International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 4114–4126

A Stage-wise Lung Cancer Detection Method Using Machine Learning

Dr. Sheshang Degadwala[1], Dhairya Vyas[2]

Associate Professor, Computer Engineering, Sigma Institute of Engineering
Vadodara, Gujarat, sheshang13@gmail.com
Managing Director, Shree Drashti Infotech LLP
Vadodara, Gujarat, dhairayvyas@live.com

Abstract

Prior, Lung threatening development is the basic driver of sickness passing worldwide among the two individuals, with more than 1 million spent each year. Lung Cancer experiences have been generally loosened up difficulty looked by individuals over a continuous couple of decades. Right when an individual has lung sickness, they have sporadic cells that pack together to shape a tumor. A destructive tumor is a social occasion of harmful development cells that can form into and pulverize near to tissue. Lung malignancy location at a beginning period has gotten significant. Presently, numerous methods are utilized dependent on picture preparing and profound learning strategies for lung malignancy characterization. For that tolerant Computer Tomography (CT) filter pictures are utilized to recognize and group lung malignant growth phase of that knobs. In this paper, we present a progressively precise technique for distinguishing lung disease stage inpatient. We have considered the shape and surface highlights of CT filter pictures for the order. The methodology is actualized utilizing different learning strategies and looked at their outcome.

Keywords: *Decision Tree, KNN, RF, DF, Machine learning*

I. INTRODUCTION

Lung Cancer is the disease that begins in the lung. It is the second most typical disease in individuals. Right when signs are discovered up close and personal, lung dangerous development tests are done to choose the sort of lung sickness and if it has spread. Exactly when threatening development spreads, this is called metastasis.[10]. Early discovery of lung tumor is finished by utilizing many imaging strategies, for example, Computed Tomography (CT), Sputum Cytology, Chest X-beam and Magnetic Resonance Imaging (MRI). Identification implies grouping tumor two classes (i)non-carcinogenic tumor (generous) and (ii)cancerous tumor (malignant)[11].The disease is the most tricky ailment in the clinical field, since it ought to be distinguished at before stage to lessen the multifaceted nature of the conclusion and treatment[12].Most frequently, this adjustment in lung cells happens when individuals take in risky, poisonous substances. Regardless of whether you were presented to these substances numerous years back, you are still in danger for lung malignant growth. In this way, it is important to characterize phases of lung.

Types of Stages [14]:

Stage I: Cancer is found in the lung, however, it has not spread outside the lung. The essential tumor is 3 centimeters (cm) or less. Medical procedure to expel a part of the lung might be all you need. Chemotherapy may likewise be suggested, particularly in case you're at high danger of repeat. Stage I indications are the brevity of breath, rapines, hacking, and chest torment.

ISSN: 2233-7857 IJFGCN

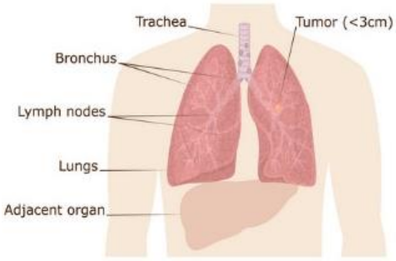

Figure 1. Stage I [13]

Stage II: Cancer is found in the lung and near to lymph center points. Fundamental tumor is 3-5 centimeters (cm). You may require clinical strategy to remove part or the sum of your lung. Chemotherapy is generally suggested. Stage II side effects are Coughing up blood or mucus, Wheezing and brevity of breath, Weight misfortune and loss of hunger, Chest torment that declines when breathing profoundly or chuckling.

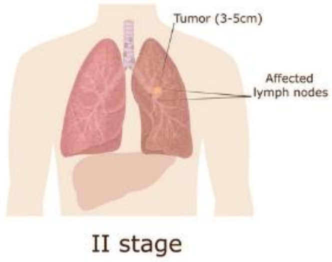

Figure 2. Stage II [13]

Stage III: Disease is in the lung and lymph center points in the chest. Basic tumor is >5 centimeters (cm).You may require a mix of chemotherapy, clinical strategy, and radiation treatment. Stage III signs are Trouble breathing, being winded or short of breath, Pain in chest domain, Wheezing sound when breathing, Voice changes (hoarser), Unexplained drop in weight, Bone anguish (may be in the back and may feel progressively unfortunate around night time), Headache.

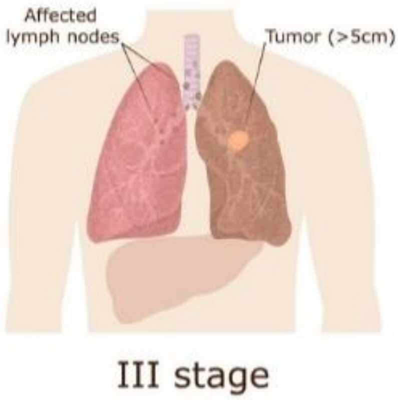

Figure 3. Stage III [13]

Stage IV: Cancer has spread to the two lungs, into the area around the lungs, or to distant organs. Fundamental tumor is >7 centimeters (cm). Particularly hard to fix. Choices join clinical methodology, radiation, chemotherapy, coordinated treatment, and immunotherapy. Stage IV appearances are Body a throbbing excruciating quality, Fatigue, Frequent headaches, Increased or decreased yearning.

ISSN: 2233-7857 IJFGCN

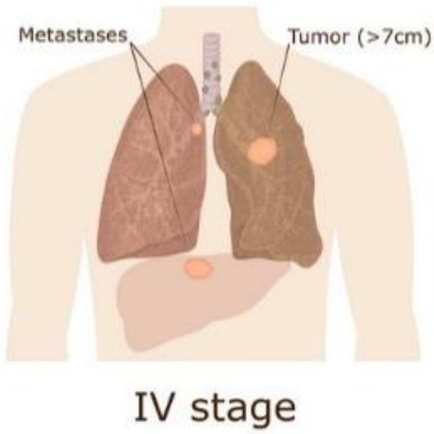

Figure 4. Stage IV [13]

II. RELATED WORKS

Janee Alam1 et al.[1] depicts lung malignant growth multistage location utilizing this strategy are SVM Classifier, GLCM highlight, and Watershed Transform. This paper confinements are, little dataset, low precision, and utilized huge element measurements.

Emine CENGİL et al.[2] depict the profound learning way to deal with lung malignancy and they utilized Convolution Neural Network (CNN). This paper has a few restrictions are grouped just dangerous or not, additionally improve engineering and framework to improve the huge dataset.

Manatee Kurkure1 et al.[6]. utilized Genetic applicant Group search (GCGS) Algorithm and Naïve base classifier. This paper has a few confinements are improved grouping strategies, low exactness and countless datasets ought to be required for better outcomes.

Suren Makajua et al.[7].used water concealed Transform and SVM Classifier. This paper has a few constraints don't group the level of the knob, not arrange various stages and furthermore improve the exactness must be expanded by pre-handling.

Mehdi Fatan Serj et al.[3] portray DCNN for the lung malignant growth recognition approach and they have utilized the DCNN classifier. This paper has a few impediments are division isn't done, large component vector and GPU are required to run the framework.

F. Ghazvinian Zanjani1 et al.[8] depict disease location in Mass Spectrometry picture information utilizing RNN(Recurrent Neural Network). A few constraints of this paper are improved engineering of RNN, Improve order results, and increment execution.

Moffy Vas et al.[4] depicts for identifying lung disease techniques that they have utilized are Morphological, Harrick, GLCM, and ANN. This paper has a few impediments are Morphological division require to change organizing component and they group just two classes.

III. BACKGROUND

In this part brief explain about tumor features. Mainly two types of feature shape and texture are used for stage tumor classification.

A. Feature Extraction:

It delineates the relevant shape and surface information contained in a model with the objective that the task of orchestrating the model is made straightforward by an appropriate strategy. In structure

ISSN: 2233-7857 IJFGCN

affirmation and in picture dealing with, incorporate extraction is a remarkable sort of dimensionality decline. At the point when features have been removed, they may be used to make AI models.

Table I: Shape Feature

Shape Features [5]	
Area	The area is the number of pixels in the shape. The total surface (or the region of square units) that occupied in a closed surface is termed as **area**. It depends on the shape of the surface.
Perimeter	The perimeter (length) is the quantity of pixels in the limit of the item. The separation of the shut shape (or figure) is named as border. The border of the square shape is the total of the lengths everything being equal.
Eccentricity	Eccentricity is the ratio of the length of the short axis to the length of the long axis of an object: Eccentricity= $\frac{axislength(\min)}{axislength(\max)}$

Table II: Texture Feature

Texture Features [4]	
Contrast	Contrast is defined as the difference between the highest and lowest intensity value of the image. CON= $\sum_I \sum_J (i-j)^2 p(i,j)$
Correlation	Correlation is the process of moving a filter mask often referred to as kernel over the image and computing the sum of products at each location. Correlation is the function of displacement of the filter. Correlation $\frac{\sum_i \sum_j (i-\mu_i)(j-\mu_j)N_d(i,j)}{\sigma_i \sigma_j}$

ISSN: 2233-7857 IJFGCN

Energy	Energy is used to describe a measure of "information" when formulating an operation under a probability framework such as MAP (maximum a priori) estimation in conjunction with Markov Random Fields. Energy= $\sum_i \sum_j N_d(i, j)$
Homogeneity	Homogeneity reflects the homogeneity of image textures and scaled the local changes of image texture. High values of homogeneity denote the absence of intra-regional changes and locally homogenous distribution in image textures. Homogeneity= $\sum_i \sum_j \frac{N_d(i, j)}{1+\lvert i-j \rvert}$

IV. PROPOSED SYSTEM

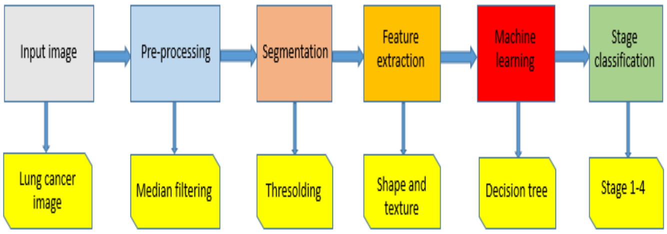

Figure 5. Proposed system

From the above figure right off the bat, we need to take the contribution of the lung malignant growth picture. At that point, after we will be pre-handling that picture by applying middle separating for lessening commotion with the goal that picture can be progressively viable. After Pre-preparing finished then division applies. Picture Segmentation of lung knobs in CT pictures is utilized to pick up the better aftereffect of pixels esteem thresholding is utilized. Afterimage division, we utilized element extraction shape and surface consolidate. Along these lines, we get increasingly precise outcomes in lung stage order. For Classification purposes, we use the Decision Tree (DT) classifier

ISSN: 2233-7857 IJFGCN

which is the most extensive classifier. In Machine Learning Algorithm Decision Tree are developed by means of an algorithmic methodology that recognizes approaches to part an informational collection dependent on various conditions. Chart book we can arrange lung stages (1-4) by utilizing this AI calculation.

A. Input:

The data picture is a little subset of pictures from the ailment imaging annal. They include the inside cut of all CT pictures taken where generous age, philosophy, and contrast names could be found. This result in 475 game plan from 69 special patients.

- **Data Header**

Images (DICOM, 18.3GB)
Tissue Slide Images (web)
Clinical Data (TXT)
Genomics (web)

- **TCIA Archive Link**

"https://wiki.cancerimagingarchive.net/display/Public/TCGA-LUAD."

B. Pre-Processing:

A separating window of 3x3 is applied on the picture where the focal pixel esteem is supplanted with the middle estimation of the relating pixels which assists with evacuating boisterous pixel of unessential worth.

252	255	252
252	2	255
255	255	253

Figure 6. Median Filtering Matrix

Neighborhood value:
2,252,252,252,253,255,255,255,255
Median value: 253
Here the value of 2 is replaced with the value 253 which removes noise and fills the faded portion in the image.

C. Segmentation:

Otsu method is one of the most successful methods for image thresholding.

ISSN: 2233-7857 IJFGCN

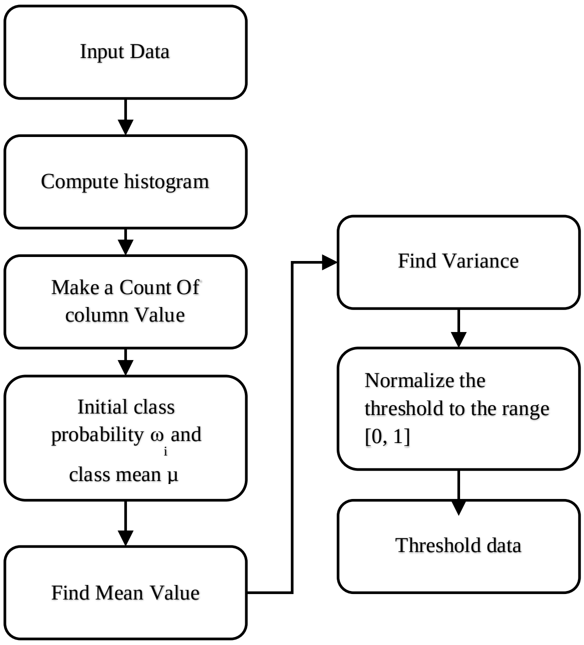

Figure 7. Otsu Segmentation

Changing over a greyscale picture to monochrome is a typical picture preparing the task. Otsu's strategy, named after its creator Nobuyuki Otsu, is one of the numerous banalization calculations. In the least complex structure, the calculation restores a solitary power limit that different pixels into two classes, forefront, and foundation.

D. Feature Extraction:

In this part, we have used a mix of two-component primarily shapes and surfaces. Fit as a fiddle highlight we have utilized territory, border, and flightiness. While on the surface we have utilized GLCM highlight for example vitality, entropy, connection, and Homogeneity. So complete component of the highlight is seven. Which will use for the train over the model.

E. Machine Learning:

AI is procedure of preparing information dependent on their group. In this part we have mostly utilized four classifier for example SVM, KNN, DT and RF.

(I) SVM (Support Vector Machine):

Bolster vector machine is regulated learning calculation which can be utilized for both Classification and relapse difficulties. It can take care of straight and non-direct issue and function admirably for any commonsense issue. The calculation makes a line or a hyper-plane which isolated the information into classes.

(ii) KNN (K-closest Neighbor):

K-Nearest neighbors is one of the clearest estimation used in learning for backslide and course of action issue. It uses for direct issues so to speak. KNN count use data and gathering new data point reliant on comparability measures.

(iii) DT (Decision Tree):

A Decision Tree Classifier works by separating a dataset into littler and littler subsets dependent on various rules. Distinctive arranging rules will be utilized to separate the dataset, with the quantity of models getting littler with each division.

ISSN: 2233-7857 IJFGCN

(iv) RF (Random Forest):

A sporadic woods (RM) is a Meta estimator that fits different decision tree classifiers on various sub-trial of the dataset and uses averaging to improve the perceptive precision and control over-fitting. The sub-test size is reliably equal to the primary information test size.

F. **Stages Classification:**

Here we are using trained models to classifying four stage of Lung Tumor: (1) Stage-I, (2) Stage-II, (3) Stage-III and (4) Stage-IV.

V. EXPERIMENTAL RESULT AND ANALYSIS

A. Result with only shape feature

1) Shape feature result for SVM:

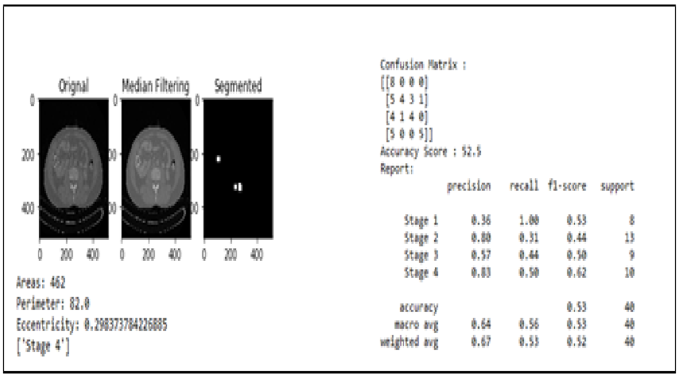

Figure 8. SVM

As appeared in the Figure-8 SVM classifier use shape highlight for lung tumor stage characterization. The disarray grid appears in first stage 8 out of 8 class genuinely characterized, two-phase 13 out of 4 really grouped, third stage 9 out of 4 genuinely ordered, and four-phase 10 out of 5 really arranged. So getting 52.5% exactness.

2) Shape feature result for KNN:

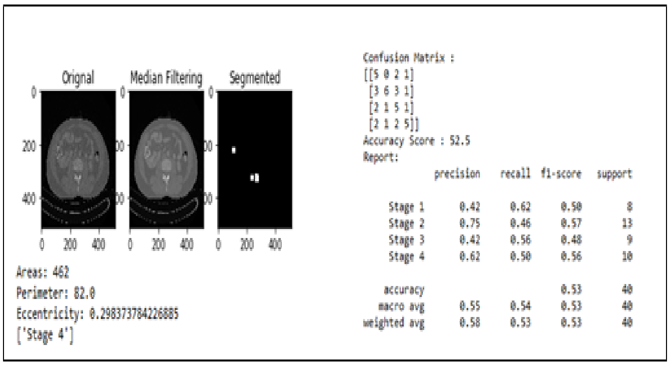

Figure 9. KNN

As appeared in the Figure-9 KNN classifier use shape highlight for lung tumor stage order. The disarray grid appears in first stage 8 out of 5 class genuinely grouped, two-phase 13 out of 6 really arranged, third stage 9 out of 5 genuinely characterized, and four-phase 10 out of 5 really ordered. So getting 52.5% precision.

3) Shape feature result for Decision Tree:

ISSN: 2233-7857 IJFGCN

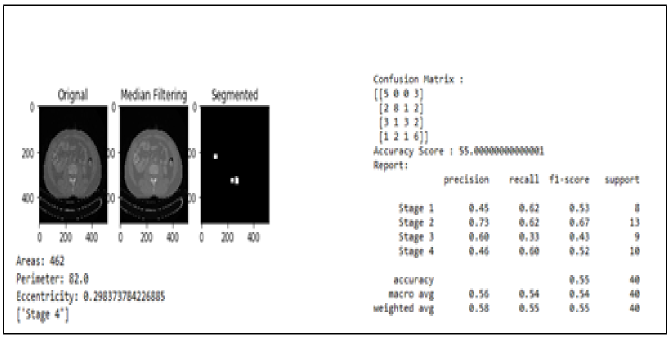

Figure 10. DT

As appeared in the Figure-10 DT classifier use shape highlight for lung tumor stage order. Disarray grid appears in first stage 8 out of 5 class really arranged, two-phase 13 out of 8 genuinely ordered, third stage 9 out of 3 genuinely characterized and four-phase 10 out of 6 genuinely grouped. So getting 55% exactness.

4) Shape feature result for Random forest:

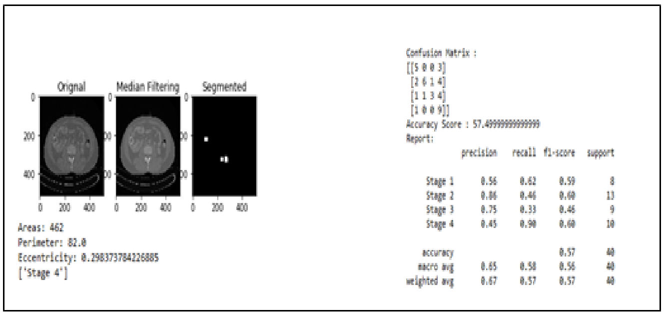

Figure 11. RF

As appeared in the Figure-11 RF classifier use shape include for lung tumor stage order. The disarray network appears in first stage 8 out of 5 class really ordered, two-phase 13 out of 6 genuinely characterized, third stage 9 out of 3 really grouped, and four-phase 10 out of 9 really arranged. So getting 57.49% exactness.

B) Result with Shape and Texture feature:

1) Shape and texture feature result for SVM:

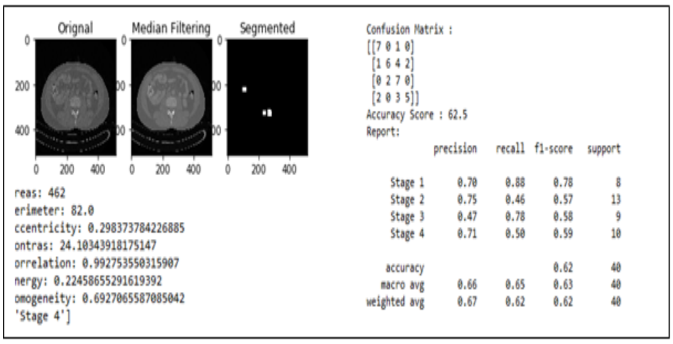

Figure 12. Shape and Texture SVM

As appeared in the Figure-12 SVM classifier use to shape and surface highlights for lung tumor stage arrangement. The disarray lattice appears in first stage 8 out of 7 class really characterized, two-phase 13 out of 6 genuinely ordered, third stage 9 out of 7 genuinely arranged, and four-phase 10 out of 5 really grouped. So getting 62.5% precision.

ISSN: 2233-7857 IJFGCN

2) Shape and Texture feature result for KNN:

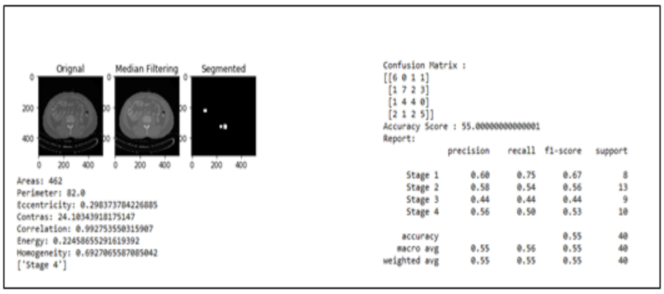

Figure 13. Shape and Texture KNN

As appeared in the Figure-13 KNN classifier use to shape and surface highlights for lung tumor stage arrangement. The disarray grid appears in first stage 8 out of 6 class genuinely ordered, two-phase 13 out of 7 really arranged, third stage 9 out of 4 genuinely grouped, and four-phase 10 out of 5 genuinely characterized. So getting 55% exactness.

3) Shape and Texture feature result for Decision Tree:

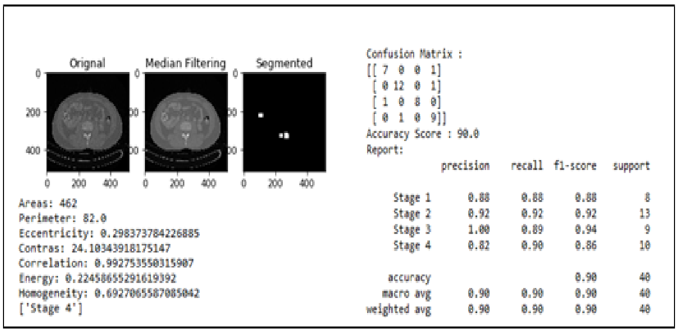

Figure 14. Shape and Texture DT

As appeared in the Figure-14 DT classifier use to shape and surface highlights for lung tumor stage arrangement. The disarray grid appears in first stage 8 out of 7 class really ordered, two-phase 13 out of 12 genuinely grouped, third stage 9 out of 8 really arranged, and four-phase 10 out of 9 really characterized. So getting 90% exactness.

4) Shape and Texture feature result for Random Forest:

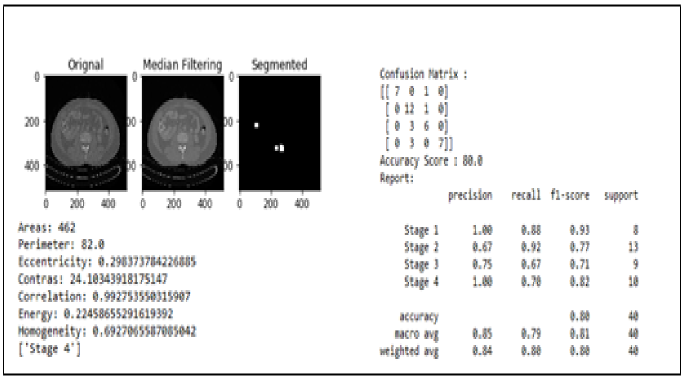

Figure 15. Shape and Texture RF

As appeared in the Figure-15 RF classifiers use to shape and surface highlights for lung tumor stage arrangement. The disarray grid appears in first stage 8 out of 7 class genuinely arranged, two-phase 13

ISSN: 2233-7857 IJFGCN

out of 12 really characterized, third stage 9 out of 6 genuinely grouped, and four-phase 10 out of 7 really ordered. So getting 80% precision.

C) Anaylsis and Graph:

Table III.Comparative Anaylsis

Method	Shape			
	Accuracy	Precision	Recall	F1
SVM	52.5	0.64	0.56	0.52
KNN	52.5	0.55	0.53	0.53
DT	**55**	**0.56**	**0.54**	**0.54**
RF	57.49	0.65	0.58	0.56

As appeared in Table III shows near investigation utilizing shape highlight with various classifier among them Decision Tree give hitter execution.

Table IV.Comparative Anaylsis

Method	Shape + Texture			
	Accuracy	Precision	Recall	F1
SVM	62.7	0.66	0.65	0.63
KNN	55	0.55	0.56	0.55
DT	**90**	**0.90**	**0.90**	**0.90**
RF	80	0.85	0.79	0.80

As appeared in Table IV shows near investigation utilizing shape and Texture include with various classifiers among them Decision Tree give player execution.

Table V. Accuracy Comparision Between Existing And Proposed Algorithm

Authors	Technique Applied	Accuracy
Janee et al, Sabrina et al,Alamgir et al.	Watershed transform,GLCM and SVM classifier	87%
Our Propose Algorithm	**Thresolding,Shape and Texture combine and Decision Tree classifier**	**90%**

As appeared in Table V shows a relative report between existing work and our proposed work. Proposed work gives hitter execution 90%.

ISSN: 2233-7857 IJFGCN

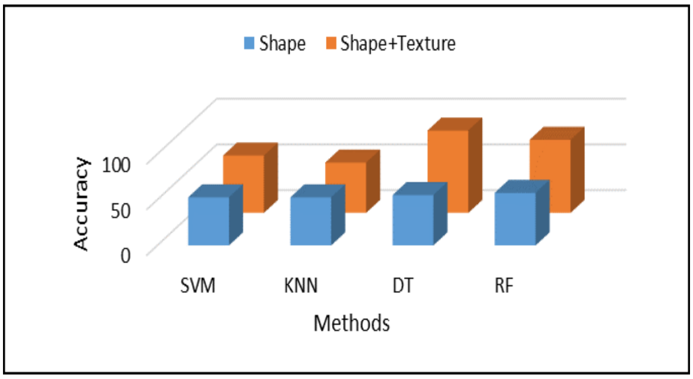

Figure 16. Analysis Graph

As appeared in Figure-15 all classifier precision expanded when utilizations shape and surface element other than just shape highlight. Additionally, Decision tree give most extreme execution than different classifiers.

CONCLUSION

In this exploration distinctive Machine learning draws near (SVM, KNN, DT, RF) are tried with shape and surface highlights. From the similar investigation, it demonstrates that shape and surface highlights perform better with choice tree classifier. Along these lines, In the learning approach getting 90% exactness. In this way, later on will utilization of profound learning approach for preferred execution over learning.

REFERENCES

1. J. Alam, A. Hossan and S. Alam, , “Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifie,” Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng. IC4ME2 2018, pp. 1–4, 2018.
2. E. Cengil and A. Çinar, “A Deep Learning Based Approach to Lung Cancer Identification,” 2018 Int. Conf. Artif. Intell. Data Process. IDAP 2018, 2019.
3. B. Lavi, “A Deep Convolutional Neural Network for Lung Cancer Diagnostic,” pp. 1–10.
4. A. Dessai and M. Vas, “Lung cancer detection system using lung CT image processing,” 2017 Int. Conf. Comput. Commun. Control Autom., pp. 1–5, 2017.
5. A. Alsadoon, S. Makaju, P. W. C. Prasad, , A. K. Singh, and A. Elchouemi, “Lung Cancer Detection using CT Scan Images,” Procedia Comput. Sci., vol. 125, no. 2009, pp. 107–114, 2018.
6. M. Kurkure and A. Thakare, “Classification of Stages of Lung Cancer using Genetic Candidate Group Search Approach,” IOSR J. Comput. Eng., vol. 18, no. 05, pp. 07–13, 2016.
7. A. Srinivasan A. Asuntha, A. Brindha and S. Indirani, “Lung cancer detection using SVM algorithm and optimization techniques,” J. Chem. Pharm. Sci., vol. 9, no. 4, pp. 3198–3203, 2016.
8. F. G. Zanjani et al., “CANCER DETECTION IN MASS SPECTROMETRY IMAGING DATA BY RECURRENT NEURAL NETWORKS Eindhoven University of Technology , SPS-VCA , 5612 AJ Eindhoven , The Netherlands Maastricht Multimodal Molecular Imaging Institute , University of Maastricht , The Netherlan,” 2019 IEEE 16th Int. Symp. Biomed. Imaging (ISBI 2019), no. Isbi, pp. 674–678, 2019.
9. M. Bharathi, S. Sasikala , and B. R. Sowmiya, “Lung cancer detection and classification using deep CNN,” Int. J. Innov. Technol. Explor. Eng., vol. 8, no. 2S, pp. 259–262, 2018.
10. X. K. Luo, Q. Z. Song, L. Zhao and X. C. Dou, “Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images,” J. Healthc. Eng., vol. 2017, 2017.

ISSN: 2233-7857 IJFGCN

11. A. Teramoto et al., “Automated classification of benign and malignant cells from lung cytological images using deep convolutional neural network,” Informatics Med. Unlocked, vol. 16, no. May, p. 100205, 2019.
12. C. A. O’Dowd, S. Tsim, R. Milroy, and S. Davidson, “Staging of non-small cell lung cancer (NSCLC): A review,” Respir. Med., vol. 104, no. 12, pp. 1767–1774, 2010.
13. Scio21,Four stage of lung cancer vector image, https://www.vectorstock.com/royalty-free-vector/four-stages-of-lung-cancer-vector-22417619
14. Winfield Boerckel, LCSW-R,Lung Cancer Program Coordinator, Cancer *Care, Lung cancer101,*https://www.lungcancer.org/find_information/publications/163-lung_cancer_101/265-what_is_lung_cancer

ISSN: 2233-7857 IJFGCN