



A NOVEL LUNG CANCER DETECTION APPROACH USING 3D CNN

¹M.Chandana, ²Dr. Punyaban Patel,

¹B. Tech Student, Dept. of CSE, CMR Technical Campus, Medchal, Hyderabad, Telangana, India, munagalachandana2002@gmail.com

²Professor, Dept. of CSE, CMR Technical Campus, Medchal, Hyderabad, Telangana, India, punyaban@gmail.com

ABSTRACT: Lung cancer is one of the most common and serious diseases present around the world, which is observed in people of all age groups ranging from children to old people. Annually it costs a lot of money for the cure and diagnosis of people with lung cancer. The existing clinical techniques such as X-Ray and other imaging procedures require complex hardware and considerable expense. Thus the most important issue is the prediction to be accurate and to use a reliable method for that. The best way to control the death cause by cancer is early detection. Many diagnosis and detection of lungs cancer has been done using various data analysis and classification techniques. Since the cause of lung cancer stay obscure, prevention become impossible, thus early detection of tumor in lungs is the only way to cure lung cancer. Hence in this work, , A Novel Lung Cancer Detection Approach Using 3D CNN is presented. In this approach, Lung Image Database Consortium (LIDC) is used to train the models. Convolutional Neural Networks (CNN) models will be trained and ready to be used for lung cancer detection. Compared to earlier approaches this approach will obtain better result for lung cancer detection.

KEYWORDS: Lung Cancer, Tumor, Deep Learning, 3-D (Three Dimensional), CNN (Convolutional Neural Network).

I. INTRODUCTION

Lung diseases are a major concern nowadays. The human body consists of diverse fundamental organs, out of which Lung is one of them. The Lungs are made up of two sections, right lung and left lung and the main function of lungs is to facilitate exchange of gas which is called as breathing or respiration. Lung disease is one of the most common disease that is

affected in our early stage to improve the rate of patients survival. For the radiologist the diagnosis of cancer is the most challenging part [1]. Lung disease causes several risk factors such as age, sex diabetes, high blood pressure, cholesterol, abnormal pulse rate and other factors. Lung cancer is the most common type of cancer in India, along with prostate, mouth, breast cancer. Due to smoking, excessive increases in pollution and the inhalation of cancerous elements cause cancer in men and women. Still, the amount of cancer in men is more compared to women. Different imaging modalities are used for diagnoses, such as CT, MRI, PET, and X-Ray. Oncologists and physicians diagnose cancer by seeing gray-scale saturation in the image. It requires experience in both anatomy and physiology of humans with imaging technique knowledge. It takes too much time for diagnosing, and accuracy is less due to a lack of imaging expertise [4].

The nature of lung disease is very complex and hence, the disease must be handled in a proper way. The opinion of lung cancer in medical science and data mining are used for discovering various metabolic syndromes. Data mining with classification and clustering plays an important role in the prediction of lung disease and data investigation. Many problems damages the lung and even causes premature death. The people who are having the highest risk of getting



cancer because of smoking. Lung cancer is divided into two types, namely non-small cell lung cancers (NSCLC) and small cell lung cancers (SCLC). It also kills more people who have affected by breast cancer, colon, prostate or ovarian. Lung cancer when their age increases to 60. Lung cancer is the growth of unusual cells in our lungs. These lung tissues divide rapidly and form tumors [2].

Cancer detection is done with the aid of the led expert docs and earlier tiers it may helpful. The opportunity for human error must be there. It produces the probability of error in lung cancer detection which necessitates an automatic manner [3]. In many cases, the lung cancer diagnosis is based on doctors' experience, which might lead to misdiagnosis and cause medical issues in patients. There have been numerous strategies and methods for predicting level of cancer malignancy using deep learning and machine learning methods. Lung cancer cannot be prevented but its risk can be reduced. So detection of lung cancer at the earliest is crucial for the survival rate of patients. The number of chain- smokers is directly proportional to the number of people affected with lung cancer.

Many techniques in data mining and neural networks have been taken to bring out the of lung disease. Machine learning - a field of artificial intelligence, works on programmed algorithms by using their past learnt experience to draw new conclusions with better accuracy. There are many learning techniques such as classification, regression, association etc. that can be applied in the applications as per the need to meet the best prediction accuracy as close as possible to the human predictions.

The choices of algorithms are based on the type of data being operated. Lot of inbuilt

libraries are supported in the ML tools and scripting languages for realizing the proposals. Machine Learning methods are currently being used in medical fields for identification and classification of various diseases. Recently Machine Learning methods have been employed in this area in predicting the cancer in early stages, which can result in better chances of survival of patients The algorithms used for this disease is classified based on various methods like K- Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Naive Bayes(NB), Random forest, Logistic Regression, Support Vector Machine(SVM) [6].

The prediction accuracy of any model depends on its learning capability. The learning ability of any algorithm depends on correctness of the training set. However still there is a need to improve the accuracy to improve the mortality rate. On the other hand, deep learning algorithms are achieving prominent results in various disease predictions. Hence in this work, A Novel Lung Cancer Detection Approach Using 3D CNN is presented. The rest of the work is organized as follows: The section II describes the literature survey. The section III presents A Novel Lung Cancer Detection Approach Using 3D CNN. The section IV evaluates the result analysis. The conclusion is provided in section V.

II. LITERATURE SURVEY

Puneet, Anamika Chauhan et. al., [7] presents Detection of Lung Cancer using Machine Learning Techniques Based on Routine Blood Indices. The purpose of this research is to predict lung cancer using machine learning techniques based on routine blood indices and have better predictive results. We have used machine learning techniques such as XGBoost, Logistic Regression, SVM, Gaussian Naive Bayes, Decision Tree and KNN



with GridSearchCV. Various scikit-learn algorithms have been used for feature selection, and only those features are selected that are important for our model. We found that XGBoost using GridSearchCV classification model is most suited for this work.

Radhika P R, Rakhi.A.S.Nair, Veena G et. al., [8] presents A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms. The lung cancer prediction was analysed using classification algorithms such as Naive Bayes, SVM, Decision tree and Logistic Regression. The key objective of this paper is the early diagnosis of lung cancer by examining the performance of classification algorithms. The predictive performance of classifiers is compared quantitatively. In the performance chart, different results are produced for each classifier on the lung cancer dataset. Looking at the correct classification (CA) and other metrics; the best result is given by the support vector machine algorithm. SVM algorithm used high dimension to classify the observation so it's performance is the best. More accurate lung cancer detection can be done using this technique.

Dendi Gayathri Reddy, Emmidi Naga Hemanth Kumar, Desireddy Lohith Sai Charan Reddy, Monika P et. al., [9] presents Integrated Machine Learning Model for Prediction of Lung Cancer Stages from Textual data using Ensemble Method. This paper conjectures the oeuvre model which is efficient in predicting the stages of lung carcinoma by applying the concepts of ML algorithms. The proposed model is combination of K-Nearest Neighbours, Decision Tree and Neural Networks models along with bagging ensemble method for enhancing the accuracy of the overall prediction. The predicted results of

the suggested model are showing better accuracy compared to individual algorithms.

Wasudeo Rahane, Himali Dalvi, Yamini Magar, Anjali Kalane, Satyajeet Jondhale et. al., [10] describes Lung Cancer Detection Using Image Processing and Machine Learning HealthCare lung cancer detection system using image processing and machine learning is used to classify the presence of lung cancer in a CT-images and blood samples. In spite of CT scan reports are more effective than Mammography, therefore patient CT scan images are categorized in normal and abnormal. The abnormal images are subjected to segmentation to focus on tumor portion. Classification done on features extracted from the images. The efficient method to detect the lung cancer and its stages successfully and also aim to have more accurate results by using SVM and Image Processing techniques.

Qing Wu and Wenbing Zhao et. al., [11] describes Small-Cell Lung Cancer Detection Using a Supervised Machine Learning Algorithm. They present a novel neural-network based algorithm, which we refer to as nentropy degradation method (EDM), to detect small cell lung cancer (SCLC) from computed tomography (CT) images. This research could facilitate early detection of lung cancers. The training data and testing data are high-resolution lung CT scans provided by the National Cancer Institute. We selected 12 lung CT scans from the library, 6 of which are for healthy lungs, and the remaining 6 are scans from patients with SCLC. We randomly take 5 scans from each group to train our model, and used the remaining two scans to test. This algorithm achieves an accuracy of 77.8%.

III. LUNG CANCER DETECTION APPROACH USING 3D CNN

In this section, A Novel Lung Cancer Detection Approach Using 3D CNN is presented. The architecture of presented approach is shown in Fig. 1. For this purpose, a web application is developed. Lung cancer data-set which is collected from Lung Image Database Consortium (LIDC) which was publicly available on the National Cancer Institute is used. Load the data-set, perform Image pre processing, select Region of Interest (ROI) and split the data set into Train and Test sets. Ratio of Train to Test is not predefined. We should choose the ratio to get the best possible accuracy of the algorithm. We are splitting the data set into Train and Test in the ratio of 8:2, as it gives better accuracy.

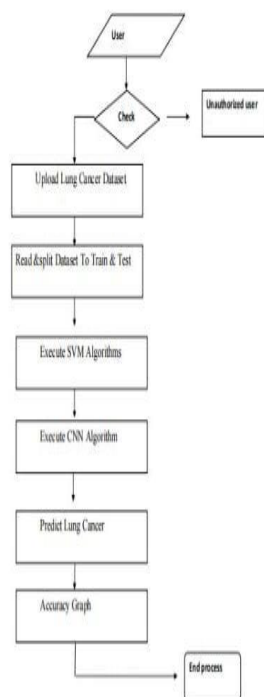


Fig. 1: The architecture of presented approach

As compared to X-ray and MRI images CT scan replica have less noise therefore CT scan image is taken as input. To get

better accuracy and less distortion, these images are used. The data is encoded or transformed into a particular form which is then made feasible for the further analysis of the machine. It allows the different algorithms to interpret the features of the data easily. Data preprocessing in Machine Learning refers to the technique of preparing (cleaning and organizing) the raw data to make it suitable for a building and training Machine Learning models. Pre-processing helps to improve the accuracy, reduce the time and resources required to train the model, prevent overfitting, and improve the interpretability of the model.

SVM Algorithm Execution: This stage entails determining whether the cancer is malignant or benign. The support vector machine (SVM) classifier is used in this case. SVM classifiers are supervised learning models that identify patterns. It is a supervised learning model that can be used for both classification and regression analysis. SVM helps in reducing the rate of misclassification and thus gives good results. It plays out a few emphasess to locate an ideal hyperplane that best isolates the various classes of a dataset in a multidimensional space. The hyperplane is framed by the support vectors that are nearest to the edge(margin). The one which is framed with the greatest edge is supposed to be the ideal hyperplane, that is, maximum marginal hyperplane(MMH). The output is then anticipated by plotting the test set data points on the obtained hyperplane. When our data has exactly two classes, we can use support vector machines. The best hyperplane for classifying the data is discovered; it generates two classes with different data points. As a result, we can find the best hyperplane classifier with the greatest margin between the two classes.

A Convolutional Neural Network (CNN) is a Deep Learning algorithm that can take an input image, assign importance (learnable weights and biases) to different aspects/objects in the image, and distinguish one from the other. CNN is influenced by the biological process, where the animal optical cortex system representation is used to display the connectivity pattern between the neuron elements. This algorithm is appropriate for a classifying multi-class problem and binary classification (for example - predicting whether a medical image has a malignant tumor or not). It comprises of a convolutional layer, activation, pooling, and output layer.

The CNN changes the pixel values of input CT image through weighing in convolutional layers and sub-sampling layers are sampled alternatively. The outcome of this algorithm is a recursive function with weighted input values. CNN in the medical area is employed auspiciously to aid disease diagnosis like breast cancer, lung cancer, cardiovascular diseases, neurological disorders, etc. Based on pathological and laboratory evidence, we designed a three-dimensional convolutional neural network (CNN) to detect pulmonary nodules and classify them as malignant or benign diseases.

Both Support Vector Machine (SVM) and Convolutional Neural Networks (CNN) models are trained and ready to be used for lung cancer detection. We then compare both algorithms accuracy and choose the most efficient algorithm to predict lung cancer.

IV. RESULT ANALYSIS

In this section, A Novel Lung Cancer Detection Approach Using 3D CNN is implemented. The result analysis of presented approach is discussed here. The training set is taken from the Lung

Image Database Consortium and includes 140 labelled CT images. The data is split into 8:2 for Train and Test. Fig. 2 shows the dataset uploading process.

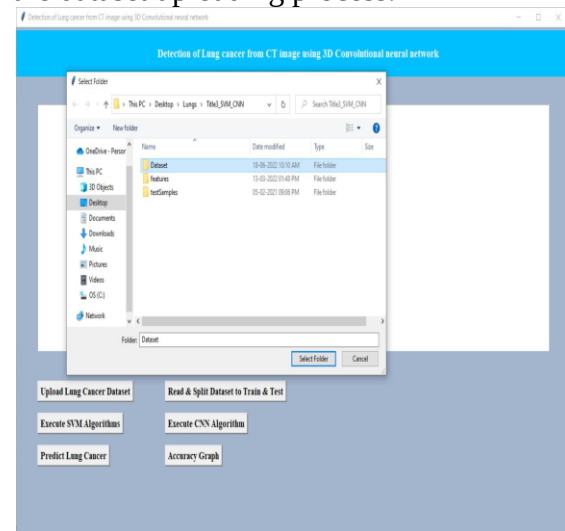


Fig. 2: Dataset Uploading Process

The Fig. 3 shows the loaded dataset.

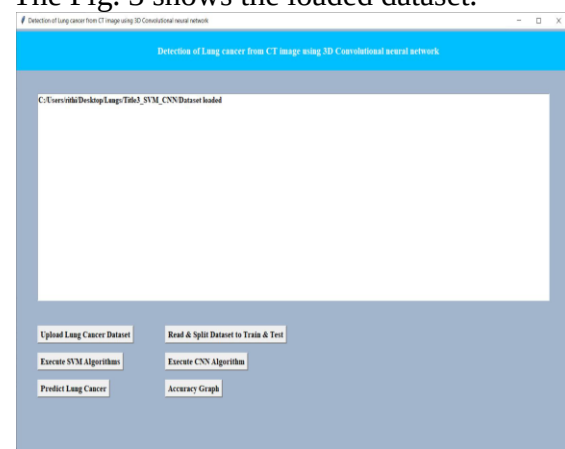


Fig. 3: Loaded Dataset

The fig. 4 shows the dataset splitting process.

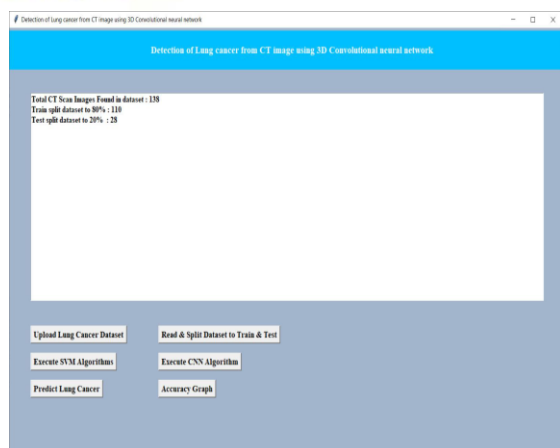


Fig. 4: Dataset Splitting

The Fig. 5 shows training process of SVM.

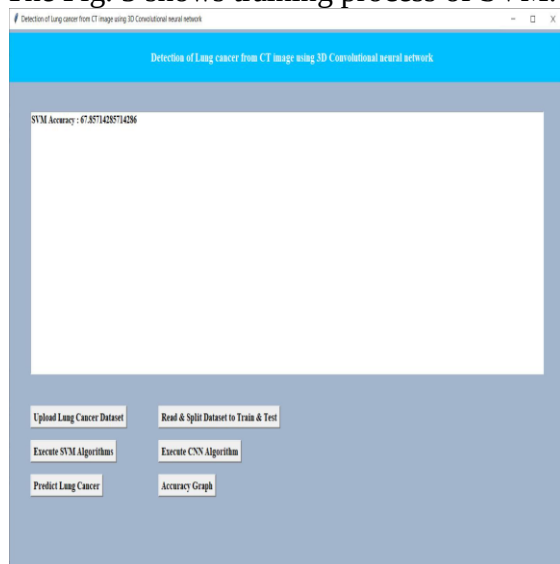


Fig. 5: Training SVM Model

The Fig. 6 shows the Training CNN model.

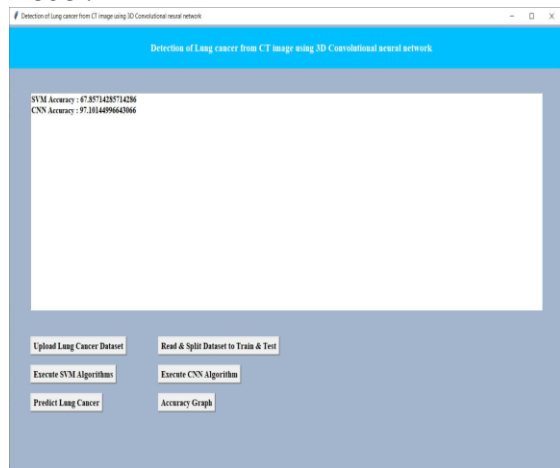


Fig. 6: Training CNN Model

The Fig. 7 shows the process of uploading the images for lung cancer detection.

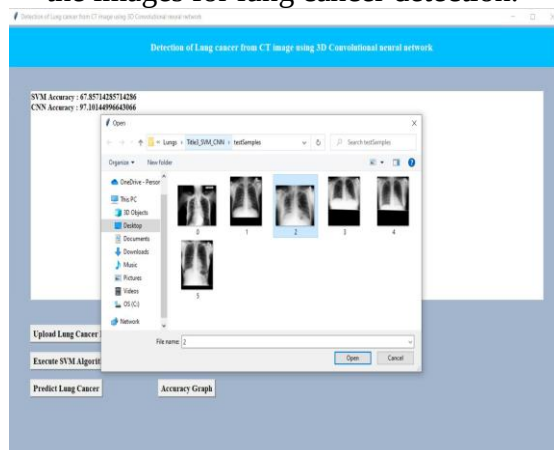


Fig. 7: Images uploading for lung cancer detection

The Fig. 8 shows the results of uploaded image.

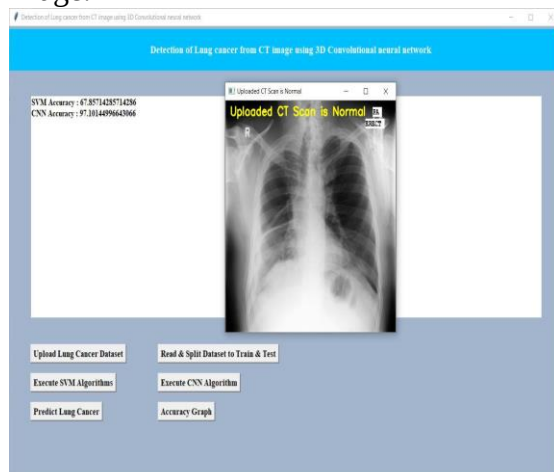


Fig. 8: Results of uploaded image

The Fig. 9 shows the accuracy comparison

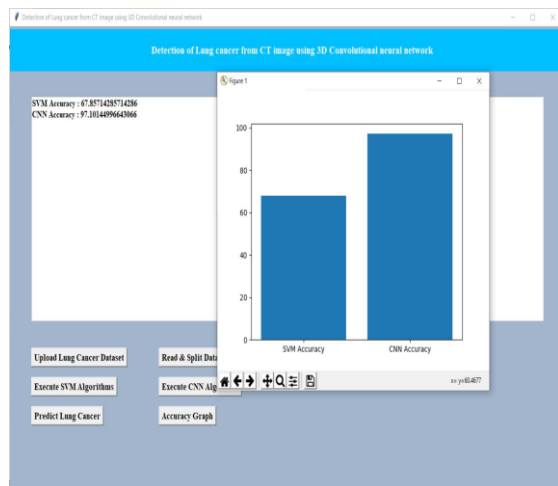


Fig. 9: Accuracy Comparison

Compared to SVM algorithm, CNN has high accuracy. This system takes 1 second to predict the output using SVM with 60% accuracy and 6 seconds to predict the output using CNN with 96% accuracy. As the accuracy of CNN is higher, we are using CNN algorithm for Lung Cancer Detection.

V. CONCLUSION

Lungs cancer is a serious disease that described by unlimited growth of cells in tissues of the lungs. Detection of lungs cancer in earlier stages is very crucial it can save many lives. In this work, A Novel Lung Cancer Detection Approach Using 3D CNN is presented. Machine learning and deep learning algorithms are being extensively used in lot of applications these days. This model uses Support Vector Machine and Three Dimensional Convolutional Neural Network for lung cancer disease detection. The training set is taken from the Lung Image Database Consortium and includes 140 labelled CT images. The data is split into 8:2 for Train and Test. This approach requires 1 second to predict the output using SVM with 60% accuracy and 6 seconds to predict the output using CNN with 96% accuracy. This technique will help doctors to detect lung cancer and may be confirmed by

going through diagnostic tests which can help in saving the precious life of a human being.

VI. ACKNOWLEDGEMENT

We thank CMR Technical Campus for supporting this paper titled “A NOVEL LUNG CANCER DETECTION APPRAOCH USING 3D CNN”, which provided good facilities and support to accomplish our work. Sincerely thank our Chairman, Director, Deans, Head Of the Department, Department Of Computer Science and Engineering, Guide and Teaching and Non- Teaching faculty members for giving valuable suggestions and guidance in every aspect of our work

VII. REFERENCES

- [1] Bharathy S, Pavithra R, Akshaya. B, “Lung Cancer Detection using Machine Learning”, 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), DOI: 10.1109/ICAAIC53929.2022.9793061
- [2] Suchismita Bandopadhyay, Anuradha C. Phadke, “Machine learning application for Lung Cancer Detection”, 2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA), DOI: 10.1109/ICCUBEA54992.2022.10010949
- [3] Shubhada Agarwal, Sanjeev Thakur, Alka Chaudhary, “Prediction of Lung Cancer Using Machine Learning Techniques and their Comparative Analysis”, 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), DOI: 10.1109/ICRITO56286.2022.9965052
- [4] Kshitija Ingle, Uttam Chaskar, Shashikant Rathod, “Lung Cancer Types Prediction Using Machine Learning Approach”, 2021 IEEE



International Conference on Electronics, Computing and Communication Technologies (CONECCT),

DOI: 10.1109/CONECCT52877.2021.9622568

[5] K.Karthick, S.Rajkumar, 3N.Selvanathan, U.K.Balaji Saravanan, M.Murali, B.Dhiyanesh, "Analysis Of Lung Cancer Detection Based On The Machine Learning Algorithm And IoT", 2021 6th International Conference on Communication and Electronics Systems (ICCES), 978-1-6654-3587-1/21, 2021 IEEE, DOI: 10.1109/ICCES51350.2021.9489084

[6] Chinmayi Thallam, Aarsha Peruboyina, Sagi Sai Tejasvi Raju, Nalini Sampath, "Early Stage Lung Cancer Prediction Using Various Machine Learning Techniques", 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), DOI: 10.1109/ICECA49313.2020.9297576

[7] Puneet, Anamika Chauhan, "Detection of Lung Cancer using Machine Learning Techniques Based on Routine Blood Indices", 2020 IEEE International Conference for Innovation in Technology (INOCON) Bengaluru, India. Nov 6-8, 2020

[8] Radhika P R, Rakhi.A.S.Nair, Veena G, "A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms", 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), DOI: 10.1109/ICECCT.2019.8869001

[9] DendiGayathri Reddy, Emmidi Naga Hemanth Kumar, DesireddyLohithSaiCharan Reddy, Monika P, "Integrated Machine Learning Model for Prediction of Lung Cancer Stages from Textual data using Ensemble Method", 2019 1st International Conference on Advances in Information Technology, 978-1-7281-3241-9/19

[10] Wasudeo Rahane, Himali Dalvi, Yamini Magar, Anjali Kalane, Satyajeet Jondhale, "Lung Cancer Detection Using Image Processing and Machine Learning HealthCare", Proceeding of 2018 IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India, 978-1-5386-3702-9/18

[11] Qing Wu and Wenbing Zhao, "Small-Cell Lung Cancer Detection Using a Supervised Machine Learning Algorithm", 2017 International Symposium on Computer Science and Intelligent Controls, DOI 10.1109/ISCSIC.2017.22