INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

LUNG CANCER DETECTION BASED ON LUNG IMAGES

N. AJAY, P.VAMSHI, Dr.A.PRATAPA REDDY, PROFESSOR

SVS GROUP OF INSTITUTIONS, Hasanparthy, Bheemaram, Hanamkonda, Telangana 506015

ABSTRACT

In this computer era we are totally going with the automation of everything, in the same way the medical industry is also automated with the help of image processing and data analytics .The best way to control the death cause by cancer is early detection. The medical image or a CT scan image is pre-processed .The contrast of the image is increased with the CLAHE Equalization technique .Then it is segmented with the help of random walk segmentation method. In segmentation the three process will happen the ROI of image is segmented and then then the border correction is done. As third part the continuous pixel change is segmented. The classification is the major portion where the cancerous and non-cancerous is identified with the pre trained model. All the methods used above deals with the traditional way of image processing and data analytics. In Future this accuracy will be boosted with the modern XGboost algorithm where less data is used to get high accuracy.

1. INTRODUCTION

Lung cancer growth has turned out to be a standout amongst the most widely recognized reasons for disease in the two people. Countless bite the dust each year because of lung malignancy. The illness has diverse stages whereby it begins from the little tissue and spreads all through the distinctive territories of the lungs by a procedure called metastasis. It is the uncontrolled development of undesirable cells in the lungs. It is assessed that around 12,203 people had lung disease in 2016, 7130 guys and 5073 females; passing from lung malignant growth in 2016 were 8839. . Biomedical image handling is the most recent rising apparatus in medicinal research utilized for the early recognition of malignancies. Biomedical image handling strategies can be utilized in the restorative field to analysis maladies at the beginning time. It utilizes biomedical images, for example, X-beams, Computed innovation and MRIs. The principle commitment of image handling in the restorative field is to analysis the malignant growth at the beginning time, expanding survival rates. The time factor is basic for tumors of the mind, the lungs, and bosoms. image handling can identify these malignant growths in the early periods of the maladies encouraging an early treatment process. The image preparing procedure comprises of four essential stages, pre-handling, division, including extraction and grouping. This paper presents image preparing procedures whereby the CT examine image is utilized as information image, is handled and beginning period lung disease is distinguished utilizing an SVM (bolster vector machine) calculation as a classifier in the grouping stage to improve exactness, affectability, and explicitness. First the image is pre-handled and divided. After that Features are removed from the sectioned image lastly the image is delegated ordinary or destructive. Advanced image handling is the utilization of PC calculations to perform image preparing on computerized images. As a subfield of advanced flag preparing, computerized image handling has numerous points of interest over simple image preparing. It permits a lot more extensive scope of calculations to be connected to the information data — the point of advanced image handling is to improve the image information (Features) by stifling undesirable mutilations as well as upgrade of some vital image includes with the goal that our AIComputer Vision models can profit by this improved information to take a shot at. Feature extraction begins from an underlying arrangement of estimated information and assembles determined qualities (Features) proposed to be useful and non-excess, encouraging the resulting learning and speculation steps, and at times prompting better human elucidations. Feature extraction is a dimensionality decrease process, where an underlying arrangement of crude factors is diminished to progressively sensible gatherings (Features) for handling, while still precisely and totally portraying the first informational collection. At the point when the information to a calculation is too substantial to be in any way handled and it is suspected to be repetitive (for example a similar estimation in the two feet and meters, or the redundancy of images introduced as pixels), at that point it very well may be changed into a decreased arrangement of Features (additionally named a component vector). Deciding a subset of the underlying Features is called include choice. The chose Features are relied upon to contain the pertinent data from the information, with the goal that the ideal undertaking can be performed by utilizing this decreased portrayal rather than the total introductory information. Feature extraction includes lessening the measure of assets required to depict a substantial arrangement of information. When performing examination of complex information one of the serious issues originates from the quantity of factors included. Examination with countless for the most part requires a lot of memory and calculation control, likewise it might make an arrangement calculation overfit to preparing tests and sum up ineffectively to new examples. Feature extraction is a general term for strategies for building mixes of the factors to get around these issues while as yet portraying the information with adequate

ISSN: 2057-5688

exactness. Many AI specialists trust that appropriately streamlined component extraction is the way to successful model development

2. LITERATURE SURVEY

2.1 Image retrieval system based on feature extraction and relevance feedback AUTHORS: D. Harini, & D. Bhaskari Abstract: The availability of huge multimedia databases and the development of information highways have urged many researchers for developing effective methods of retrieval based on their content. The traditional way of searching the available huge collections of multimedia data was by keyword indexing or simply by browsing, where by the user's main interest lies in the maximum retrieval of similar data. Digital image databases however, opened the way to content-based searching and retrieval. A lot of research has been done in retrieving the content based on image features like color, texture, and shape. In this paper an attempt is made to design a methodology for an efficient image retrieval system by extracting low level and high level features from images through relevance feedback. In order to reduce the computational complexity and to achieve efficiency, a two phase approach is adapted. In the first phase color segmentation and GLCM of second order statistics for texture are performed. The second phase takes the feedback obtained from phase1 and involves the usage of wavelets combined with PCA for a refined search and subsequent retrieval of similar images. 2.2 Tumor detection in non stationary backgrounds AUTHORS: R.N. Strickland Abstract: The authors introduce two detectors which they use to locate simulated tumors of fixed size in clinical gamma-ray images. The first method was conceived when it was observed that small tumors possess an identifiable signature in curvature feature space, where "curvature" is the local curvature of the image data when viewed as a relief map. Computed curvature values are mapped to a normalized significance space using a windowed statistic. The resulting test statistic is thresholded at a chosen level of significance to give a positive detection. Nonuniform anatomic background activity is effectively suppressed. The second detector is an adaptive prewhitening matched filter, which uses a form of preprocessing known as statistical scaling to adaptively prewhiten the background. Tests are performed using simulated Gaussian-shaped tumors superimposed on twelve clinical gamma ray images. When the tumors to be detected are small-less than 3 pixels in diameter-the curvature detector out-performs the matched filter in true positive/false positive tests. A mean true positive rate of 95% at one false positive per image is achieved when the local signal-to-noise ratio of the tumor-background is /spl ges/2. At larger tumor sizes the best performance is displayed by a different form of matched filter, namely the statistical correlation function proposed by Pratt (1991).

2.3 A novel computer-aided lung nodule detection system for ct images AUTHORS: M. Tan, R. Deklerck, et al., Abstract: The paper presents a complete computer-aided detection (CAD) system for the detection of lung nodules in computed tomography images. A new mixed feature selection and classification methodology is applied for the first time on a difficult medical image analysis problem. The CAD system was trained and tested on images from the publicly available Lung Image Database Consortium (LIDC) on the National Cancer Institute website. The detection stage of the system consists of a nodule segmentation method based on nodule and vessel enhancement filters and a computed divergence feature to locate the centers of the nodule clusters. In the subsequent classification stage, invariant features, defined on a gauge coordinates system, are used to differentiate between real nodules and some forms of blood vessels that are easily generating false positive detections. The performance of the novel feature-selective classifier based on genetic algorithms and artificial neural networks (ANNs) is compared with that of two other established classifiers, namely, support vector machines (SVMs) and fixed-topology neural networks. A set of 235 randomly selected cases from the LIDC database was used to train the CAD system. The system has been tested on 125 independent cases from the LIDC database. The overall performance of the fixed-topology ANN classifier slightly exceeds that of the other classifiers, provided the number of internal ANN nodes is chosen well. Making educated guesses about the number of internal ANN nodes is not needed in the new feature-selective classifier, and therefore this classifier remains interesting due to its flexibility and adaptability to the complexity of the classification problem to be solved. Our fixed-topology ANN classifier with 11 hidden nodes reaches a detection sensitivity of 87.5% with an average of four false positives per scan, for nodules with diameter greater than or equal to 3 mm. Analysis of the false positive items reveals that a considerable proportion (18%) of them are smaller nodules, less than 3 mm in diameter. A complete CAD system incorporating novel features is presented, and its performance with three separate classifiers is compared and analyzed. The overall performance of our CAD system equipped with any of the three classifiers is well with respect to other methods described in literature.

2.4 Learning orientation invariant contextual features for nodule detection in lung CT scans AUTHORS: J. Bai, X. Huang, et al Abstract: This work combines model-based local shape analysis

ISSN: 2057-5688

and data-driven local contextual feature learning for improved detection of pulmonary nodules in low dose computed tomography (LDCT) chest scans. We reduce orientation-induced appearance variability by performing intensity-weighted principal component analysis (PCA) to estimate the local orientation at each candidate location. Random comparison primitives defined in a local coordinate system are used to describe the local context around a nodule candidate. A random forest is trained to learn and combine a subset of these primitives into discriminative orientation invariant contextual features and classify nodule candidates. Validation using 99 CT scans from the publicly available Lung Image Database Consortium (LIDC) demonstrates the benefit of combining geometric modeling and data-driven machine learning. The proposed method reduces more than 80% of false positives of the baseline model-based method consistently over a wide range of sensitivity levels (70%-90%).

3. PROBLEM STATEMENT

In existing paper, a picture handling procedures has been utilized to recognize beginning time lung malignant growth in CT examine pictures. The CT filter picture is preprepared pursued by division of the ROI of the lung. Discrete waveform Transform is connected for picture pressure and highlights are extricated utilizing a GLCM. The outcomes are encouraged into a SVM classifier to decide whether the lung picture is carcinogenic or not. The SVM classifier is assessed dependent on a LIDC dataset.

3.1 Limitation Of System

1.The CT filter picture is pre-prepared pursued by division of the ROI of the lung. 2.Discrete waveform Transform is connected for picture pressure and highlights are extricated utilizing a GLCM.

4. PROPOSED SYSTEM

The proposed model applies a range of algorithms to the different stages of image processing. In this proposed model, first the CT scan image is pre-processed and the ROI (region of interest) is separated in preparation for segmentation.[17] At the segmentation stage, Discrete Wavelet Transform (DWT) is applied and the feature is extracted by using a GLCM (Gray level co-occurrence matrix) such as correlation, entropy, variance, contrast, dissimilarity and energy. After the feature extraction stage, classification is carried out by an SVM (support vector machine) for classification of cancerous and non-cancerous nodules

4.1 Advantages:

1. The classification is the major portion where the cancerous and non-cancerous is identified with the pre trained model.

5. IMPLEMENTATION

5.1 Data Collection Module:

Purpose: This module is responsible for collecting the dataset required for training the lung cancer detection model. Functionality: It includes a function uploadDataset() that allows the user to select the directory containing the dataset. After selecting the directory, it displays a message confirming that the dataset has been loaded. Implementation: The function utilizes the filedialog module from the tkinter library to open a dialog box for directory selection. Once the directory is selected, it updates the global variable filename and displays a message in the Text widget confirming the dataset loading.

5.2 Data Handling Module:

Purpose: This module handles the preprocessing of the dataset, including loading, splitting into training and testing sets, and performing dimensionality reduction using Principal Component Analysis (PCA). Functionality: It includes a function splitDataset() that loads the dataset stored in the specified directory, splits it into training and testing sets (using an 80-20 split), and performs PCA to reduce the dimensionality of the feature space. Implementation: The function loads the dataset files X.txt.npy and Y.txt.npy, reshapes the features, performs PCA, splits the dataset, and updates the global variables X_train, X_test, y_train, and y_test. It also displays messages in the Text widget regarding the dataset size and split ratios.

5.3 SVM Module:

Purpose: This module implements the Support Vector Machine (SVM) algorithm for classification of lung cancer nodules. Functionality: It includes a function executeSVM() that initializes an SVM classifier, trains it using the training data, makes predictions on the test data, calculates the accuracy, and displays the SVM accuracy in the Text widget. Implementation: The function creates an SVM classifier using svm.SVC(), fits it to the training data, predicts on the test data, calculates the accuracy using accuracy_score, and updates the global variables classifier and svm_acc. It displays the SVM accuracy in the Text widget.

5.4 K-Means Module:

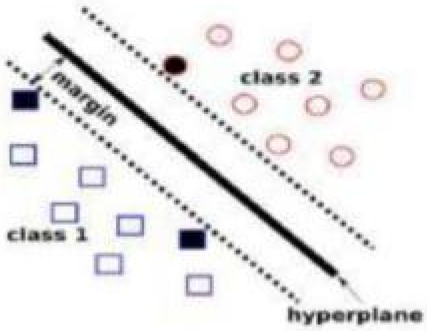
Purpose: This module implements the K-Means clustering algorithm for classification of lung cancer nodules. Functionality: It includes a function executeKmeans() that initializes a K-Means clustering model, fits it to the training data, makes predictions on the test data, calculates the accuracy, and displays the K-Means accuracy in the Text widget. Implementation: The function creates a K-Means model using KMeans(), fits it to the training data, predicts on the test data, calculates the accuracy using accuracy_score, and updates the global variable kmeans_acc. It displays the K-Means accuracy in the Text widget.

5.5 Deployment Module:

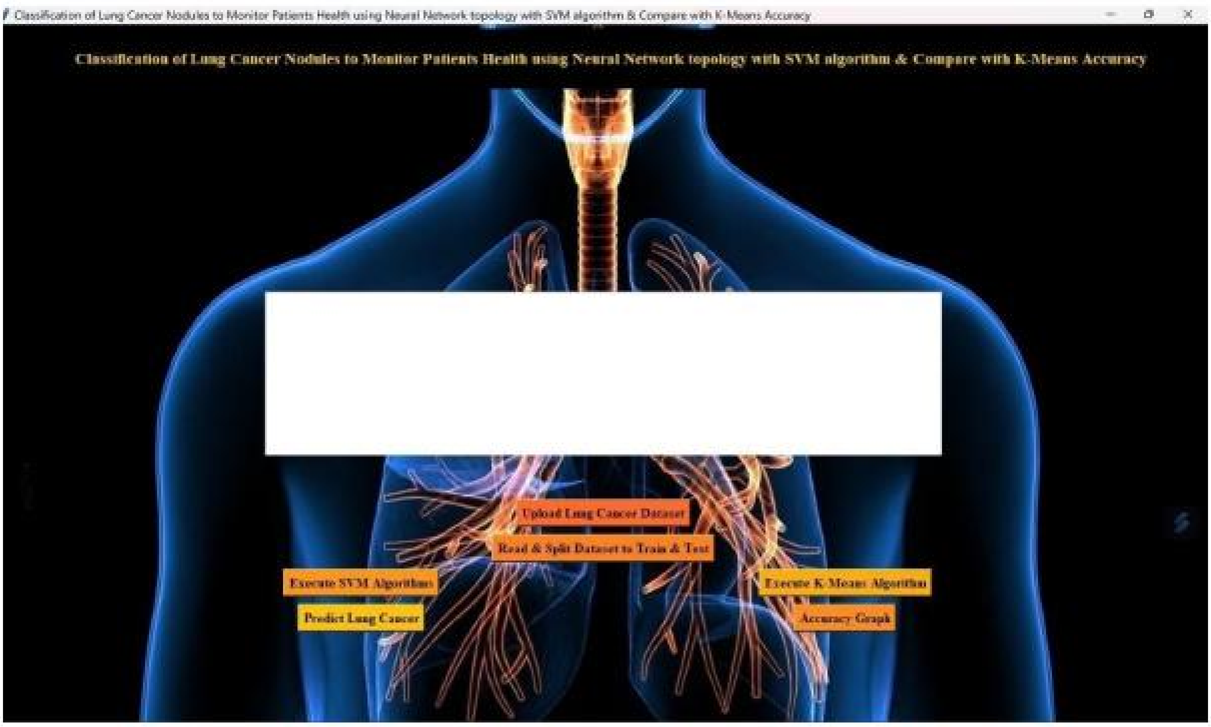
Purpose: This module handles the deployment of the lung cancer detection system, including predicting lung cancer on new images and visualizing accuracy. Functionality: It includes functions for predicting lung cancer on new images (predictCancer()), displaying the accuracy graph (graph()), and configuring the GUI interface (tkinter). Implementation: The deployment module integrates all the functionalities of data collection, data handling, SVM, and K-Means modules into a graphical user interface (GUI) using tkinter. It provides buttons for uploading the dataset, splitting it, executing SVM and K-Means algorithms, predicting lung cancer on new images, and visualizing accuracy. The accuracy results are displayed in a Text widget, and the GUI layout is designed for user interaction..

6. ARCHITECTURE DIAGRAM

ISSN: 2057-5688



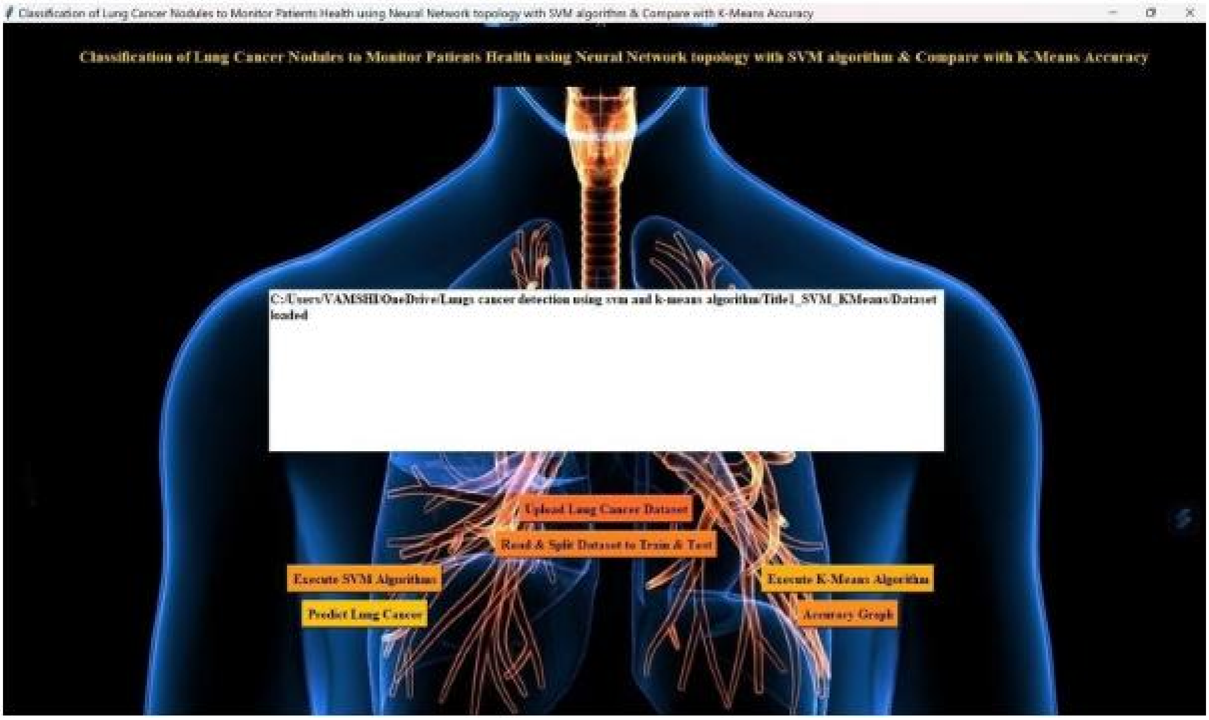
7. EXPECTED RESULTS



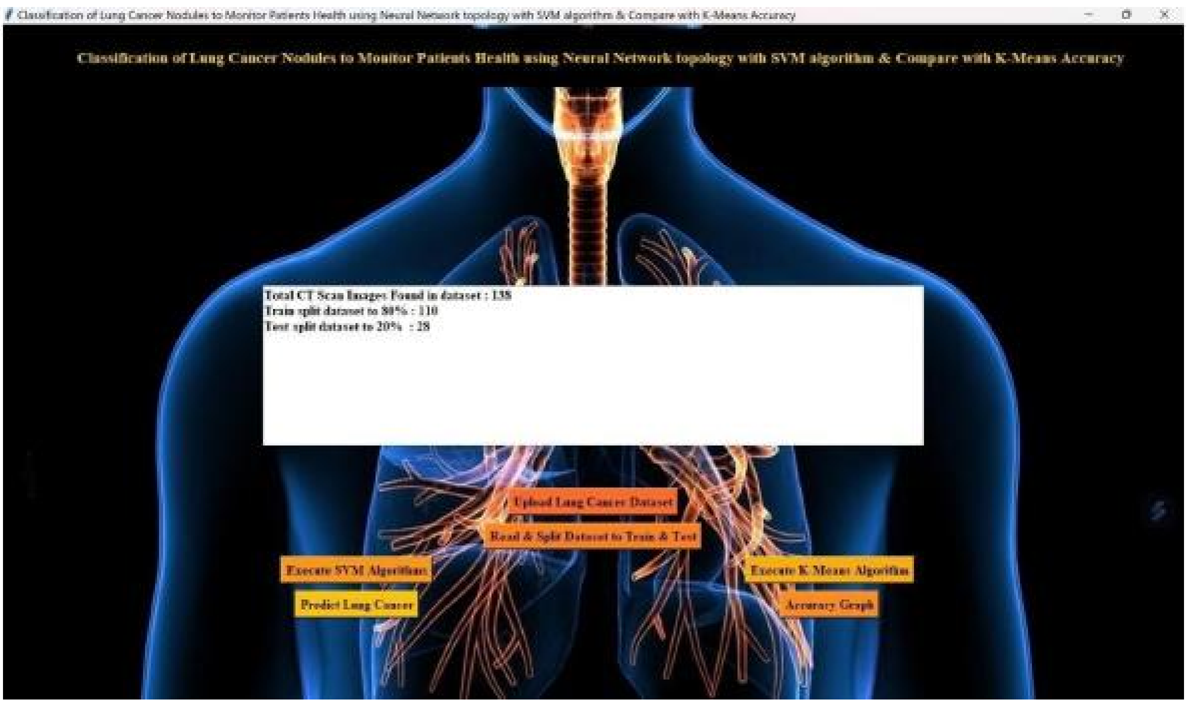
In above screen click on ‘Upload Lung Cancer Dataset’ button and then upload dataset folder

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

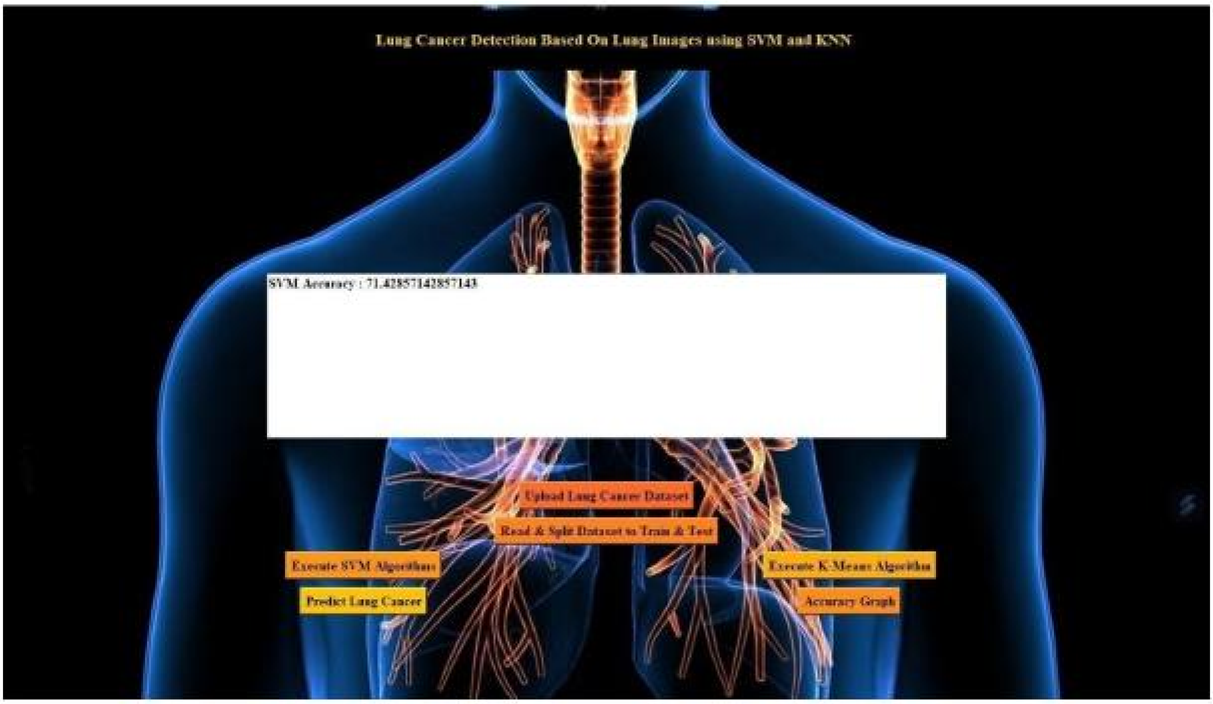


In above screen dataset loaded and now click on 'Read & Split Dataset to Train & Test' button to split dataset into train and test parts and application split 80% dataset for training and 20% dataset to test trained model



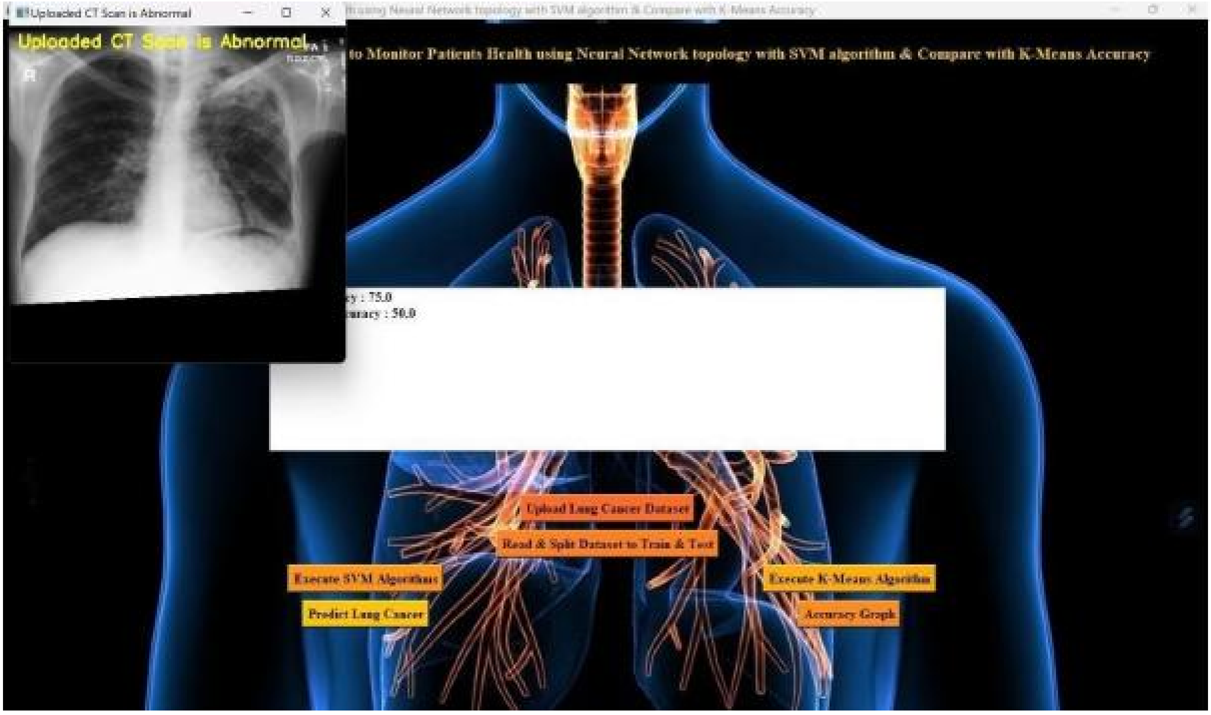
ISSN: 2057-5688

In above screen we can see dataset contains total 138 images and then application using 110 images for training and 28 images for testing and now data is ready and now click on 'Execute SVM Algorithm' button to run SVM on loaded dataset and to get below accuracy



In above screen SVM accuracy is 60% and now click on "Execute K-Means Algorithm" button to run KMEANS algorithm on loaded dataset and to get below screen

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688



In above screen uploaded image predicted as Abnormal and now test with another image

8. CONCLUSION

In the principal period of the venture the Region of Interest in a picture distinguished. The Identified district is situated in an item. The highlights in the picture are distinguished by utilizing some picture handling system. In second period of the task the component removed information is then used to arrange the picture is destructive or not utilizing a portion of the SVM – bolster vector machine grouping. At that point some boostin calculation is utilized to expand the exactness of the instrument.In the principal period of the venture the Region of Interest in a picture is distinguished. The Identified \district is situated in an item. The highlights in the picture are distinguished by utilizing some picture handling system. In second period of the task the component removed information is then used to arrange the picture is destructive or not utilizing a portion of the SVM – bolster vector machine grouping. At that point some boosting calculation is utilized to expand the exactness of the instrument.In the principal period of the venture the Region of Interest in a picture is distinguished. The Identified district is situated in an item. The highlights in the picture are distinguished by utilizing some picture handling system. In second period of the task the component removed information is then used to arrange the picture is destructive or not utilizing a portion of the

SVM – bolster vector machine grouping. At that point some boosting calculation is utilized to expand the exactness of the instrument.

9. REFERENCES

1. M. Debois, "TxNxM1 : the anatomy and clinics of metastatic cancer.Boston," Kluwer Academic Publishers, 2006.

2. J.M. Fitzpatrick, & M. Sonka, "Medical imaging 2003: image processing. USA: SPIE Society of PhotoOptical Instrumentation Engineers," 2003.

3.C. M. Haskell, & J.S. Berek, " Cancer treatment. Philadelphia: W.B. Saunders, 2001

4.K.T.Manivannan, "Development of gray level co-occurrence matrix based support vector machines for particulate matter characterization,"Retrieved fromhttp://rave.ohiolink.edu/etdc/view?acc_num=toledo1341577486, 2012.

5.Sathish Kumar R, R. Logeswari, N. Anitha Devi, S. DivyaBharathy Efficient Clustering using ECATCH Algorithm to Extend Network Lifetime in Wireless Sensor Networks. International Journal of Engineering Trends and Technology. 45. 476-481.

6.B. H. Boyle, "Support vector machines : data analysis, machine learning, and applications," Retrievedhttp://public.eblib.com/choice/publicfullrecord.aspx?p=3021500, 2011.

7.A.A. Abdulla, S.M. Shaharum, Lung cancer cell classification method using artificial neural network, Information Engineering Letters 2 (March) (2012) 50–58.

8.Sathish Kumar R , Nive tha M, Madhu mita G & Santhoshy P. Image Enhancement using NHSI Model Employed in Color Retinal Images. International Journal of Engineering Trends and Technology. 58. 14-19. 10.14445/22315381/IJETT-V58P203- April 2018.

9.D. Harini, & D. Bhaskari, "Image retrieval system based on feature extraction and relevance feedback," ACM, 2 Penn Plaza, Suite 701, New York,NY 10121-0701, 2012.