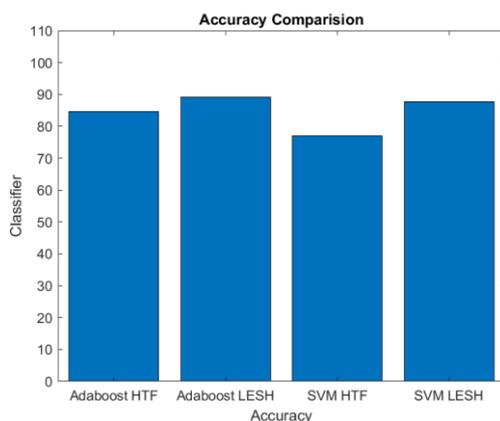


Figure 4.2: Comparison of accuracy between different classifiers

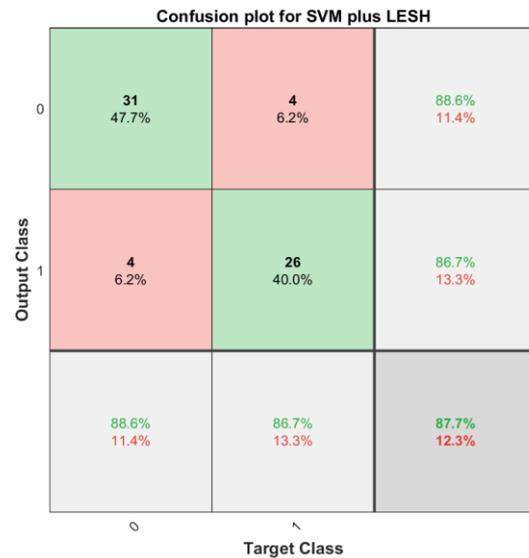


Figure 4.3: Confusion plot for LESH feature extraction technique with SVM Classifier

A comparative performance analysis of LESH, wavelet and SVM classifiers using t-test. The LESH feature extraction technique quality was compared along with Cristiane et al's [18] suggested benchmarked wavelet features extraction. They used Daubechies wavelets to transform the image into 4 levels as prescribed by S Mallat by decomposing the image first [22].

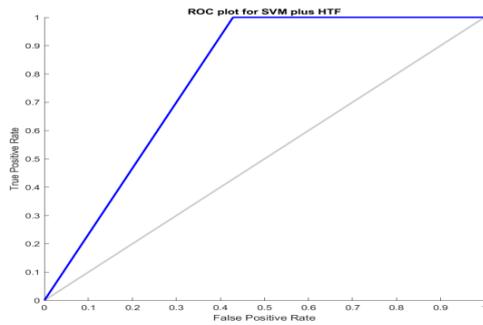
Low and high pass filter convolution helped us to achieve the decomposition of the image. Later low frequency coefficients were subsequently considered as features and were fed to the classifier as they display a better ability to distinguish different patterns of abnormalities. We also chose the largest 100 wavelet features to replicate those findings as presented in [30].

Cristiane et al. worked with a variety of selected wavelet coefficients of high-degree and categorized them using SVM. Through comparing types of the resulting classification accuracy measured using the LESH feature extraction technique (100 selected features) and the wavelet feature extraction technique (100 selected features) at a meaning rate of 0.05, the value of the LESH-based feature extraction technique was evaluated.

#### V. DISCUSSION

Due to inhomogeneous lung areas, lung cancer detection systems face multiple challenges; similarities in the thickness of bronchi, ribs, veins, bronchioles and arteries in the lung region; and various shapes of nodule such as cavities nodules and ground glass nodules. Because of the presence of all these organs, feature extraction and classification of nodule segments is a tedious task. Nodule features extraction such as form, volume, texture, and others, can help in the prediction of malignancy. Certain variables that may contribute to system performance to detect lung anomalies include: parameters for image acquisition and reconstruction, nodule position, data set size, device optimization through cross-validation.

# Lung Cancer Detection using Local Energy-Based Shape Histogram (LESH) Feature Extraction Using Adaboost Machine Learning Techniques

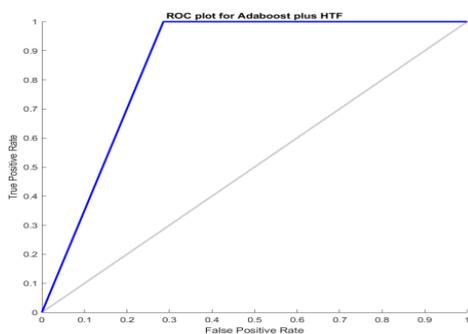


**Figure 5.1: ROC Plot for HTF Feature technique using SVM Classifier**

Most state-of - the-art methods used by researchers to extract features from ROI lung nodules do not represent discontinuities along curves and edges and are therefore unable to present a resilient set of feature vectors that may help to classify various types of abnormalities.

On the other hand, our proposed LESH extraction technique is on the basis of calculating local energy histogram using phase congruence, thus this preserves the important parameter of change in image intensity data. It is therefore able to mark significant variations in the pattern of medical images. The higher LESH coefficients refer to the most critical set of characteristics, when selected leads to nearly the same accuracy classification while reducing the dimensionality of the curve. We conducted experiments along various subgroups of the highest degree LESH coefficients and determined that in order to improve the performance of the classification,  $N = 100$  is the most appropriate number in both lung classification with and without nodules and distinguishing between malignant and benign nodules.

LESH outperforms wavelet technique compared to state-of - the-art wavelet extraction technique, albeit with a small margin. Because successful detection of malignancies is crucial in the diagnosis of cancer — to avoid unnecessary surgery — these findings are seen as an improvement..



**Figure 5.2: ROC Plot for HTF Feature using Adaboost Classifier**

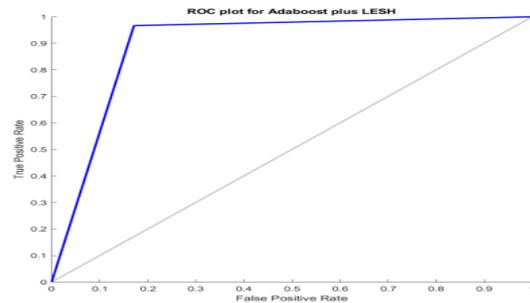
## 5.1 A multi-approach segmentation

Many methods of segmentation are proposed in the literature. The problem is data loss. Each method generates a separate segmentation image. In our proposed paper we apply several methods of segmentation and get a single segmented image with advantages of fusion theory to

combine their results. This helps discover new knowledge not discovered when applying only one method.

## 5.2 Selection of features

The role of feature extraction is important in enhancing classification performance while reducing the curve of dimensionality.



**Figure 5.3: ROC Plot for LESH Feature technique using AdaBoost Classifier indicates greater area under the curve**

## VI. CONCLUSIONS AND FUTURE WORK

The CDSS presented in our paper outperforms other state-of - the-art methods provided in the literature, since it achieves the highest degree of accuracy of 100% when classification with the polynomial SVM kernel is performed. The results preliminary presented in our paper should be considered carefully since they show the ability to use small data sets to detect LESH characteristics for lung cancer. To order to further determine the quality of the proposed method, further comparative analysis of large-scale medical datasets now needs to be carried out.

In the future, we aim to evaluate our proposed system. In the future, we plan to test our proposed system using additional clinical datasets that are benchmarked against other state-of - the-art feature selection and classification approaches recently proposed, namely Arbitrary Standard SVM and sparsely connected SVM by Huang et al. and Multi-Layered Echo State Machine by Malik et al. We have already extended our proposed framework for the detection of cancer in three-dimensional medical images, resulting in a research paper on breast cancer detection through magnetic resonance imaging (MRI) analysis. In addition, it is possible to evaluate a hybrid approach of combining various techniques for state-of - the-art extraction with LESH features while selecting significant features to enhance classification efficiency.

Eventually, we intend to use sensitivity analysis (SA) in order to identify LESH characteristics which have less impact on the classification results. SA makes these feature evaluation based on the effect of these features on system performance.

## REFERENCES

1. M. Gomathi and Dr.P.Thangaraj, "A Computer Aided Diagnosis System For Detection Of Lung Cancer Nodules Using Extreme Learning Machine", International Journal of Engineering Science and Technology, vol. 2, issue. 10, 2010.
2. J. Tong, Z. Da-Zhe, W. Ying, Z. Xin-Hua, and W. Xu, "Computer-Aided Lung Nodule Detection Based On CT Images", IEEE/ICME International Conference on Complex Medical Engineering, 2007.

3. P. Bhuvaneshwari and A. B. Therese, "Detection of Cancer in Lung with K-NN Classification Using Genetic Algorithm," *Procedia Materials Science*, vol. 10, pp. 433-440, 2015.
4. C. Bhuvaneshwari, P. Aruna, and D. Loganathan, "A new fusion model for classification of the lung diseases using genetic algorithm," *Egyptian Informatics Journal*, vol. 15, Issue 2, July 2014, pp. 69-77.
5. I. Bush, "Lung Nodule Detection and Classification," [http://cs231n.stanford.edu/reports2016/313\\_Report.pdf](http://cs231n.stanford.edu/reports2016/313_Report.pdf).
6. M. A. Jaffar, A. Hussain, F. Jabeen, M. Nazir, and A. M. Mirza, "GA-SVM Based Lungs Nodule Detection and Classification," *Signal Processing, Image Processing and Pattern Recognition*. Springer Berlin Heidelberg, 2009: pp. 133-140.
7. K. Suzuki, "Pixel-based Machine Learning in Computer-Aided Diagnosis of Lung and Colon Cancer," *Machine Learning in Healthcare Informatics*, vol. 56 (of the series Intelligent Systems Reference Library), pp 81-112, 10 December 2013.
8. R. Sammouda, J. Abu Hassan, M. Sammouda, and A. Al-ZuhairatemabouElAbbas, "Computer Aided Diagnosis System for Early detection of Lung Cancer Using Chest Computer Tomography Images," *GVIP O5 Conference*, 19-20 December 2005, CICC, Cairo, Egypt.
9. L.P Wong and H.T Ewe, "A Study of Lung Cancer Detection using Chest X-ray Images.," 3rd APT Telemedicine Workshop, 27-28th Jan., 2005, Kuala Lumpur, Malaysia, pp. 210-214.
10. Bastos R. Ferreira, and D. Leandro Borges, "Analysis of mammogram classification using a wavelet transform decomposition", *Pattern Recognition Letters*. vol. 24, pp.973-982, 2003.
11. S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation.," *IEEE Trans. Pattern Anal. Machine Intelligence*, vol. 11 (7), pp.674-693, 1989.
12. American Cancer Society. *Cancer Facts & Figures 2016*. Atlanta, Ga: American Cancer Society; 2016.
13. K. Huang, D. Zheng, J. Sun, Yoshinobu Hotta, Katsuhito Fujimoto and Satoshi Naoi, "Sparse learning for support vector classification," *Pattern Recognition Letters*, vol. 31, no. 13, pp. 1944-1951, Oct., 2010.
14. K. Huang, D. Zheng, I. King, and M. R. Lyu, "Arbitrary Norm Support Vector Machines. *Neural Computation*", Vol. 21, No. 2: 560-582, 2009.
15. S. K. Wajid and A. Hussain, "Local Energy-based Shape Histogram (LESH) Based Clinical Decision Support System for Breast Cancer Detection using Magnetic Resonance Imaging (MRI)", *Expert Systems with Applications*, submitted
16. Z. K. Malik, A. Hussain, and J. Wu, "Multi-Layered Echo State Machine: A novel Architecture and Algorithm for Big Data applications," *IEEE Transactions on Cybernetics*, 2016 (in press).
17. Imed Riadh Farah, Wadii Boulila, Karim Saheb Ettabaâ, Basel Solaiman, Mohamed Ben Ahmed, "Interpretation of Multisensor Remote Sensing Images: Multiapproach Fusion of Uncertain Information," *IEEE Trans. Geoscience and Remote Sensing*, vol. 46, no. 12, pp. 4142-4152, 2008.

Coordinator. Provided guidance for more than 25 projects at Undergraduate level and more than 5 projects at Post graduate level. Provided STTP on "Design & Development of application for smart city ". Attended IIT Bombay's E-yantra workshop. Strong education professional with a BE Bachelor of Engineering focused in biomedical signal processing, industrial automation from Govt. college of engineering, Amravati.

## AUTHORS PROFILE



**Sharvani Bhargava** Currently pursuing Master's degree in Electronics and TeleCommunication Engineering at Thakur Collage of Engineering and Technology. Completed Bachelor of Engineering in Biomedical from Watumull Institute of Engineering and Technology. Completed project in Hand gesture recognition for

American Signed language using Matlab. Currently having skilled knowledge in Information Technology domain with specialization in SAP Process Integration. Deep knowledge and trained in various programming languages like Java, SAP ABAP, C++, SQL and HTML. Formally worked as a system engineer at Infosys Technology Pvt. Ltd. In personal life, a diploma holder in Bharatnatyam classical dance form and hobbies are book reading and gardening.



**Hemant Kasturiwale** Experienced Associate Professor with a demonstrated history of working in the higher education industry. Skilled in Research, Matlab, Digital Electronics, Power Electronics, Industrial Electronics and Control

Biosignal Processing, Management, Leadership, and Public Speaking. Certified in Signals and Systems IIT workshop. Was reviewer for international journal IRQS-ISO 9001:2015 Quality Management System Lead auditor (Jan 2017), Degree certificate Distribution Ceremony