

An Efficient Novel Machine Learning Approach for An Early Diagnosis of Bronchogenic Carcinoma

H.S. Saraswathi
Dept of CSE
JIT, Davanagere, India

Sriharsha G A
Dept of CSE
JIT, Davanagere, India

Sanath Kumar A L
Dept of CSE
JIT, Davanagere, India

Srujan K R
Dept of CSE
JIT, Davanagere, India

Manoj A Banakar
Dept of CSE
JIT, Davanagere, India

Abstract — Lung and bronchus cancer may be a common and sometimes aggressive sort of cancer. This type of cancer often has no noticeable symptoms in its early stages. Detecting whether the lesion is malignant or benign more reliably and without the necessity for a biopsy would be challenging. So getting an accurate prediction of disease outcome is most vital for oncologists. To boost the advancement and drug of destructive conditions, AI procedures are used on account of their precise results. So as to extend accuracy, conducted a study here during which AI deep learning and machine learning algorithms are trained to sight malignant neoplastic disease tumors in laptop imaging scans. The algorithms analysis was then compared with that of half-dozen radiologists and so the result showed that AI was a lot correct once previous CT imaging wasn't accessible. So AI shows promise in detecting carcinoma.

Keywords — Lung cancer disease, Image processing, Machine learning, Support Vector Machine.

I. INTRODUCTION

Lung cancer is a disease of irregular cells multiplying and increasing into tumors. Malignancy cells continue to increment and structure new, abnormal cells. Malignant tumors that begin in some portion of the lung are called primary lung disease.

Following are the types of lung disease:

1. Small cell cancer- it's conjointly referred to as oat cell hurt. It creates and spreads a lot of quickly than non-smallcell respiratory organ malignancy development. It abundant of the time spreads to numerous elements of the body at a strategic planning stage.
2. Non-small cell cancer -This cancer has three varieties.
 - a. Adenocarcinoma- it is the most generally recognized quite respiratory organ sickness in non-smokers. In any case, it's found even additional in smokers or past smokers.
 - b. Squamous cell malignant neoplastic disease (epidermoid carcinoma). These diseases can generally start within the centerpiece of the lungs close to the

quality aviation routes (the bronchi).

c. Large cell carcinoma-It can generally develop and unfold to varied organs. this may create lungs more durable to treat.

The primary side effects are torment in the chest or rib, the hack can be constant, dry, with mucus or with blood, respiratory contaminations, the brevity of inhaling, wheezing, entire body weakness, or loss of craving, and swollen lymph hubs.

Tumor cells are those cells that create, in spite of the fact that when the body doesn't require them, and other than typical old cells, they don't pass. The malignancy cells present in the lung cause lung development illness.

Malignant tumors: 14 % of all new malignant growth analyses are primary lung diseases. All the blood in the body goes from the heart through the lungs, so malignant growth can undoubtedly spread to various parts of the body. Benign tumors are a less basic reason for respiratory infection. One model is a hematoma. These can pack encompassing tissue, however, they are typically asymptomatic.

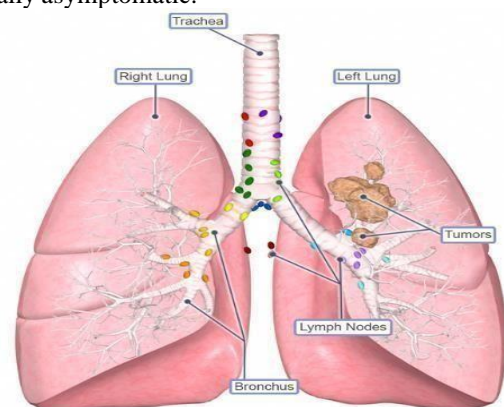


Figure 1: Lung affected by tumors

At the point when air enters the nose or mouth, it goes down the trachea, additionally called the windpipe. After this, it arrives at an area called the carina. At the carina, the windpipe parts into two, making two principle stem bronchi. One prompts the left lung and the other to the right lung. From that point, similar to branches on a tree,

the channel-like bronchi split again into littler bronchi and afterward much littler bronchioles.

This ever-decreasing pipework eventually terminates in the alveoli. Lung malignant growth spreads when cells sever or from a tumor and travel through the circulatory system the lymphatics to far-off locales of the body and develop. This cell disclosure is a basic issue for medical specialists.

ARTIFICIAL INTELLIGENCE IS BETTER THAN HUMANS AT SPOTTING CARCINOMA

Lung disease is among the foremost widely known malignant growths around the world. A majority of patients are determined to possess non-small cell carcinoma growth (NSCLC) and have a five-year endurance pace of 18%. The group made different profound learning models utilizing picture Net, a neural system that distinguishes normal articles from pertinent highlights.

People prepared the models utilizing sequential CT outputs of 179 patients with stage III NSCLC who had gotten chemoradiation treatment. Every patient created up to four pictures taken before treatment and at one, three, and half year follow-up for a sum of 581 pictures.

Inside clinical oncology, AI has progressively been applied to tackle the intensity of the electronic health record (EHR) especially, AI-based characteristic language preparing methods have indicated a guarantee in anticipating the advancement of maladies across huge medicinal services frameworks. DL-based AI calculation displaying EHR had the choice to foresee the development of an assortment of maladies with 93% precision generally speaking, including malignant growths of the prostate, rectum, and liver.

II. RELATED WORK

Sanjukta aristocrat Battle of Jena, Dr. Thomas Saint George, and Dr. Narain Ponraj projected a model regarding prevailing carcinoma detection techniques that square measure available within the literature.

Local Binary Pattern (LBP) :

It is an important however helpful surface government that names the pels of a picture by thresholding the region of every pixel and considers the result a binary variety. In lightweight of its discriminative force and procedure straightforwardness, LBP surface administrator has become a regular methodology in varied applications it would be seen as a binding approach to manage the typically dissimilar quantitative and auxiliary models of surface examination.

Perhaps the foremost important property of the LBP administrator in evident applications is its capability to monotonic gray-scale changes caused, for instance, by enlightenment assortments. Another large property is its process straightforwardness, which makes it doable to separate photos in testing continuous settings. At the lower place documentation is employed for the LBP operator: LBPP, Ru2. The subscript addresses the exploitation of the administrator in an exceedingly (P,R) neighborhood.

Superscript u2 speaks to victimization merely uniform models and naming every extraordinary model with a

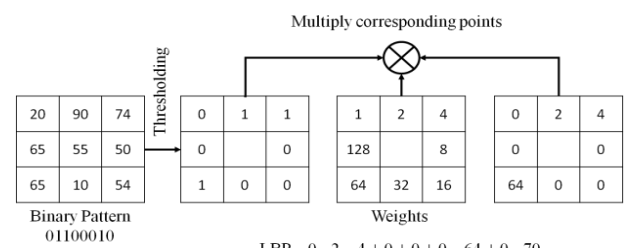
singular mark. when the LBP marked image $fl(x,y)$ has been gotten, the LBP bar chart will be delineated as

$$Entropy = \sum_{i,j=0}^{N-1} -\ln(P_{ij}) P_{ij} \quad (1)$$

in which n is the quantity of various names sent by the LBP administrator, and I is one if Associate in Nursing is substantial and zero if Associate in Nursing is invalid. At that point, once the image fixes whose histograms square measure to be thought of as having completely different sizes, the histograms should be normalized to induce a wise depiction:

$$H_i = \sum x, y \{ fl(x, y) = i \}, i = 0, \dots, n-1, \quad (2)$$

$$N_i = H_i \sum n-1 j = 0 H_j. \quad (3)$$

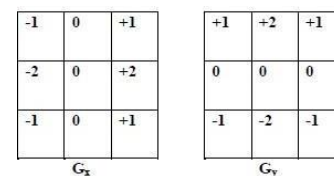


Mamta Joshi proposed a model for the Comparison of the Canny edge detector with the Sobel edge detector.

Sobel edge identification :

Sobel edge recognizable procedure system includes two or three 3 convolution portions. One half is barely the opposite turned by 90° as shown up in Figure.

These elements are expected to retort maximally to edges running vertically and on A level plane similar to the element network of the image, with one kernel for all of the 2 inverse headings. The kernels may be applied freely to the information image, to create separate estimations of the slope section toward every path. These would then have the choice to be coagulated along to seek out the whole extent of the slope at every purpose and also the course of that inclination.



Canny technique:

The Canny edge area computation is referred to broadly as the perfect edge pointer. The Canny calculation uses a perfect edge discoverer subject to a lot of standards which join finding the most edges by restricting the mistake rate, stepping edges as eagerly as possible to the genuine edges to enhance confinement, and checking edges just once when a solitary edge exists for a negligible response[14]. As demonstrated by Canny, the ideal channel that meets

each of the three standards can be gainfully approximated using the primary auxiliary of a Gaussian capacity.

The principal stage includes smoothing the image by convolving with a Gaussian channel. this is often trailed by finding the slope of the image by managing the smoothened image through a convolution action with the auxiliary of the Gaussian in each of the vertical and horizontal headings. This methodology mitigates problems connected with edge discontinuities by perceiving robust edges, and economical the many weak edges,

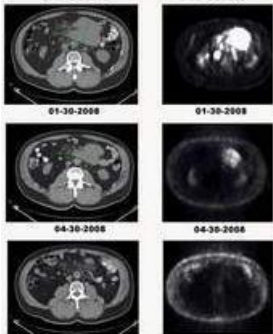


Figure 4: Output of the edge detection techniques

Finally, hysteresis is used as strategies for taking out streaking. Streaking is the isolating of an edge structure achieved by the administrator yield fluctuating above and underneath the edge. The figure shows the output of the distinctive edge discovery procedure of the given info picture as appeared.

Emmanuel Adetiba and Oludayo .O. Olugbara proposed a model for an artificial neural network for lung cancer prediction.[5]

Artificial neural network:

The data structures and quality of neural nets square measure are supposed to simulate associate memory. Neural nets learn by getting ready models, each one of that contains a glorious "info" and "result," framing probability-weighted associations between the two, that square measure place away within the information structure of cyberspace itself. Such structures "learn" to perform tasks by thinking models, typically while not being altered with task-explicit rules. For example, in image recognition, they may make sense of how to recognize pictures that contain malignant development by investigating model pictures that have been truly named as "disease" or "no disease" and using the results to recognizetumors in various pictures[12].

As delineated in Figure, an artificial neuron has a great deal of neurotransmitters related to the data sources and each piece of information has a related weight. A sign at the input is increased by the weight, the weighted inputs are included, and a linear combination of the weighted inputs is gotten. A bias, which isn't connected with any data, is added to a linear combination and a weighted sum is obtained as

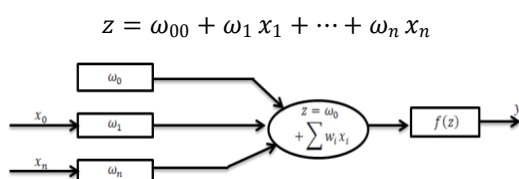


Figure 5: The neurons with the functional elements.

Rivansyah Suhendra's approach is a model of Support Vector Machine

It is one of the most productive conventional techniques utilized for grouping 'n' some features. The classification is finished by finding a hyperplane. SVM passes a linear distinguishable hyperplane through a dataset so as to arrange information into two gatherings. The hyperplane is utilized as a separator for any measurement. The best hyperplane is the one that amplifies the sting. The edge is the separation between the hyperplane and a few shutpoints. These nearby points control the hyperplane. This is the greatest edge classifier. Maximal edge classifiers help in expanding the edge of the hyperplane. This is best since it sums up the mistake.

Parameshwar R. Hegde proposed a model for binary and multiclass classification for the comparison of ML algorithms. Binary classification

Binary Classification is the task of characterizing the elements of a given set into 2 gatherings (envisioning that bundle each encompasses a spot with) seeable of an associate classification rule. Contexts requiring a choice regarding whether or not a factor has some abstract property, some predefined trademark, or binary classification include:

- Medical testing to settle on whether or not a patient has sure unwellness or not – the arrangement property is the proximity of the ill.
- A "leave or come behind short" test technique for instance picking if a specific has or has not been met – a go/no go gathering.

Here, the two classes considered were malignancy and Non-disease. Highlights were thought about in three blends; first, the color, texture, both color, and texture joined. Support vector machine gave the most elevated precision when similar classifiers were contrasted with texture highlights. Later when we joined both the features and tried SVM got the most noteworthy precision pace of 82.26%.

Numerous researchers have proposed strategies to develop computer-aided diagnosis frameworks for lung malignant growth recognition. The regular stage that is frequently utilized in conclusion framework advancement includes image preprocessing, region of interest (ROI) division, feature extraction, feature choice, classification, and evaluation. Moreover, the vast majority of the analysts have utilized just one ML algorithm with a solitary element mix. We have tried to elucidate, take into account and assess the show of various AI that square measure being applied to malignant development want and illustration. Expressly we tend to recognize different examples with relation to such AI systems getting used, the types of coaching information being expedited, the types of termination predictions being created, the types of malignant developments being inspected, and therefore the general execution of those procedures in imagery malady vulnerability or results. The framework was fabricated utilizing a model-based methodology based with respect to $L^a \times b \times$ color space. The above study contains color, texture component and SVM, CNN classifier model is viewed more frequently for

lung malignant growth recognition. At last, we found that the correlation of various AI algorithms will give a remarkable outcome. In this way, in the proposed examination we have tried four ML algorithms execution with two features set in three distinct blends. The Support Vector Machine (SVM) is utilized to prepare the 3D highlights joined with 2D asymmetry, outskirts abnormality, shading variegation, and distance across highlights.

In this proposed strategy, distinguishing proof of the lung malignancy with the assistance of CT pictures is finished. The strategy has a few phases wherein it starts with image acquisition, trailed by image preprocessing, segmentation, feature extraction, combination, classification, and calculating accuracy using a confusion matrix. The examination indicated the precision execution was around 87.8%. The general outcome demonstrated that the classifier effectively arranged the image into two classes. The best classifier was SVM with the exactness of up to 99.92%.

III. DATA GATHERING

CT images were acquired from the Malnad cancer hospital Shimoga, Cancer care unit Tumkur, Karnataka 2000 images were collected having different types of stages in lung cancer CT pictures incorporate less clamor when compared with X-beam and MRI pictures.

IV. PROPOSED METHODOLOGY

A. Image Preprocessing

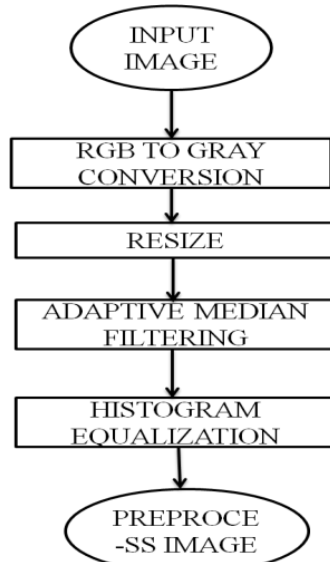


Figure 6: Flow Chart Of Preprocessing of an image

We are making utilization of straight Adaptive Median Filtering. It will directly extend the first advanced estimations of the remotely detected information into new appropriation. Toward the beginning, the original picture is converted from RGB to Gray and afterward, it is resized to the required Size. Later For which adaptive Median Filter and Histogram Equalization Techniques are applied. Finally, we got the preprocessed Output.

B. Image segmentation

Image Segmentation is the path toward analyzing a picture into completely different elements. this is often usually want to see objects or alternative noteworthy knowledge in progressive photos[18]. the target of division is to enhance or probably amendment the portrayal of an image into one thing that's dynamically vast and less exhausting to investigate.

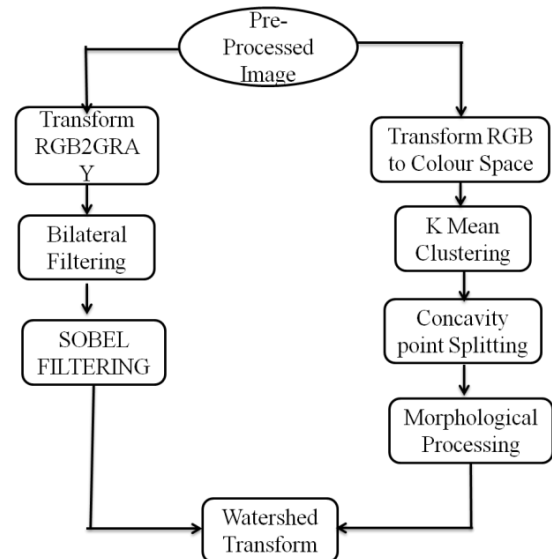


Figure 7:Flowchart of Marker Controlled Watershed Segmentation.

The figure shows the detailed Step By Step flow of watershed segmentation. In the beginning, the preprocessed image is given as input to the watershed segmentation. In beginning, the flow starts with two parts. In the first part, the preprocessing of the image is done as shown in the figure.

In the second stage, the preprocessed image is transformed into L*a*b Color space from RGB for which k-mean Clustering is obtained to find Concavity point splitting which is ended with Morphological processing. In the end output of both preprocessing and marker extraction is used to find the watershed transform of an image.

C. Feature Extraction

Color feature: Shading highlights are one of the significant parts of infection part detection and thus, in AI-based arrangement tools. Red, inexperienced, and Blue (RGB) color house is used during this investigation[19]. RGB colors were isolated and the normal worth was noted for additional examinations

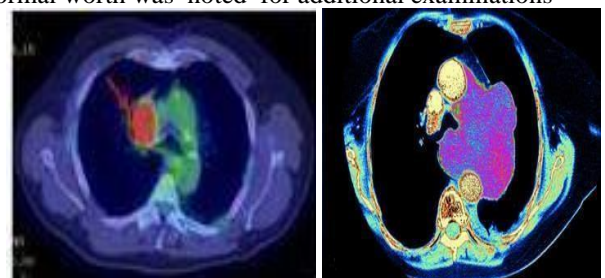


Figure 8: Extraction of Color Feature from CT scan

Texture feature: One of the most normally used techniques to extract textural information from pictures is the Grey Level Co-occurrence Matrix (GLCM). The GLCM technique offers wise surface information of an image that may be nonheritable simply from 2 pixels. It utilizes applied mathematics ways to seem at the surface by pondering the spatial association between pixels.

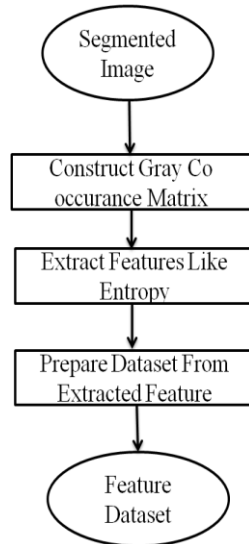


Figure 9: Flowchart of GLCM Algorithm

The accompanying highlights below are determined utilizing the GLCM algorithm:

$$Entropy = \sum_{i,j=0}^{N-1} -\ln(P_{ij}) P_{ij} \quad (eqn1)$$

$$Energy = \sum_{i,j=0}^{N-1} -\ln P_{ij}^2 \quad (eqn2)$$

$$Contrast = \sum_{i,j=0}^{N-1} P_{ij}(i-j)^2 \quad (eqn3)$$

$$Local Homogeneity = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1 + (i-j)^2} \quad (eqn4)$$

$$Correlation = \sum_{i,j=0}^{N-1} P_{ij} \frac{(i-\mu)(j-\mu)}{\sigma^2} \quad (eqn5)$$

D. Machine Learning Algorithms

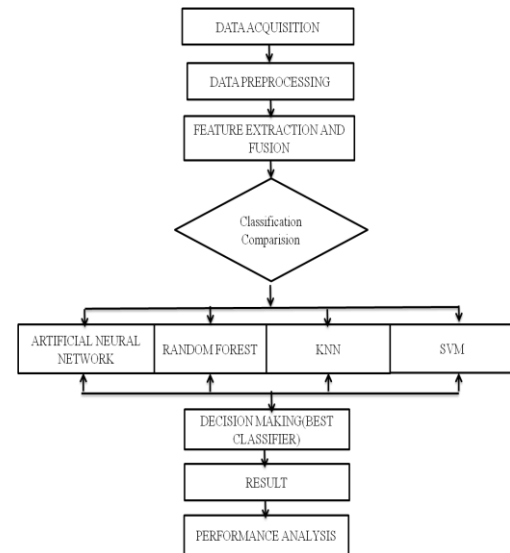


Figure 10: Block Diagram

- [1] Artificial Neural Network (ANN): ANN could be a factual model and works obsessed with the natural neural network guideline.[5] A neural network model has 3 layers; the most layer is the input layer, trailed by hidden layers, and eventually the output layer. The essential thinking ability increments aboard the number of hidden layers. This was dead with one hundred neurons (concealed layer's size), alpha incentive as zero.1 and two hundred cycles.
- [2] Random Forest: These are a gathering learning strategy for portrayal, drop off, and varied tasks that job by increasing unnumbered callhairstyle at coaching time and yielding the the category that's a technique for the categories ormeans wants for the individual trees[6].
- [3] K- Nearest Neighbor (KNN): Its motivation is to utilize a database wherein the information focuses are separated into a few classes to predict the characterization of a near sample Vvector Machine (SVM). It is onamongst the regular arrangement calculations, that utilize a choice plane to isolate a knowledge set having numerous categories. we've got dead it with the linear kernel, catching highlight, simply one cycle and with no organization.

D. Decision Making

Table 1: Accuracy of all classifiers.

CLASSIFIER	ACCURACY RATE
ANN	73.23
RANDOM FOREST	79.17
KNN	80.72
SVM	81.61

$$Cluster Shade = \sum_{i,j=0}^{N-1} P_{ij} (i - M_x + j - M_y)^2 \quad (eqn6)$$

$$Cluster prominence = \sum_{i,j=0}^{N-1} P_{ij} (i - M_x + j - M_y) \quad (eqn7)$$

$$\text{Standard Deviation}(\sigma^2) = \sum_{i=1}^{L-1} (x_i - m)^2 P(x_i) \quad (\text{eqn8})$$

SVM classifier showed better performance with an accuracy of 81% among them. So SVM is used to train the model in further process.

E. Testing

At the training stage, the dataset consists of varieties of lung cancer CT scan images and then followed by preprocessing and feature extraction process takes place. We have built a model by using the SVM algorithm[19]. At the testing stage, one patient image had given as input to the model. The model was successfully classified into a particular class with an accuracy performance of around 81%.

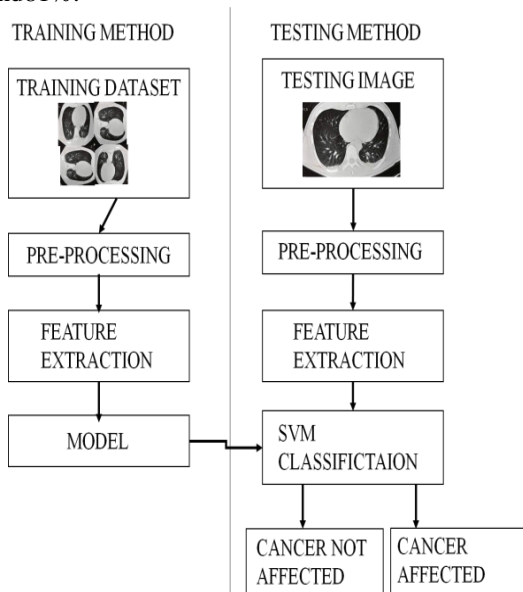


Figure 11: Flow Diagram

F. Support Vector Machine

Support vector machines (SVMs) are supervised learning models with connected learning calculations that examine information used for arrangement and fall away examination.

Even additional formally, a support vector machine builds up a hyperplane or set of hyperplanes in a very high-or perpetual dimensional house, which might be used for gathering, relapse, or varied tasks like exceptions recognizable proof. Normally, a far better than an average division is cultivated by the hyperplane that has the largest distance to the closest coaching datum of any category.

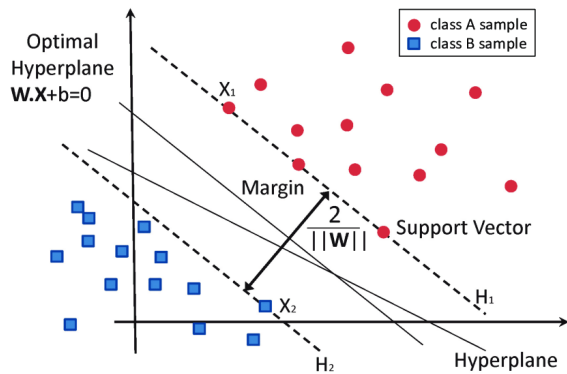


Figure 12: Hyperplane of Support Vector Machine

Definitions of SVM and Margin:

SVM works by mapping data to a high-dimensional featurespace so that data points can be categorized, even when the data are not otherwise linearly separable (also called the support vectors) and for other points,

$$|W^T X_i + b| > 1$$

that w is a vector perpendicular to the hyperplane, so we have:

$$f(x) = f\left(X_p + \frac{w}{\|w\|} \cdot r\right) = W^T X_p + W^T \frac{w}{\|w\|} \cdot r + b = \frac{W^T w}{\|w\|} \cdot r + b \quad (\text{eqn1})$$

Solve for margin length p :

$$p = \frac{f(x_+) - f(x_-)}{\|w\|} = \frac{2}{\|w\|} \quad (\text{eqn2})$$

Hypothetical Defence:

$$h \leq \frac{R^2}{\rho^2} + 1 \quad (\text{eqn3})$$

G. Bolster Vector Machine

Given the qualities $\alpha_1, \alpha_2, \dots, \alpha_N$ acquired by the arrangement of the double issue, the last SVM indicator can be communicated as

$$f(x) = W^T X_i + b = \sum_{i=1}^N \alpha_i y_i X_i^T X + b \quad (\text{eqn4})$$

SVM answer for the gathering:

Denote $\phi: \mathcal{R}^M \rightarrow F$ as a mapping from the M -dimensional credit space to the particularly dimensional space F .

Find α that expands subject to

$$\sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i x_j^t x_j \sum_{i=1}^N \alpha_i y_i = 0 \quad (\text{eqn5})$$

SVM answer by order:

Indicate $\phi: \mathcal{R}^M \rightarrow F$ as a mapping from the M -dimensional credit space to the very dimensional characteristic space F .

$$f(x) = W^T \phi(x_i) + b = \sum_{i=1}^N \alpha_i y_i \phi(x_i^T) \phi(x) + b \quad (\text{eqn6})$$

Viable Problem: Although SVM is effective in managing profoundly dimensional characteristic spaces, the way that the SVM prepares scales directly with the number of properties, and considering restricted memory space

could to a great extent confine the decision of mapping Φ .
Arrangement: Kernel Trick, By introducing the kernel
trick: The dual problem: find α that maximizes

$$f(x) = w^T \phi(x_i) + b = \sum_{i=1}^N \alpha_i y_i \phi(x_i)^T \phi(x) + b \quad (eqn7)$$

$$\sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (eqn8)$$

subject to $\sum_{i=1} \alpha_i y_i = 0$

The Resulting SVM is:

$$f(x) = w^T \phi(x_i) + b = \sum_{i=1}^N \alpha_i y_i K(x_{i,x}) + b$$

F. Performance Measurement

The outcome execution is estimated utilizing recall, precision, f-score, and precision. These measures are determined utilizing the confusion matrix that contains data about the actual and predicted characterization. The estimation is done to watch the performance of the proposed framework.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \\ = (100 + 45) / (100 + 45 + 9 + 6) = 0.90$$

Recall: Recall gives us an idea about when it's actually yes, and how often does it is predicted yes.

$$\text{Recall} = TP / (TP + FN) \\ = 100 / (100 + 6) = 0.94$$

Precision: Precision tells us about when it predicts yes, and how often is it correct.

$$\text{Precision} = TP / (TP + FP) \\ = 100 / (100 + 9) = 0.91$$

$$\text{F-measure} = (2 * \text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}) \\ = (2 * 0.94 * 0.91) / (0.94 + 0.91) = 0.92$$

V. RESULTS

Lung disease is the most widely recognized fatal malignancy in grown-ups around the world. CT imaging is a basic symptomatic apparatus to quantify the area, degree, size, and state of lung injuries to manage remedial choices for patients with lung malignant growth. Be that as it may, examination of CT images is restricted to what exactly is obvious to the natural eye, which can bring about contrasts in clinical consideration across various oncology communities. We had the option to show that the ML model was as precise as an accomplished radiologist in diagnosing the infection, and far better now and again.

Using both SVM and other classifiers increases the accuracy of detection and diminish false detection. Overall, this investigation pleasantly represented how the best possible structure, cautious usage, proper information determination, and approval of various machine learners can create a vigorous and exact malignancy hazard expectation tool.

Even more, starting late AI has been applied to harmful development expectations and figures. Among the better arranged and endorsed examinations indisputably AI methodologies can be used to altogether improve the precision of anticipating malignant development defenselessness, rehash, and mortality. This proposed technique gives very promising results compared with different used techniques.

VI. CONCLUSION

Lung disease is one of the riskiest illnesses in the world. The focal point of this exploration is an assessment of AI classifiers just as the forecast for lung disease discovery. Correct Diagnosis and early prediction of lung malignant growth can build the continuance rate. It is seen that this calculation had given extraordinary execution among all other individuals and utilized for additional procedures. The proposed procedure gives exceptionally encouraging outcomes compared with other utilized methods. The overall accuracy achieved by the SVM classifier is 96%. The early detection can be done and hence it will help the doctor to plan and do the better treatment for the patient. This can be implemented in many hospitals.

FUTURE WORK

Future work can embody the modification of the system to acknowledge pictures that have multiple scattered tumor on the respiratory organ. The projected work is extended by extra vary of algorithms on extra vary of datasets from immense medical info. The performance analysis of various classification algorithms will be done to offer higher results.

REFERENCES

- [1] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," Université de Montréal, Tech. Rep., 2012.
- [2] H. Larochelle, M. Mandel, R. Pascanu, and Y. Bengio, "Learning Algorithms for the Classification Restricted Boltzmann Machine," Journal of Machine Learning Research, vol. 13, pp. 643–669, Mar. 2012.
- [3] G. Desjardins and Y. Bengio, "Empirical Evaluation of Convolutional RBMs for Vision," Université de Montréal, Tech. Rep., 2008.
- [4] M. Norouzi, M. Ranjbar, and G. Mori, "Stacks of Convolutional Restricted Boltzmann Machines for Shift-Invariant Feature Learning," in IEEE Conference on Computer Vision and Pattern Recognition, 2009.
- [5] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in ICML, New York, New York, USA, 2009, pp. 609–616.
- [6] "Unsupervised learning of hierarchical representations with convolutional deep belief networks," Communications of the ACM, vol. 54, no. 10, pp. 95–103, Oct. 2011.
- [7] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [8] A. Depeursinge, A. Vargas, A. Platon, A. Geissbühler, P.-A. Poletti, and H. Müller, "Building a reference multimedia database for interstitial lung diseases," Computerized Medical Imaging and Graphics, vol. 36, no. 3, pp. 227–38, Apr. 2012.
- [9] A. Depeursinge, A. Foncubierta-Rodríguez, D. Van de Ville, and H. Müller, "Lung Texture Classification Using Locally-Oriented Riesz Components," in MICCAI, 2011, pp. 231–238.

-
- [10] —, “Multiscale Lung Texture Signature Learning Using the Riesz Transform,” in MICCAI, 2012, pp. 517–524.
 - [11] A. Depeursinge, D. Van de Ville, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Muller, “Near-affine- invariant texture learning for lung “ tissue analysis using isotropic wavelet frames.” IEEE transactions on information technology in biomedicine: a publication of the IEEE Engineering in Medicine and Biology Society, vol. 16, no. 4, pp. 665–75, Jul. 2012.
 - [12] A. Depeursinge, T. Zimec, S. Busayarat, and H. Muller, “3D lung “ image retrieval using localized features,” in SPIE Medical Imaging, 2011.
 - [13] Y. Song, W. Cai, Y. Zhou, and D. D. Feng, “Feature-Based Image Patch Approximation for Lung Tissue Classification,” IEEE Transactions on Medical Imaging, vol. 32, no. 4, pp. 797–808, Apr. 2013.
 - [14] Y. Song, W. Cai, H. Huang, Y. Zhou, D. Feng, Y. Wang, M. Fulham, and M. Chen, “Large Margin Local Estimate with Applications to Medical Image Classification,” IEEE Transactions on Medical Imaging, vol. 34, no. 6, pp. 1362–1377, Jun. 2015.
 - [15] Choy G, Khalilzadeh O, Michalski M, et al. Current applications and future impact of machine learning in radiology. Radiology.
 - [16] Yu, Z., Chen, X. Z., Cui, L. H., Si, H. Z., Lu, H. J., & Liu, S.H. (2014). Prediction of lung cancer based on serum biomarkers by gene expression programming methods. Asian Pacific Journal of Cancer Prevention, 15(21), 9367-9373.
 - [17] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," IEEE Trans Neural Netw, vol. 12, pp. 181- 201, 2001.
 - [18] Zhang, F., Song, Y., Cai, W., Zhou, Y., Fulham, M., Eberl, S., Shan, S., Feng, D.: A ranking-based lung nodule image classification method using unlabeled image knowledge. In: IEEE ISBI. pp. 1356–1359. IEEE (2014)
 - [19] Kumar, D., Wong, A., Clausi, D.A.: Lung nodule classification using deep features in CT images. In: Computer and Robot Vision (CRV), 2015 12th Conference on. pp. 133–138. IEEE (2015)
 - [20] Hussein, S., Cao, K., Song, Q., Bagci, U.: Risk Stratification of Lung Nodules Using 3D CNN-Based Multi-task Learning. In: International Conference on Information Processing in Medical Imaging. pp. 249–260. Springer (2017).