Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 132 (2018) 525–533

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2018)

An Innovative Approach for Investigation and Diagnosis of Lung Cancer by Utilizing Average Information Parameters

*Kale Vaishnaw G.[a], Vandana B. Malode[b]

[a] Research Scholar, Dr. Babasaheb Ambedkar University, Aurangabad, Maharashtra, Pin 431004, India

[b] Department of Electronics and Telecommunication, Jawaharlal Nehru Engineering College, Aurangabad, Maharashtra, Pin 431003, India

Abstract

In this paper, an Average Information based approach for lung cancer analysis and diagnosis has been proposed. Suggested methodology is established on average information parameters by utilizing image processing tools for lung cancer investigation. The real issue for the lung cancer diagnosis is the time constrictions for physical diagnosis that expands the death possibilities. Henceforth essentially proposed technique is an approach that would help the medical practitioners for precise and superior decision against the lung cancer discovery. Microscopic lung images are taken for analysis and investigation by using digital image processing with MATLAB. The statistical and mathematical parameters under statistical analysis are selected on the basis of the principle working of Average information technique. The input parameters like Entropy, Standard Deviation, Mean, Variance and MSE for average information method are implemented over a large microscopic lung image database. The individual statistical and mathematical parameter analysis with its impact on lung cancer images is successfully carried out and finally the accuracy, selectivity, and sensitivity of the proposed method is calculated by implementing the standard diagnostic test on the proposed method. This method also successfully rejects null hypothesis test by implementing one of the standard statistical methods.

Corresponding author: vaishnaw25@rediffmail.com

1877-0509
10.1016/j.procs.2018.05.005

Keywords: Average Information , Lung cancer, Fuzzy Logic, ANOVA, Diagnostic Test

1. Introduction

Lung cancer seems to be a very common cause of death among the people all over the world. Lungs are the essential element of respiration system of the human body. Lung malignancy reported in [1]-[2] suggests that the point, when cells inside the lungs increase in wild rate is known as lung cancer. Today lots of medical imaging techniques [1]-[3] are accessible by radiologists and medical practitioners for the diagnoses of lung cancer for example X-ray, CT, HRCT, MRI, PET and Lung Biopsy. As these radiological techniques have some advantages with some shortcomings, which doesn't provide any complete declaration around the lung cancer. One of the successful medical imaging techniques is the Lung Biopsy [4]. The Biopsies and surgical operations imaging strategies aid are exceptionally imperative in the dissection about lung malignancy and cancer. However, no test will be ideal, and no examine might diagnose lung cancer; in any case, the biopsy could do that. Yet all over again biopsy also have some drawbacks that incorporate trouble over breathing, excessive bleeding oozing out and additionally there will be a constantly chance for spreading of growing cells in the lungs and moreover different parts of the body due the removal of small part of tissue and hence considered as the last choice for the cancer diagnosis. It is often recommended, when no different scan meets expectations.

The microscopic lung picture is acknowledged for the analysis and diagnosis, is acquired through biopsy and observed under Electron Microscope [5], which is a capable magnifying instrument that permits the researchers to examine the lungs at nano gauge level. Magnifications of these depicted images could be up to 400 times or indeed additional, which is exceptionally of service to the medicinal analysis. It is extremely troublesome to visualize those microscopic and take decisions as it may go wrong in a number of cases, subsequently obliges a robust strategy. Image transforming and processing with MATLAB will be excellently useful package in taking care of the microscopic lung images. The proposed method is based on the statistical analysis due to the random behaviour and unpredictable pattern of lung cancer as a disease. The average information method using statistical and mathematical parameters is carried out for lung cancer analysis as well as diagnosis. The image classifier tool is utilized for final decision which is reported in next sections.

2. Average Information Method

The technique is based on averaging of the intensity levels for every pixel position in the image. Each scanned image has two segments, one is the signal component and the other is the noise component. In the averaging process, signal component stays unchanged, yet the noise component part varies from frame to frame. Since the noise is irregular, it tends to neutralize while, performing out the summation. At this point it is figured out that the picture signal part has an impact over the summation, when contrasted with the noise segment. In light of a similar recommendation, all the factual-statistical and mathematical parameters under this are chosen. The after effect of the parameters breaks down the harmful and ordinary lung pictures. The statistical and mathematical parameters [6]-[8] of average information method are explained as

a) *Entropy*: It indicates average information of the image. The lowest value of entropy means, no uncertainty of the image information. It is zero if the event is sure or impossible [7].

$$E = -\sum_{x}^{m}\sum_{y}^{n} P[x,y]\log P[x,y] \tag{1}$$

b) Mean: It calculates the mean of the gray levels in the image. The Mean is supposed to be high, if the sum of the gray levels of the image is high. Mean depends on the first moment of the data. The mathematical expression of mean is given as in equation (2)

$$\mu = 1/N * M\sum_{x=0}^{M}\sum_{y=0}^{N} P[x,y] \tag{2}$$

c) Variance: It explains about the distribution of gray levels over the image. The value of the Variance is expected to be high, if the gray levels of the image are spread out extensively. It explains about the probability of distribution, describing how far the value lies from the mean that is the anticipated value, which can also be defined as the moments of a distribution.

d) Standard Deviation: The value of standard deviation is assigned to the center pixel of the image. All the steps are similar to calculate the standard deviation as the variance, only the last step is the addition of the square root, hence is also known as the square root of the variance [7] represented by equation (3)

$$SD = \sigma = \sqrt{\sigma^2} \tag{3}$$

e) MSE (Mean Square Error): The MSE represents the averaging of the squares of the errors between the two images [9]. The error is the amount by which the values of the reference image differ from the test image which is calculated from equation (4)

$$MSE = \sum_{0}^{m-1}\sum_{0}^{n-1}\left\| f(i,j) - g(i,j) \right\|^2 \tag{4}$$

3. Proposed Methodology

The proposed methodology is divided into two sections; the first is pre-processing of the input image with input parameter calculation and second is lung cancer diagnosis using Fuzzy Logic. Normalized grayscaled image with a size of 255*255 is formed from the microscopically captured lung image and utilized for further processing including image enhancement and investigation practice. These images would be rightfully separated into cancer and non-cancer state of affairs. The flow diagram is shown in figure1 (a) & (b)

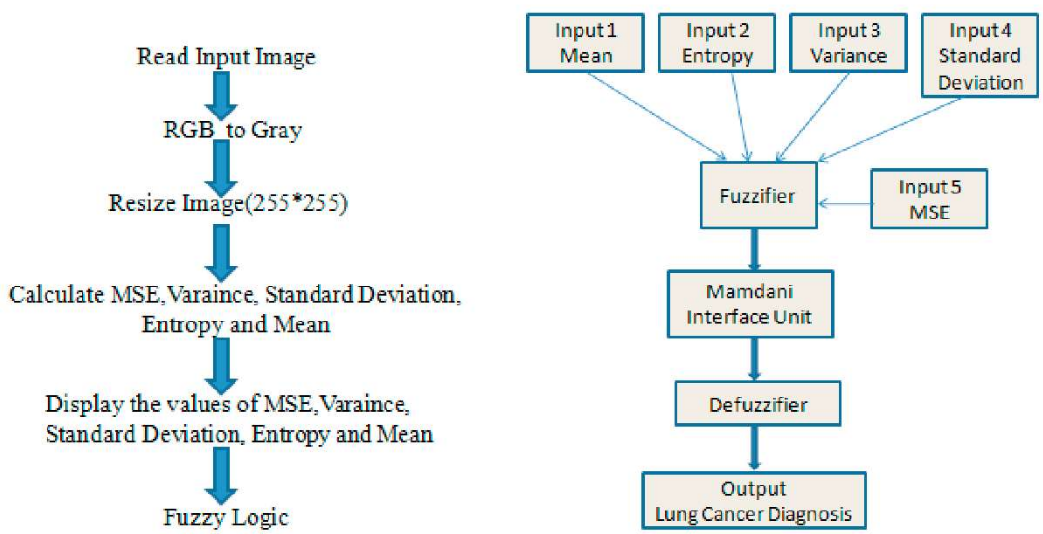


Fig.1 a) Flow Diagram Input image with Pre-processing b) Flow Diagram of Fuzzy Logic for Lung Cancer Diagnosis

Image parameters have been enhanced by adopting the Median filter, which not only removes the noise contents but also provides smoothing by removing those pixels which have shown sudden changes from the captured microscopic lung image. Average information technique is applied on these filtered-noise free images to extract the valuable statistical information of cancer lung particles and finally given to image classifier for final diagnosis.

3.1 Image Classifier

Fuzzy Logic is best used as an image classifier [10]-[11]. Fuzzy Logic is the approach utilized as an image classifier in this work for lung cancer conclusion. The different values obtained for various measurable statistical parameters of Average Information method for cancer and non-cancer microscopic lung images overlap, which settle on it hard to take a choice, whether the picture is tainted or not, henceforth Fuzzy Logic as an image classifier is utilized as a leader for lung cancer determination.

4. Results and Discussions

This session is divided into 05 sections

4.1 Calculations of Input Parameter Range

By applying Average Information statistical and mathematical parameters over 323 microscopic lung images for the number of iterations in real time, the range is calculated for each parameter. If the input parameter value of a particular image lies in that identified range, then the image is infected lung otherwise a normal lung image. In this way, each statistical parameter is applied over a particular image to get the range as observed in table1.

Table 1.Identified Statistical Parameter Range for Cancer Lung Images

Average Information Parameters	Minimum to maximum value for cancer lung	Average Range	Range for cancer lung from graph	Range for Noncancer lung from graph
Mean	97 to 123	110	215.506	218.189
Standard Deviation	50 to 71	60.50	47.752	29.3785
Mean Square Error	55 to 262	108.50	228.958	245.856
Variance	$2.25*10^3$ to $4.84*10^3$	$3.54*10^3$	$2.25*10^3$	$8.17*10^2$
Entropy	7.43 to 7.965	7.697	6.174	6.481

4.2 Statistical Response of Individual Parameter

Now the most important thing is to analyze each statistical and mathematical parameter by implementing them on both the types of microscopic lung images. At this point, when these computations are completed by applying the algorithm over a high image database, the range of all the parameters under Average Information strategy is acquired. This identified range is the distinguishing factor for an image to be cancerous or non-cancer lung image. In spite of the fact that it is very difficult as it looks since some of the statistical and mathematical parameter values overlap and is by all accounts is similar for both the types of lung images, subsequently this disarray is overcome by the Fuzzy Logic, which prepares and test the images with the input parameters for the number of iterations in real time. Table 2 and Table 3 shows the calculations of some identified input parameter values for the particular of cancerous and Noncancerous microscopic lung images.

Table 2.Input Parameter values for the Particular of Cancer Lung Images

Average Information Parameters	Cancer Image1	Cancer Image2	Cancer Image3	Cancer Image4	Cancer Image5
Mean	186.783	132.90	159.83	169.28	140.52
Standard Deviation	32.68	48.554	52.813	51.21	43.69
Mean Square Error	220.78	118.89	162.88	179.72	136.92
Variance	$1.024 *10^3$	$2.308 *10^3$	$2.426 *10^3$	$2.465 *10^3$	$1.841 *10^3$
Entropy	5.96	7.51	7.56	7.48	7.47

Table 3 Input Parameter values for the Particular of Noncancer Lung Images

Average Information Parameters	Noncancer Image1	Noncancer Image 2	Noncancer Image 3	Noncancer Image 4	Noncancer Image 5
Mean	187.36	222.52	217.89	158.44	186.78
Standard Deviation	40.08	30.05	36.75	40.03	45.91
Mean Square Error	235.92	244.05	176.21	228.7	199.01
Variance	$1.567*10^3$	866.20	$1.121*10^3$	$1.540*10^3$	$2.072*10^3$
Entropy	6.12	6.39	7.00	6.24	7.28

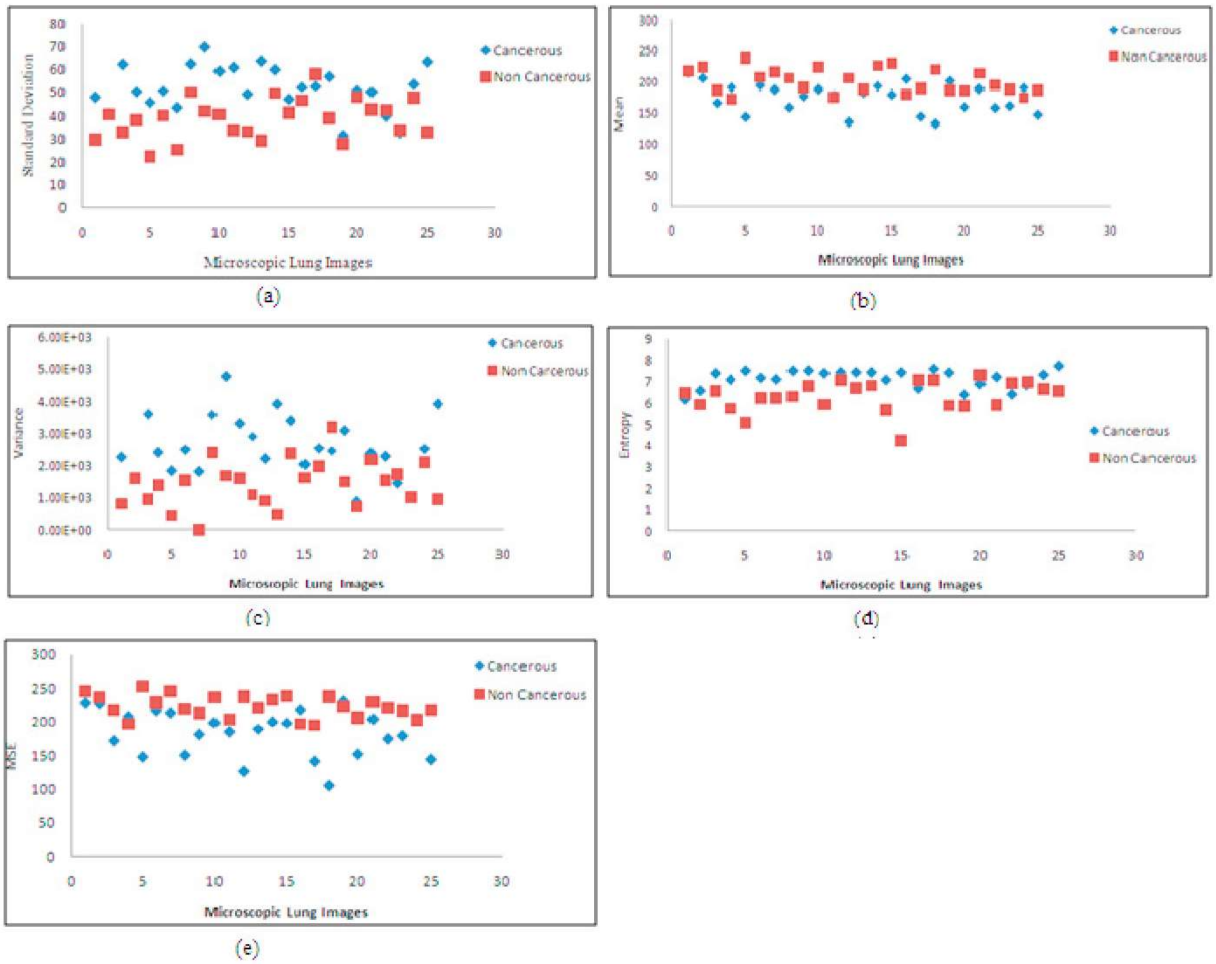
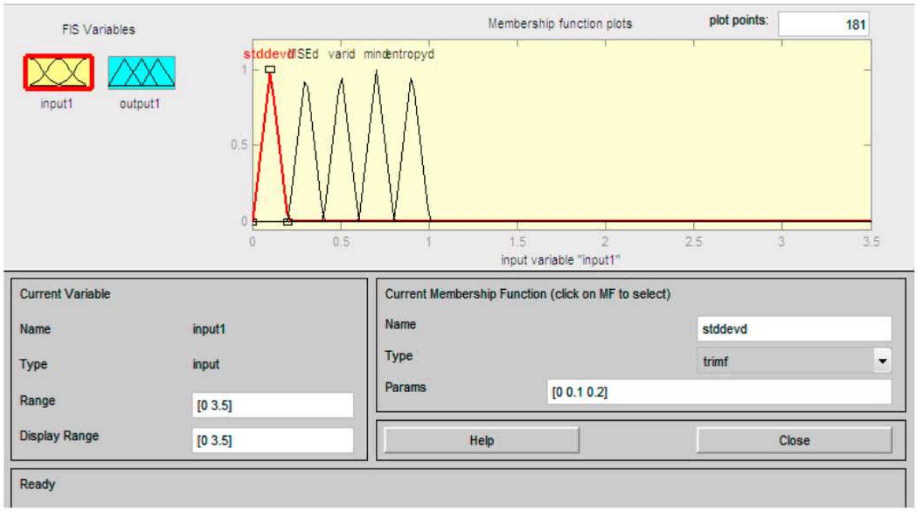


Fig. 2Statistical Response of (a)Standard Deviation,(b) Mean,(c) Variance,(d)Entropy and (e) MSE for Cancer and Noncancer Lung Images

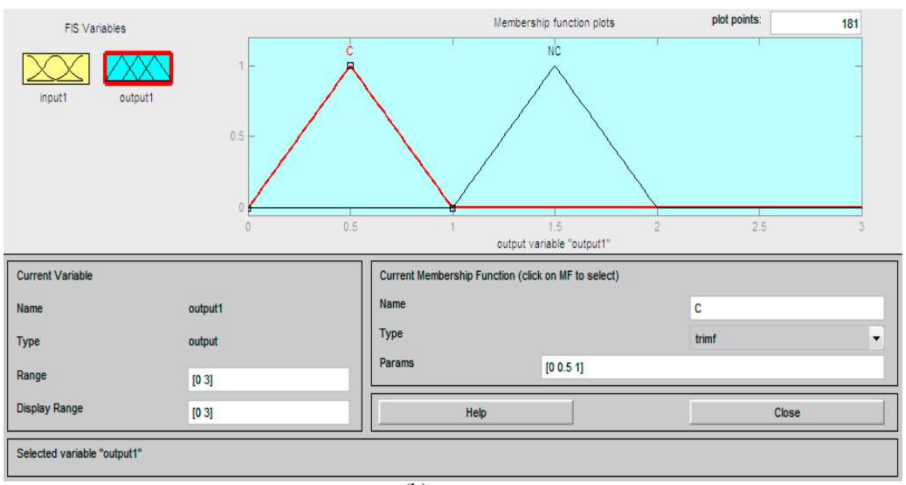
The examination of the considerable number of parameters can be seen measurable from the statistical response as seen in Fig.2 (a) (b) (c) (d) (e) which shows the actual response of all five statistical parameters for Cancer and Noncancer lung images. Graphical analysis of 25 images has been shown for each parameter. The graph shows the display of scattered points of each statistical and mathematical parameter for Cancer and Noncancer lung images. It can be observed from the graph that lesser the overlap of these statistical points for Cancer and Noncancer lung images, better is the usefulness of the input parameter for cancer investigation and it is found that each and every parameter has played a very crucial role in lung cancer diagnosis. On an average, all the average information parameters are having approximately the same statistical response for Cancer and Noncancer lung images, which also imply that all of them play an equally important role for lung cancer analysis which leads towards a good diagnosis.

4.3 Performance Analysis of Fuzzy Logic

The result of fuzzy Logic is observed in figure 3. The whole working of the Fuzzy depends on the input parameters. The input to the Fuzzy is the parameter range of five identified statistical parameters through current membership function. The triangular membership function (trimf) is used as the current membership function. The output is obtained using the Fuzzy Rules. Fuzzy rules are the simple statements implemented using logic 'OR'. The membership function actually is used to map the membership values between zero and one. Lung cancer diagnosis works in two steps. Step 1 is Input image pre-processing with the calculation of input parameters range and step2 is the decision making by Fuzzy Logic.



(a)



(b)

Fig. 3(a) Input of Fuzzy Logic (b) Output of Fuzzy Logic

The decision making of Fuzzy Logic to classify the image into cancerous or noncancerous is based on Fuzzification and Defuzzification process using a set of Fuzzy Rules. The result clearly indicates that, Fuzzy Logic successfully classifies the microscopic lung images into Cancer and Noncancer.

4.4 Accuracy of Input Parameters

The accuracy of each individual statistical and mathematical parameter is calculated on the basis of how well each parameter differentiates a test image into Cancer and Noncancer lung image. The calculated accuracy of each statistical and mathematical parameter for the lung cancer images is shown in Table 4 and Fig.4.

Table 4 Accuracy of Average Information Parameters

Sr no.	Average Information Parameters	Accuracy (%)
1	Standard. deviation	62.23
2	Variance	63.78
3	Mean	56.96
4	Entropy	58.20
5	Mean Square Error	61.92

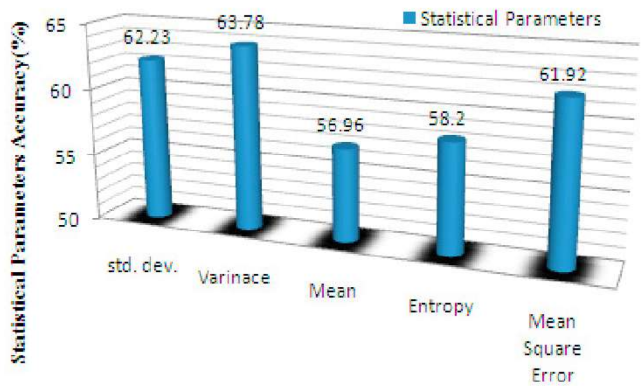


Fig.4 Accuracy of input Parameters for Lung Cancer

The accuracy of Entropy, Mean, Standard Deviation, Variance and MSE are almost similar ranging from 56.96% to 63.78% which also indicates the equal importance of each parameter for lung cancer analysis and diagnosis. Variance is having the higher response, whereas the Mean parameter has a lesser response as compared to other parameters. Here the two types of accuracies are calculated, one is individual parameter accuracy which is calculated on the basis of how well each parameter is able to differentiate the images into cancerous and noncancerous images and the other is the overall accuracy of the Average Information Method which is calculated using standard diagnostic test.

4.5 Standard Diagnostic Test

The three most significant measures of the diagnostic test are sensitivity, specificity, and accuracy [12], which together is used to verify the performance of the proposed method. The following standard equations are used in the diagnostic test to analyze the performance of the proposed method.

$$Sensitivity = \frac{T1}{T1+T4} \tag{5}$$

$$Specificity = \frac{T2}{T2+T3} \tag{6}$$

$$Accuracy = \frac{T1+T2}{T1+T2+T3+T4} \tag{7}$$

Where T1 is true positive, T3 is the false positive, T2 is true negative and T4 is false negative values of the image. The calculated values of true positive, false positive, true negative and false negative is shown in table 5

Table 5 Calculations of Standard Diagnostic Test for the Proposed Method

Proposed Method	T1	T3	T2	T4
Average Information	165	77	55	25

Table 6.Standard Diagnostic Test of the Proposed Method

Proposed Method	Sensitivity	Specificity	Accuracy
Average Information	86%	41%	68%

The Sensitivity is 86%, Specificity is 41% and Accuracy of the proposed method obtained is 68% as shown in Table 6 and fig 5. The standard performance test clearly indicates that the proposed methodology is a successful method for lung cancer diagnosis. Again to be more precise, the proposed method is also checked through ANOVA, which is one of the standard statistical methods. Both standard diagnostic tests from fig5and ANOVA results from fig6 are fantastic for the proposed method.

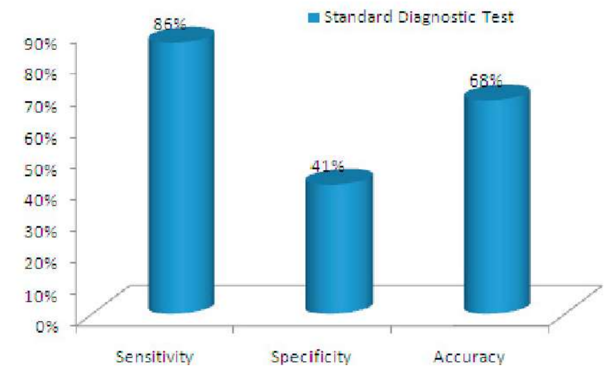


Fig5.Standard Diagnostic Test Results of the Proposed Method

4.6 Results of ANOVA

After the successful implementation of standard diagnostic test, this proposed is also checked by yet another standard statistical method called as ANOVA (Analysis of Variance). In proposed system five groups are considered according to used input parameters. ANOVA [13]-[14] is applied on the proposed system with the important calculations as follows:

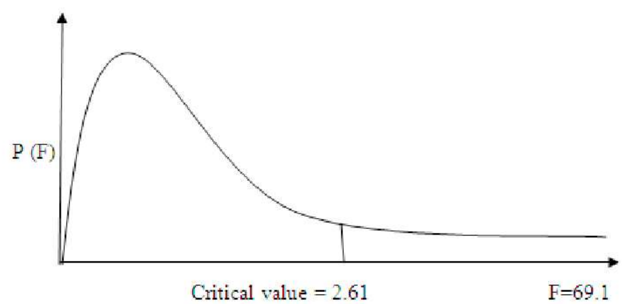


Fig 6 Critical value calculation on F distribution

Total sum of squares (TSS) = 66050543

Sum of squares between the groups (SSB) = 5.68E+07, Sum of squares with in the groups (SSW) = 9.25E+06

F ratio=SSB/SSW

$F (4, 45) = 69.1$, $p< .05$(p= significance factor)

Critical value = **2.61 (**approximately according to F-Distribution table for F (4, 45))

F (4, 45) is relative frequency

F test value > Critical value, which can be observed from Fig.6, i.e. 69.1 >2.61, hence the proposed method successfully rejects a null hypothesis.

5. Conclusions

Among various average information parameters, selective statistical and mathematical parameters are recognized for lung cancer investigation and diagnosis. For the determination of these input parameters, iteration method over predetermined lung cancer microscopic images is utilized. These parameters are tested and verified for the database of 323 microscopic lung pictures including cancer and non-cancer by utilizing image processing techniques with MATLAB. It is observed that each of the measurable statistical and mathematical parameter is ultimately an imperative part in lung cancer determination. The overall exactness of each individual statistical parameter in terms of accuracy ranges from 56% to 64% for cancer investigation. The Sensitivity is 86%, Specificity is 41% and Accuracy of the method is calculated as 68%, which is ascertained on the basis of the number of pictures analyzed and diagnosed accurately and checked successfully by standard diagnostic results. The proposed method also successfully rejects a null hypothesis test by implementation of ANOVA. Fuzzy Logic functions excellently as an image classifier for this proposed technique. The proposed strategy is working palatably and can likewise be improved in the future scope using a half-breed mix of scientific, measurable, basic parameters or even designing a new statistical and mathematical parameter enhancing the execution and precision of the proposed calculation. In any case, doubtlessly this strategy is one of the significant progressive developments towards the therapeutic research field for lung cancer investigation and analysis.

References

[1] Joes Vilar (2013). Breathe Easy, How Radiologic Helps to Find and Fight Lung Diseases, *European Society of Radiological*, Chapter1-4, 1-60.

[2] J. B. Walter & D. M. Pryce (1961). The Histology of Lung Cancer. *National Library of Medicine, National Institute of Health*, 15(3), 107-116.

[3]Kale Vaishnaw G. (2014). Imaging Techniques for Lungs Analysis, *International Journal of Scientific & Engineering Research (IJSER)*, 5(4), 1-4

[4] Muhammad Qurhanul Rizqie, Nurul Shafiqa Mohd Yusof, Rino Ferdian Surakusumah, Dyah Ekashanti Octorina Dewi, Eko Supriyanto and Khin Wee Lai (2015)," Review on Image Guided Lung Biopsy", IJN-UTM Cardiovascular Engineering Center, *Springer Science and Business Media Singapore*:41-50

[5]Vaishnaw Gorakhnath Kale (2015).An Overview of Microscopic Imaging Technique for Lung Cancer & Classification, IEEE *International Journal of Innovation in Engineering, Research and Technology [IJIERT]*, ICITDCEME'15 ,1-4.

[6].Kale Vaishnaw G., Vandana B. Malode (2016). New Approach of Statistical Analysis for Lung Disease Diagnosis using Microscopy images" *International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT,* 378-383.

[7] K.Punithavathy, M.M.Ramya, Sumathi Poobal (2015).Analysis of Statistical Texture Features for Automatic Lung Cancer Detection in PET/CT Images, *International Conference on Robotics, Automation, Control and Embedded Systems–RACE2015,*14-18, https://doi.org/10.1109/RACE.2015.7097244.

[8] Jay L. Devore, Kenneth N. Berk (2012). Modern Mathematical Statistics with Applications, *Springer Science + Business Media.*

[9] Kale Vaishnaw G (2016). Lung Cancer Analysis by Quality Measures" *International Journal of Modern Trends in Engineering and Research*, 3(4), Special Issue of ICRTET conference, 738-741.

[10] Information Resources Management Association (2017). Fuzzy Systems: Concepts, Methodologies, Tools and Applications*, IGI Global Publications*: 1-447.

[11] Atın¸C Yilmaz, K¨ur¸sat Ayan (2013). Cancer Risk Analysis by Fuzzy Logic Approach and Performance Status of the Model, *Turkish Journal of Electrical Engineering & Computer Sciences*: 897 – 912.

[12]Wen Zhu, Nancy Zeng, Ning Wang (2010). Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical Implementations, *NESUG2010*, Health care and life sciences, 1-9.

[13] Kaouther, E. K., Fares, B., & Souhir, E. (2013).Detect the Tumour with Numerical Analysis and With ANOVA Technique for MRI Image, *International Journal of Engineering, 3*(1), 257–260.

[14] El.Kourd Kaouther, Seif eddine Khelil , Saleh Hammoum (2015).Study With RK4 & ANOVA The Location Of The Tumour at the Smallest Time for Multi-Image,1-6.