JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Using an Ensemble Machine Learning Algorithm to Detect Lung Cancer

[1]Ms. Yogita Gupta, [2]Ch. Yashwanth Kumar, [3]P. Karthik Reddy, [4]B. Narayan Singh

[1]Asst. Professor, Department of CSE (AI&ML), CMR College of Engineering & Technology, Hyderabad, Telangana

[2,3,4] UG Students, Department of CSE (AI&ML), CMR College of Engineering & Technology, Hyderabad, Telangana

ABSTRACT:

An aberrant proliferation of malignant cells in the respiratory tract of a person causes lung cancer, a deadly hereditary illness. Cancer of the lung can have major consequences because the lungs are one of the most important organs in the human body. We have concentrated on early identification of lung cancer in this work to benefit both physicians and patients. Histopathology pictures and other diagnostic instruments can be used to identify lung cancer. The suggested study uses histopathology images and CT scans to identify lung cancer by combining a hybridised model of neural networks using convolution with a group of machine learning algorithms, including a Support Vector Classifier, KNN Algorithm, Decision Tree, and Adaboost. 96.12 percent accuracy was attained overall with this job.

INTRODUCTION

With cancer mortality rates rising daily, it is a significant global health concern. Lung cancer is the most prevalent and lethal type of cancer that affects both men and women. Lung cancer, sometimes referred to as carcinoma, is the development of cancerous nodules, or malignant lung tumours, because of unchecked cell growth in the lung tissues. The two main risk factors for developing malignant lung nodules are tobacco consumption and smoking. During a period of around 5–6 years, the combined survival percentage of patients with lung cancer in all stages is low—roughly 14%. Lung cancer is primarily problematic because most cases are discovered after the disease has progressed, making treatment more difficult and drastically decreasing the chances of survival. Because of this, early detection of lung cancer can reduce mortality by up to 70% and boost survival chances by giving patients access to the necessary quick treatment. Based on the features of the cells, there are two primary categories of lung cancer: small cell lung cancer and non-small cell lung cancer. About 80–

85% of all cases of lung cancer are non-small cell lung cancer, while small cell lung cancer accounts for 15-20% of instances. Tumour size and the extent of malignancy in the lungs determine the stage of lung cancer. In terms of severity, lung cancer is primarily divided into four stages: Stage I lung cancer is limited to the lung; Stages II and III lung cancer is limited to the chest; and Stage IV lung cancer is limited to the chest and has spread to other areas of the body. Chest X-rays, Computed Tomography (CT), nuclear medicine (MRI), PET (positron emission tomography), and MRI, or magnetic resonance imaging, can all be used to diagnose lung cancer. Most people prefer CT scan images over those from other types of imaging because they are less distorted, more dependable, and have superior clarity. The process of visually interpreting a database is laborious, time-consuming, and extremely individualistic. This increases the likelihood of human error and may cause cancer to be misclassified. In order to assist the radiologist in correctly diagnosing lung cancer, an automated method is therefore crucial. The pre-processing, feature extraction, lung segmentation, dataset gathering, and classification are all part of the technique designed for this system.

RELATED WORK

Lung Cancer Detection Using Image Processing Techniques

In order to improve images during the initial stages of diagnosis and treatment, image processing techniques are becoming more popularly used in the medical field. Here, time is particularly important in identifying any abnormalities present in target images, specifically in cases of cancer tumours like breast and lung cancer. The two main parts of this research are photo quality and precision checking. The evaluation and development of image quality depend on the advancement level, where low-preprocessing methods are used based on the Gabor filter inside Gaussian rules. An upgraded region of the object of interest is attained by sticking to the segmentation principles, which function as the main grounds for feature extraction. A comparison of normality is performed, depending on common features. Pixel percentage and mask-labelling are the main detected features in this study for precise image comparison.

Using Image Processing Techniques to Identify Various Stages of Lung Cancer in CT-Scan Images

Lung cancer, both small cell and non-small cell varieties, is the most common cancer in all genders. Early detection of lung cancer boosted the survival rate of patients. The most complicated issue in lung cancer prediction is that most cancer cells overlap with one another due to their structural makeup. The medical field has recently made substantial use of image processing methods for early-stage identification and treatment, where time is of the essence. Patients with lung cancer whose cancer cells are discovered early have an overall 5-year survival rate that rises from 14 to 49%. With an eye towards accuracy, the authors of this study have presented a system for detecting lung cancer employing image processing techniques to classify the existence of cancer cells in the lung and their stages from CT-scan pictures using a variety of enhancement and segmentation approaches.

Methods for analysing CT scan data through image processing in order to detect lung cancer early.

The use of image processing methods to analyse CT scan pictures that show lung cancer cells has become more popular in recent years. Thus, it is interesting to talk about how a Computer-Aided Diagnosis (CAD) system that uses Computed Tomography (CT) pictures can aid in the early detection of lung cancer by differentiating between benign and malignant tumours. We go over and investigate the purpose and architecture of a CAD-CT image processing model for cancer detection.

A Review of CT Image Processing Techniques for Lung Cancer Detection

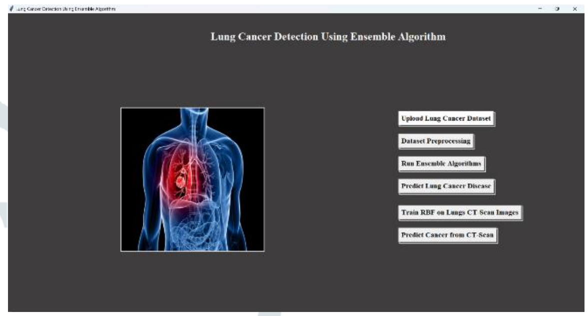
In summary: Among all cancers, lung cancer is the most common cause of death, and the most effective way to image lung cancer is with a CT scan. Many studies on the CAD system for CT image-based lung cancer diagnosis have been conducted in the past. It is broken down into four phases. These comprise nodule identification, segmentation, and classification; they also involve lung segmentation or preprocessing. In-depth reviews of the literature on numerous methods for nodule segmentation, classification, and pre-processing are presented in this work.

METHODOLOGY

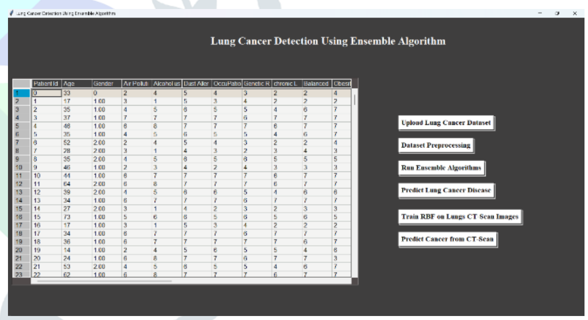
In this project, we are using Lung Cancer dataset to train the Ensemble algorithm by combining AdaBoost, KNN, SVM and Decision Tree and then after training, when we upload test data then this algorithm will predict lung cancer stage as HIGH, LOW and MEDIUM and if HIGH is detected, then application will ask the user to go for CT-SCAN.

Here we designed another algorithm using RBF and lung cancer CT-SCAN images and this CT-SCAN images will be trained using RBF algorithm and then after training when user upload test image then application will predict whether uploaded CT-SCAN is normal or abnormal.

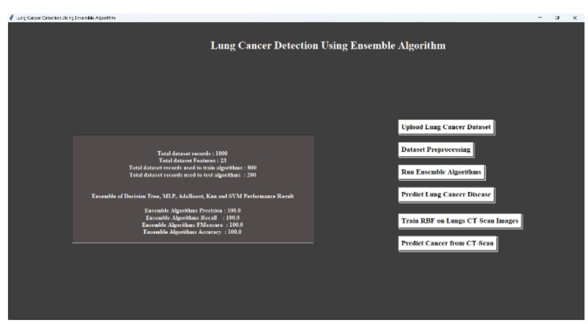
RESULT AND DISCUSSION



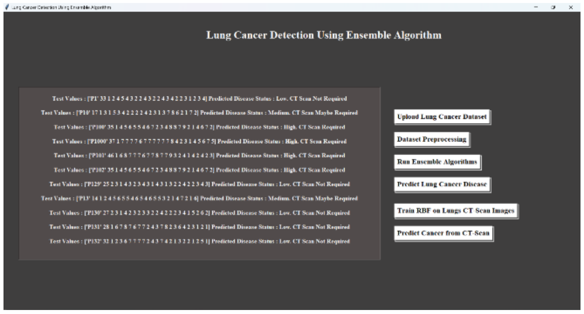
In above screen click on 'Upload Lung Cancer Dataset' button to upload dataset



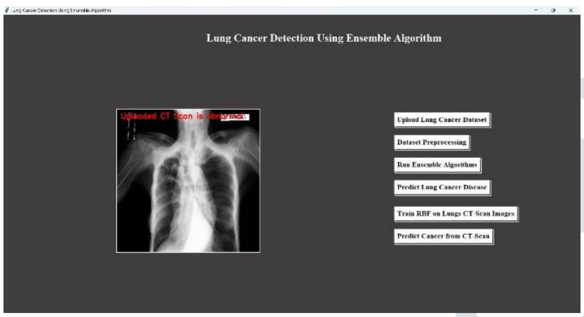
In the above screen we have done 'Dataset Preprocessing' where the processed data is shown.



In the above screen we have to click on 'Run Ensemble Algorithms.'



Followed by the previous step now we click on ‘Predict Lung Cancer Disease’ and move to next step to run RBF.



In above screen in the image title bar or you can see prediction result as ‘ABNORMAL’ indicating in RED colour and now assess with another image.

CONCLUSION

Using the Lung Cancer dataset, our ensemble algorithm is trained in this project using a combination of AdaBoost, Multilayer Perceptron, and Decision Tree. Once trained, the algorithm predicts the stage of lung cancer as HIGH, LOW, or MEDIUM, and if a HIGH stage is detected, the application will prompt the user to undergo a CT scan. This time, we created a new algorithm that makes use of RBF and CT-SCAN images from lung cancer patients. The CT-SCAN images are trained using the RBF method, and once the user uploads a test image, the program determines if the uploaded CT-SCAN was normal or abnormal.

REFERENCES

1. S. R. Nayak and K. Jena, "MRI Brain Tumour Image Analysis Using Fuzzy Rule based Approach", vol. 50, no. 98, pp. 112.

2. L. Mann and S. R. Nayak, "Recent Advances on Mammogram Imaging for Breast Cancer Analysis: A Technological Review", Compute. Intel. Pattern Recognition. Adv. Intel. Syst. Compute. vol 1349, 2021, [online] Available: https://doi.org/10.1007/978-981-16-2543-5_46.

3. B. Jehangir and S.R. Nayak, "A Hybrid-Based Ensemble Learning Model and DNN for Automated Diagnosis of Brain Tumour from MRI Images", Innovation in Electrical Power Engineering Communication and Computing Technology. Lecture Notes in Electrical Engineering, vol. 814, 2022, [online]

Available: https://doi.org/10.1007/978-981-16-7076-3_27.

4. H. Zhao, Y. Sun and H. Li, "Retinal vascular junction detection and classification via deep neural networks", Compute. Methods Programs Biomed., vol. 183, pp. 105096, 2020.

5. B. Gupta and S. Tiwari, "Lung Cancer Detection using Curvelet Transform and Neural Network", Int. J. Compute. Appl., vol. 86, no. 1, pp. 15-17, 2014.

6. Cancer, Sep. 2021, [online] Available: https://www.who.int/news-room/fact-sheets/detail/cancer.

7. C. R. Smitten Aar, K. A. Petersen, K. Stewart and N. Mitt, "Cancer incidence and mortality projections in the UK until 2035", Br. J. Cancer, vol. 115, no. 9, pp. 1147-1155, 2016.

8. M. N. Gurcan, L. E. Boucheron, A. Can, A. Adbhuta, N. M. Rajpoot and B. Yener, "Histopathological Image Analysis: A Review", IEEE Rev. Biomed. Eng., vol. 2, pp. 147-171, 2009.

9. S. R. Nayak, D. R. Nayak, U. Sinha, V. Arora and R. B. Pachauri, "Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study", Biomed. Signal Process. Control, vol. 64, pp. 102365, October 2020.

10. S. R. Nayak and J. Mishra, "Fractal dimension-based generalized box-counting technique with application to grayscale images", Fractals, vol. 29, no. 3, 2021.

11. S. R. Nayak, V. Arora, U. Sinha and R. C. Paonia, "A statistical analysis of COVID-19 using Gaussian and probabilistic model", J. Interduce. Math., vol. 24, no. 1, pp. 19-32, 2021.

12. S. Panigrahi, J. Das and T. Swarnkar, "Capsule network-based analysis of histopathological images of oral squamous cell carcinoma", J. King Saud Univ. - Compute. Inf. Sci., no. 2020.

13. S. R. Nayak, J. Nayak, S. Vimal, V. Arora and U. Sinha, "An Ensemble AI-Enabled MIoT for Automated Diagnosis of Malaria Parasite Expert Systems", 2021, [online] Available: https://doi.org/10.1111/exsy.12906.

14. S. Shandilya and S.R. Nayak, "Analysis of Lung Cancer by Using Deep Neural Network", Innovation in Electrical Power Engineering Communication and Computing Technology. Lecture Notes in Electrical Engineering, vol. 814, 2022, [online] Available: https://doi.org/10.1007/978-981-16-7076-3_37.

15. A. A. Borkowski, M. M. Bui, L. B. Thomas, C. P. Wilson, L. A. DeLand and S. M. Masto rides, "Lung and Colon Cancer Histopathological Image Dataset (LC25000)", Sep. 2019,