

# Comparative Study: Statistical Approach and Deep Learning Method for Automatic Segmentation Methods for Lung CT Image Segmentation

**Dr. Akey Sungheetha**

Data Science SIG member,  
Image Processing SIG member,  
Computer Science and Engineering,  
School of Electrical Engineering and Computing,  
Adama Science and Technology University,  
Adama, Nazret, Ethiopia.

**Dr. Rajesh Sharma R**

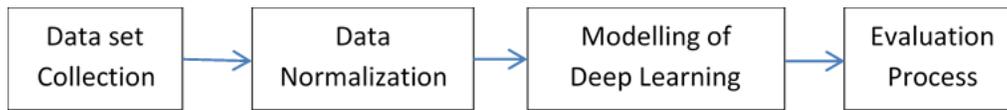
Image Processing SIG member,  
Computer Science and Engineering,  
School of Electrical Engineering and Computing,  
Adama Science and Technology University,  
Adama, Nazret, Ethiopia.

**Abstract:** Recently, deep learning technique is playing important starring role for image segmentation field in medical imaging of accurate tasks. In a critical component of diagnosis, deep learning is an organized network with homogeneous areas to provide accurate results. It is proved its superior quality with statistical model automatic segmentation methods in many critical condition environments. In this research article, we focus the improved accuracy and speed of the system process compared with conservative automatic segmentation methods. Also we compared performance metrics such as accuracy, sensitivity, specificity, precision, RMSE, Precision- Recall Curve with different algorithm in deep learning method. This comparative study covers the constructing an efficient and accurate model for Lung CT image segmentation.

**Keywords:** Image segmentation, Probability density function, deep learning

## 1. INTRODUCTION

In a recent medical research, early prediction of lung cancer is enormous research. There are many algorithms to detect the lung nodule automatically in early diagnosis. Though there are many algorithms used in the field of radiology, oncology, neurology for early prediction of lung cancer, they fails to attain good accuracy and efficiency [8]. Because of earlier systems were built on outdated algorithm such as edge detection filters with mathematical methods. The machine learning techniques are constraint with small data set and providing good results [9]. Now a day, these segmentation techniques are altered by deep learning method algorithm intensively. The deep learning contains multiple nonlinear layers for processing which consist of input layer, inner layer and output layer. These layers are interconnected with each other layers. This figure 1 is showing simple flow of deep learning method analysis in image segmentation.



**Figure 1** Deep learning approach for image segmentation

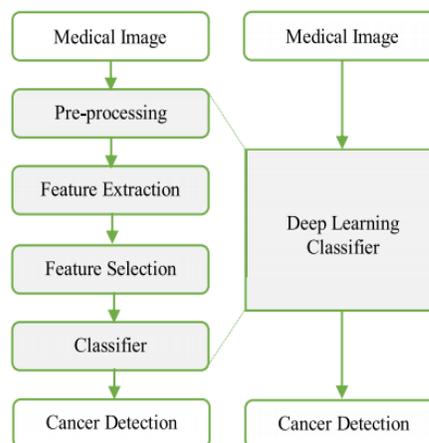
The owing to improve the accuracy of the process, deep learning approaches addresses correlation of the critical challenges during the segmentation process in a trained manner. In this process, there are three main attributes as frame the network, collect the data and trained the network [9].

## 2. RELATED WORK

Statistical analysis based automatic image segmentation methods are not feasible in current practical real time scenario in medical image analysis. Previously, lot of researches had been done about automatic image segmentation with the help of mathematical modelling, statistical analysis. It is acceptable that performance of process is lagging in efficient and accuracy with current recent trends of image segmentation. So the automatic image segmentation is directed towards machine learning methods currently.

There are many researches has done with machine learning methods for early prediction for lung nodule segmentation. In order to increase the accuracy level of image segmentation, randomly splitting Ts and Tr using machine learning methods proposed by Yu et al [1]. With the help of support vector machine algorithm, accuracy level was improved a bit due to splitting Ts and Tr. in a machine learning process. This one gave high accuracy with small amount of data set with homogenous function which was proposed by yuan et al [2]. Splitting training data (Tr) and testing data (Ts) will be improving performance of the system. Manually 70% and 30% was divided the Tr and validation data respectively in the data set proposed and implemented in machine learning algorithm by Hart et al [3].

Similarly, Hazra et al [4] contains logistics regression and implemented with 80% T and 20% Ts provided 78% accuracy in a validation. Ye et al [5] proves in his research with the help of support vector machine algorithm with 99.4% of accuracy. It is pretending high accuracy with splitting of 70% Tr and 30% Ts. Machine learning with decision tree methods performs good results in accuracy with 93.5% was proposed by gunaydin et al [6]. Here they split Tr and Ts in 70% and 30% respectively for the evaluation. Purely, statistical approach automatic segmentation can segment the nodule with percentages of error [7]. Here RMSE parameter is a measuring the error of the model. The noisy image creates individual section to deal it and process getting slower down. So this process leads to perform poorly with lot of subdivisions in the processes [10]. Subsequently, the deep learning process can improve and rectify that older automatic segmentation process in terms of speed and accuracy of the process [11]. This survey has motivated us to do this article to study the comparative analysis with state of art methods.



**Figure 2** Deep Learning algorithm for detect lung cancer

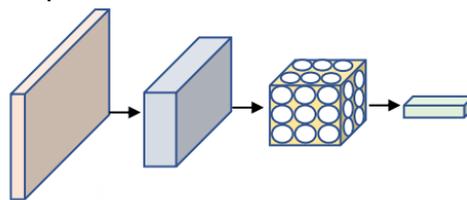
### 3. THEORETICAL ANALYSES

Segments an image by assuming pixel by pixel value distribution of each region will be comprised with the distribution function and selecting a density function according to the statistical principles and estimating the parameters [17]. Calculating the probability of malignancy is the one of the statistical approach to find the proper segmentation of images. Based on this measuring metrics, the overall system performance can be calculated. These property parameters are explained enough from the past research [16]. The following distribution are normal, uniform, beta, gamma is not such a simple problem to select a probability density function; but it will simulates the distribution of pixel values from the actual image to be segmented. Generally, the distribution of pixel values from the image will be acquired. Subsequently, computing the error between this distribution and above mentioned six common distribution factors are using least square error. The image should be divided into several regions by cumulative density function for segmenting lung nodule in the image. After apply the cumulative density function probability band to the grey scale images of lung cancer CT images, it divides pixel with same probability or same interval as one region. This pixel of same probability band will not affect the image with noise. So here we have to approach the pre-processing to eliminate the noise distribution in the CT images.

#### Overview of Deep Learning Method

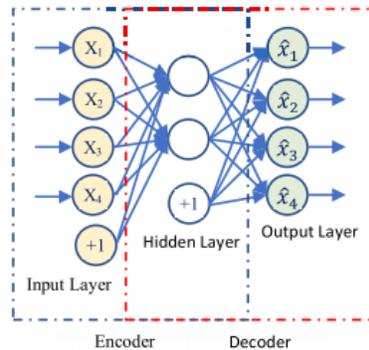
The deep learning method is used to classify the region of interest in the medical images for segment the infected region and healthy region. Generally, this deep learning method is associated with denoising and contrast filters to enhance the input images. So it can show the supremacy than other conservative methods easily. With the help of principal component analysis the features are calculated in this method. The alternative way to determine the feature is statistical analysis. Subsequently, input features are considering these determined features for the process by machine learning. The learning methods are classifying the input feature analysis vector along with class labels. After determine the model boundary in the images that splits each class for training of machine learning classifier [12].

Deep learning classifier doesn't need any type of preprocessing such as feature extraction and segmentation. But in order to improving the accuracy level in this approach, it requires image resizing due to the limit on the input values. Theoretically, this representation learning approach in the each layer transforms the input data from the previous one. So these input data's are modeled and learn both input layer and hidden layer in a hierarchical structure. The representation of the data between the layers of a deep learning process is a result of nonlinear function [13]. Generally, the first layer is used to identify the edges in the ROC of the image. This edges position will be recognized by second layer and discards the discontinuities edges if any unwanted noisy state in the same position. The third layers associates all those patterns to transform into detect the objects successfully with good accuracy level. The feature representation learning is used to lead in the many artificial intelligence approaches [14]. The construction of deep learning with CNN is shown in the picture below.



**Figure 3** Convolutional Neural Network Construction design

This research construction work contains the two main parts such as network structure, training techniques of deep learning method. Finally it is concluding the challenges by the methods. Input and output layers are used for encoder and decoder functions respectively. This process is that stack of convolution and de-convolution process of input vectors. The encoder and decoder are used to compress and reconstruct the inputs in the networks [15].



**Figure 4** Architecture of Deep Encoder-Decoder of vector inputs

### Construction of convolutional Neural Network architecture

While construction of CNN architecture, the layers are bundled of nonlinear activation layer which comprises non-linear unit layer, pooling layer, fully connected layer. In order to obtain the volume of feature maps, convolution operation is the only way between pixel of the input image and a filter. All constructed layers are performing the specific function apiece. For example, pooling layer is taking care of improve the training speed due to down sampling the input values. Moreover, Also it prevents the over fitting problem and are translation invariant problem which are common challenges in training deep models. Usually, the translation invariant is elevating due to similarity in the neighbouring pixels. Subsequently, the fully connected layer neurons linked with preceding layer neuron well. In order to classify the input values, the CNN starts classification from the centre pixel of the region. Obviously, this approach moves towards next region centre pixel even though, this method is leading to inefficient and loss of spatial information of the input image. This problem also can solve by fully convolutional network which is concentrating up sampling in the less spatial region in the maps to retaining the good resolution in the segmentation. These glitches are not elucidated in the conservative automatic segmentation methods.

### Remarks

The dataset is splitting in to two training and testing data in order to improve the accuracy and efficiency of machine learning algorithm. Also identify the parameters easily with the output linear variable.

**Consideration:** Cost and selection of Tr and Ts should be narrowed based on the network structure.

## 4. RESULTS AND DISCUSSIONS

Deep learning algorithm comprises training, validation and testing set of images; obviously very superior than older methods which was having minimum data set. The statistical approach automatic segmentation shows poor performance for large data set comparison.



**Figure 5** a & b Different approach for same segmented image by different algorithm for accuracy measurement

Also the statistical approach methods are failed to detect the lung cancer in slight reforms in the medical images as shown in above figure 5. It loses the accuracy of the model when it goes big data evaluation as shown in all the graphs. In these graphs in a figure 6 and 7, X axis are mentioned less, medium, high no of images for testing for segmentation purpose. Here we are considering 1000 images samples are in “less” measures.

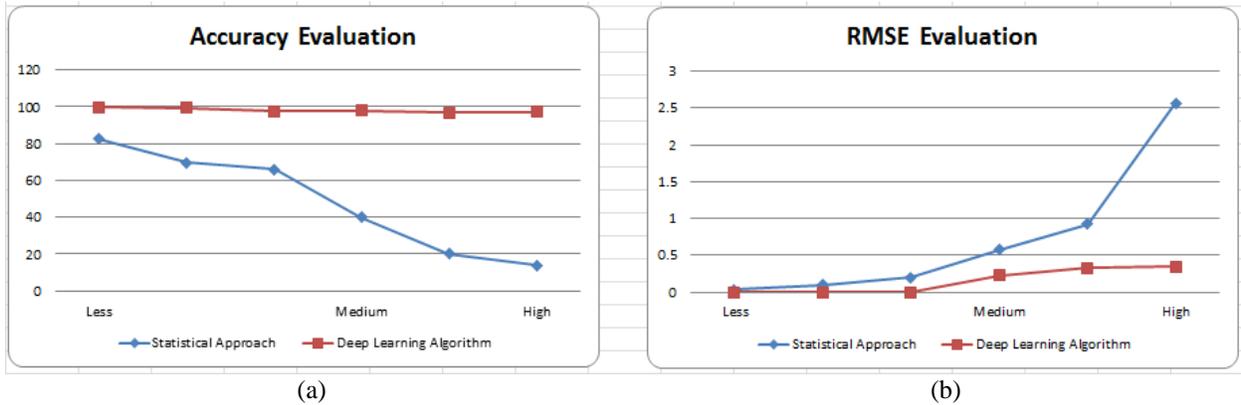


Figure 6 (a, b) Accuracy & Root Mean Square Error evaluations between two methods

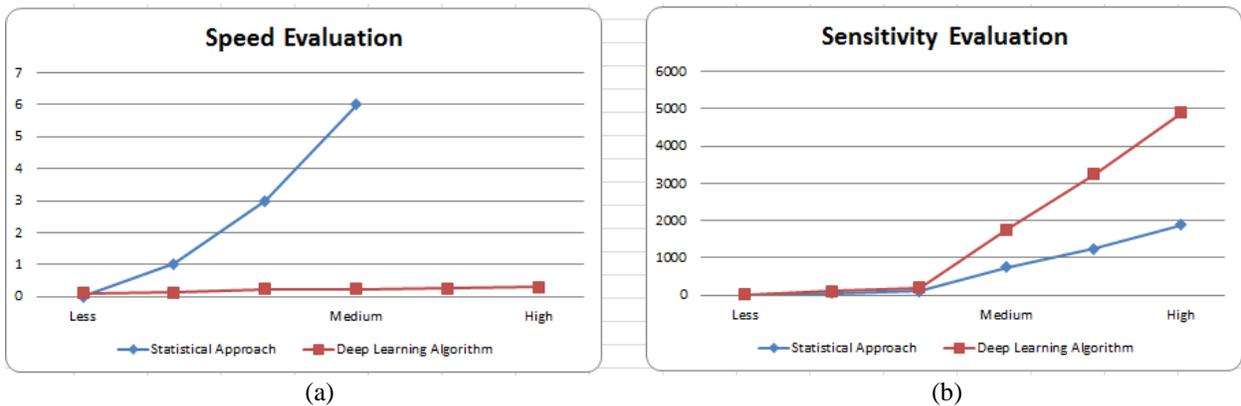


Figure 7 (a, b) Speed & Sensitivity evaluations between two methods

From these graphs, we tabulated the detailed description of the performance metrics as shown below in table 1. Some of the parameter are analysed with deep learning algorithm alone. So it is not mentioned in graphs.

S.No	Performance Measuring Metrics	Statistical approach for Automation Segmentation	Deep learning Algorithm for segmentation
1	Accuracy	Decaying	Stable
2	Sensitivity	Less	High
3	Specificity	Bad	Good
4	Precision	Low	High



5	Root Mean Square Error	High	Less
6	AUC – ROC curve	*NA	Good
7	PR Curve	*NA	Good

\*NA – Not Available

The above table 1 shows comparison between automation segmentation and deep learning method with the measuring metrics. Deep learning method shows the good amenities in all measuring metrics. Both accuracy and sensitivity tells deep learning method is superior to conservative methods.

The percentage of accuracy is 99.6% by using deep learning method compared with traditional statistical segmentation methods. This accuracy measure phenomenon is used to measure the performance of a model. Sensitivity also superior as shown in table which is indicating that how many are relevant data out of correctly classified data. Precision parameter also how many data's are correctly classified by the model out of all data's. Minimum RMSE value can state in other words of better accuracy of the model also standard way to determine the error of the model. In order to measure the multiclass classification problems, the Area Under Curve – Receiver Operating Characteristics metrics are consider in deep learning approach which is not feasible in statistical process. Finally PR curve is used for evaluate output quality of the classifier.

## 5. CONCLUSION

Our test images are 512\*512 pixels, 16-bit grey level matrices with ranges from 0.7mm and 1.2mm; in addition, minimum set of CT scan image with and without lung cancer for automatic segmentation from the Lung Image Database Consortium (LIDC) [6] which is a database for lung images headed by the US National Cancer Institute.

We compared both the technique and evaluated with different types of dataset. Also, we show the compared performance metrics to evaluate between these approaches. Clearly, we can conclude that deep learning methods are very effective algorithm for image segmentation that measured by performance metrics compared with statistical approach. Deep learning approach is applied to wide range of different medical image segmentation tasks with maximum accuracy. Different type of deep learning algorithms should be constructed and compared for image segmentation process in future. Also, we strongly believe that our study will help many researchers to identify deep learning technique in segmentation process that provides more accuracy and efficiency.

## ACKNOWLEDGEMENT

We thank Dr. Edriss babiker Hamdan, Eritrea Institute of Technology, Eritrea, North east Africa for his great support to complete this paper.

## REFERENCES

- [1] Yu L, Tao G, Zhu L, Wang G, Li Z, Ye J and Chen Q, "Prediction of pathologic stage in non-small cell lung cancer using machine learning algorithm based on CT image feature Analysis", <https://doi.org/10.1186/s12885-019-5646-9>.
- [2] Yuan F, Lu L, Zou Q, "Analysis of gene expression profiles of lung cancer subtypes with machine learning algorithms. BBA - Molecular Basis of Disease", <https://doi.org/10.1016/j.bbadis.2020.165822>.
- [3] Hart G R, Roffman D A, Decker R, Deng J, " A multi-parameterized artificial neural network for lung cancer risk prediction", <https://doi.org/10.1371/journal.pone.0205264>.
- [4] Hazra A, Bera N, Mandal A, "Predicting Lung Cancer Survivability using SVM and Logistic Regression Algorithms", International Journal of Computer Applications (0975 – 8887) Volume 174 – No.2, September 2017.
- [5] Ye Z, Sun B, Xiao Z, "Machine learning identifies 10 feature miRNAs for Lung squamous cell carcinoma", Gene. Volume 749, July 2020, 144669.



- [6] Gunaydin O, Gunay, Ozgur Sengel , “ Comparison of Lung Cancer Detection Algorithms ”, doi: 10.1109/EBBT.2019.8741826.
- [7] Bartholomai J A, Frieboes H B, “Lung Cancer Survival Prediction via Machine Learning Regression, Classification and Statistical Techniques. ISSPIT 2018”, doi: 10.1109/ISSPIT.2018.8642753.
- [8] Sumathipala Y, Shafiq M, Bongen E, Brinton C, Paik D, “Machine Learning to Predict Lung Nodule Biopsy Method Using CT Image Features,” <https://doi.org/10.1016/j.compmedimag.2018.10.006>.
- [9] Wang G, et al, Interactive medical image segmentation using deep learning with image specific fine tuning, *IEEE Trans Med Image* 2018; 37(7):1562-73.
- [10] Taha AA, Hanbury A, Metrics for evaluating 3D medical image segmentation; analysis, selection, and tool. *BMC Med Imag* 2015;15(1).
- [11] Wang G, A perspective on deep imaging, *IEEE Access* 2016;4:8914-24
- [12] Volkenandt T, Freitag S, Rauscher M, Machine Learning powered image segmentation, *Microsc Microanal* 2018;24(S1):520-1.
- [13] Havaei M, et al. Brain tumor segmentation with deep neural networks. *Med Image Anal* 2017; 35:18-31.
- [14] G.Litjens et al, “A survey on deep learning in medical image analysis” *Med. Image Anal.*,vol.42, pp. 60-88, Dec.2017
- [15] Mathieu Hatt, et al, Machine (Deep) Learning Methods for Image Processing and Radiomics, *IEEE Trans Radiation and plasma Medical sciences* 2019; 3(2): 105-109.
- [16] S. Nedeveschi, C. Pantilie, T. Marita and S. Ducea, "Statistical methods for automatic segmentation of elastographic images," 2008 4th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, 2008, pp. 287-290, doi: 10.1109/ICCP.2008.4648388.
- [17] Seon-Do Kang et al “Image Segmentation using Statistical approach via Perception-based Color Information”, published in *IJCSNS International Journal of Computer Science and Network Security*, VOL.8 No.4, April 2008.

