iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** VI **Month of publication:** June 2024

DOI: https://doi.org/10.22214/ijraset.2024.63367

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue VI June 2024- Available at www.ijraset.com

Lung Cancer Detection using Ensemble Algorithm

Dr. R. Bhargav Ram[1], Dr. E. N. V. Purna Chandra Rao[2], J. Anjali[3], S. Akash[4], S. Karthik[5]
[1]*Assistant Professor, Electronics & Communication Engineering, CMR College of Engineering & Technology, Kandlokoya, Hyderabad, India*
[2]*Professor & HOD, Electronics & Communication Engineering, CMR College of Engineering & Technology, Kandlokoya, Hyderabad, India*
[3, 4, 5]*Students, Electronics & Communication Engineering, CMR College of Engineering & Technology, Kandlokoya, Hyderabad, India*

Abstract: ***A deadly hereditary illness, lung cancer is caused by an aberrant proliferation of malignant cells in the lungs of an individual. As one of the most important organs in the human body, the lungs might have major consequences from lung cancer. In order to help both patients and physicians, we have concentrated on the rapid diagnosis of lung cancer in our work. Histopathology pictures and other diagnostic methods can also be used to diagnose lung cancer.In existing methodology, it is identified that the methods are time taking process and less Accuracy. The proposed work contains a hybridized model of Convolution neural networks and an ensemble of Machine Learning algorithms: Support Vector Classifier, Random Forest, and Ada Boost that detect the lung cancer using histopathology images.***

Keywords: Cancer Detection, Pre-Processing, Random forest, Ada Boost, Support Vector Machine (SVM), CT-Scan images, Prediction.

I. INTRODUCTION

Disease is a critical general heath issue overall with death rates expanding step by step. Cellular breakdown in the lungs, among any remaining disease types is the most widely recognized and lethal that happen both in people. Cellular breakdown in the lungs, furthermore realized carcinoma is arrangement of dangerous lung growths' (malignant knobs) because of uncontrolled development of cells in lung tissues. Eating tobacco and smoking are the main gamble factors for causing carcinogenic lung knobs. The endurance pace of cellular breakdown in the lungs patients consolidating all stages is exceptionally less generally 14%with a stretch of time of around 5-6 years. The fundamental issue with cellular breakdown in the lungs is that the majority of these malignant growth cases are analyzed in later phases of disease making medicines more hazardous and altogether lessening the endurance possibilities. Consequently location of cellular breakdown in the lungs in its previous stages can build the endurance chances up to 60-70% by giving the patients essential quick treatment and hence it controls the death rate. Little cell cellular breakdown in the lungs and non-little cell cellular breakdown in the lungs are two fundamental kinds of cellular breakdown in the lungs arrangements in light of cell qualities. The most normally happening is non-little cell cellular breakdown in the lungs that makes up around 80-85% of all cases, while 15-20% of disease cases are addressed by little cell cellular breakdown in the lungs. Lung cancer staging depends on spread of cancer in the lungs and tumor size.

Cellular breakdown in the lungs is essentially arranged into 4 phases arranged by earnestness: Stage I-Malignant growth is bound to the lung, Stage II and III-Disease is kept to the chest and Stage IV-Cellular breakdown in the lungs has spread from the chest to different pieces of the body. Cellular breakdown in the lungs finding should be possible by utilizing different imaging modalities such Positron Outflow Tomography (PET), Attractive Reverberation Imaging (X-ray), Processed Tomography (CT) and Chest X-beams. CT check pictures are generally liked over different modalities since they are more dependable, have better clearness and less bending.

II. LITERATURE REVIEW

A lot of works have already been proposed for the prediction of carcinoma through numerous researchers amongst them, paper projected several approaches for police to detect carcinoma in the early stages. They implemented different ML models like K-NN, Artificial Neural Networks, Support Vector Machine, Naive Bayes, and Decision Trees to know the idea of using these approaches, and comparison of results are performed while pre-processing and after it is done.

©IJRASET: All Rights are Reserved

[1]Sumathipala, deliberate a version wherever the picture statistics is taken from LIDC-IDRI, as soon as grouping the picture statistics picture filtration has been enforced, filtration is finished supported the affected person United Nations company went through diagnostic check and module degree is good enough to thirty and so snapshots whose module degree is good enough to thirty is split and for predicting they used random forest and logistic regression model. Paper gives a carcinoma detection machine victimization picture technique and device learning are hired to categorize the presence of carcinoma at some point of a CT- scan images and blood samples. [2]Fenwa, proposed a version in which they extracted the properties like brightness, the contrast from picture dataset. The use of extracting texture-primarily based features and on the one's varieties of ML set of rules are carried out one is ANN any other one is SVM, after which overall performance has been evaluated on each the set of rules to examine which set of rules is giving extra accuracy. In the work goal of the work is to propose a version for early detection and the right designation of the illness which could facilitate the health practitioner in saving the life of the affected person.[3]Maisa Daouda, in their work discussed how neural network models can be used in the prediction of cancer and also addressed some issues that are generally encountered while building neural network models for predicting cancer.[4]M.Siddardha Kumar, projected pre-managing strategies are likewise applied at some point of this paintings to urge accurate outcomes. In preprocessing approach, the morphological approach has been applied to expel the unwanted statistics from the image. The feature extraction system it's an accustomed restriction the only in all a type dataset via way of means of manipulating a few changed over alternatives. Exceptional Strategies were implemented to find the different ways for extracting geometrical and measurable properties from the image. [5]Swati Mukherjee, the evaluation and study of respiration organ sicknesses has been the most interesting research area of docs from time to present. They addressed this issue by introducing a novel methodology using deep neural mechanisms such as CNN and AI. The paper aims at detection, prediction, and diagnosing of carcinoma has to turn out to be vital as it simplifies resultant clinical board.[6]Wasudeo Rahane, Kyamelia Roy, aims to erect the development and pills of cancerous situations machine learning strategies are applied because of their accurate outcomes. Various sorts of ML algorithms like, Naive Thomas Bayes, SVM, logistic regression, are implemented in the core region for evaluation and diagnosis of carcinoma. [7]Sanjukta blue blood Jena, projected a model consisting of five kinds of feature extraction strategies that had been applied in individual class method to expect at that alternatives extraction approach that system gaining knowledge of method is giving quite a few accuracies.[8]Şaban Oztürk, type of histopathologic snapshots and identity of cancerous regions is kind of difficult due to picture heritage first-rate and determination. The difference between conventional tissue and cancerous tissue is extraordinarily tiny in a few cases. So, the alternatives of the tissue patches within the picture have key significance for computerized type. [9]Dendi Gayathri Reddy, projected a version this is cost-effective in predicting the ranges of respiration organ malignant neoplastic disorder through applying the thoughts of cc algorithms. It is a mixture of KNN, decision tree, and NN models beside fabric ensemble technique that helps to boost the accuracy. The obtained results were better than other existing models. Paper talks about the numerous devices gaining knowledge of algorithms that are applied to predict the survivability price of a person, and performance is estimated primarily based on root suggest square error. Using those concepts, we introduce a unique technique to are expecting cancer the use of ensemble strategies which might be mentioned in element within the subsequent section.

III. METHODOLOGY

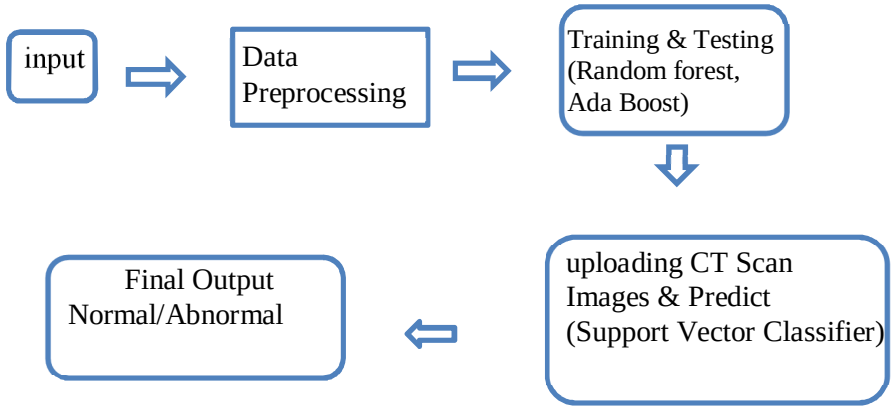


Fig 1:Block Diagram

In existing methodology,it is identified that the methods are time taking process and less Accuracy.So to overcome this problems we are using lung histopathology pictures, the proposed study employs a hybridised model of Convolution neural networks with an ensemble of Machine Learning algorithms, including Support Vector Classifier, Random Forest, and XG Boost. In all, this job has a 99.13% success rate.

1) *Upload Lung Cancer Dataset:* using this module we will Upload Lung Cancer Dataset
2) *Dataset Preprocessing:* using this module we will Dataset Preprocessing
3) *Run Ensemble Algorithm:* using this module we will Run Ensemble Algorithm
4) *Predict Lung Cancer Disease:* using this module we will Predict Lung Cancer Disease
5) *Train RBF on LUNGS CT-SCAN Image:* using this module we will Train RBF on LUNGS CT-SCAN Image.
6) *Predict Cancer from CT-Scan:* using this module we will Predict Cancer from CT-Scan.

Upload Lung Cancer Dataset. Dataset loaded and dataset contains total 1000 records and 25 columns and in above dataset we can see some string values are there but machine learning algorithms accept only numeric values so we need to preprocess above dataset to convert string values to numeric by assigning unique id's to each string value and in above graph in x-axis we can see disease stage as HIGH or LOW and number of patients detected in that stage are plotting in y-axis as bar graph. In the Data Preprocessing method the data converts string values to numeric values. After converting the data, use Ensemble algorithm to train the data. Bracket we can see test values and then by analysing those test values ensemble has given prediction result as LOW, HIGH or MEDIUM and we can see this result after square bracket and if disease high then application asking user to go for CT-SCAN image. Train Radial Basis Function(RBF) on LUNGS CT-SCAN Image to train Radial Basis Function(RBF) with lungs CT-SCAN. Radial Basis Function trained on images and its got prediction accuracy as 99 and always this accuracy may vary as Radial Basis Function calculate accuracy on random test data. Now click on 'Predict Cancer from CT-Scan' button to upload test image and then get prediction result.

IV. ALGORITHMS IMPLEMENTED

Support Vector Machine used for prediction, regression, and classification for the given input data. It classifies the input dataset via way of means of introducing a boundary called a hyper-plane that separates the dataset into components. Support vector mechine is used to differentiate linear and non-linear areas. A linear separation classifier is hired to split the affected and non-affected areas within the image. In non-linear separation, we're going to separate the affected element or place via way of means of representing the non-linear type. Logistic Regression is a famous modeling process utilized in the evaluation of epidemiologic datasets. The Logistic Regression approach initially calculates the usage of logistic characteristics, learns the coefficients for an Logistic Regression model, makes accurate predictions. Logistic Regression is also referred to as a binary classifier that calculates the category of classification which is primarily based on the speculation that works using the sigmoid function.

Random Forest is a technique that generates plenty of decision trees that are allowed to break up arbitrarily from a seed. This finally ends up like a forest of arbitrarily generates the decision trees. Final results of decision trees are ensembled using Random-Forest algorithmic utility that is anticipating to offer extra accuracy in preference to one tree will alone be giving. Individual decision trees are like if then-else tips that may be generated from the dataset directly.

V. RESULTS

In this section, we discuss and inspect the results for the built ensemble classifier with Random-Forest based on the different parameters. To analyze the result, we have used the data that consists of CT scanned images of Lungs.

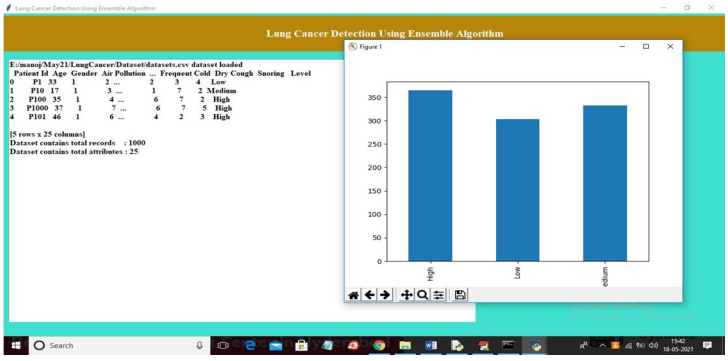


Fig 2: Data Preprocessing

It has 561 images belonging to class 1 and 416 images belonging to class 0 where class 0 refers to Benign and 1 refers to Malignant.

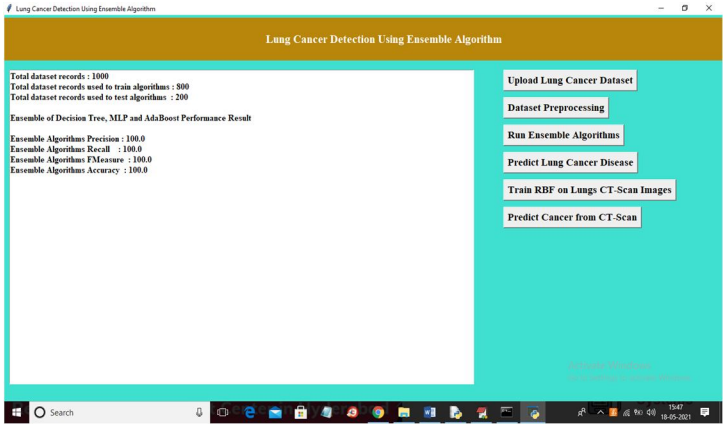


Fig 3: Run Ensemble Algorithm

Images from the dataset are preprocessed and segmented inorder to extract region based features and this data is used as dataset for classification. The dataset is split into an 8:2 ratio for training and testing.

In below screen in square bracket we can see test values and then by analysing those test values ensemble has given prediction result as LOW, HIGH or MEDIUM and we can see this result after square bracket and if disease high then application asking user to go for CT-SCAN image. Now click on 'Train Radial Basis Function on LUNGS CT-SCAN Image' to train Radial Basis Function with lungs CT-SCAN.

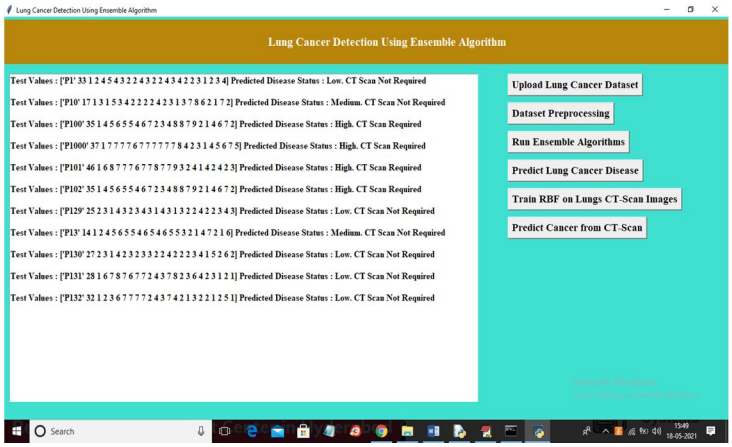


Fig 4: Predict Lung cancer From Dataset

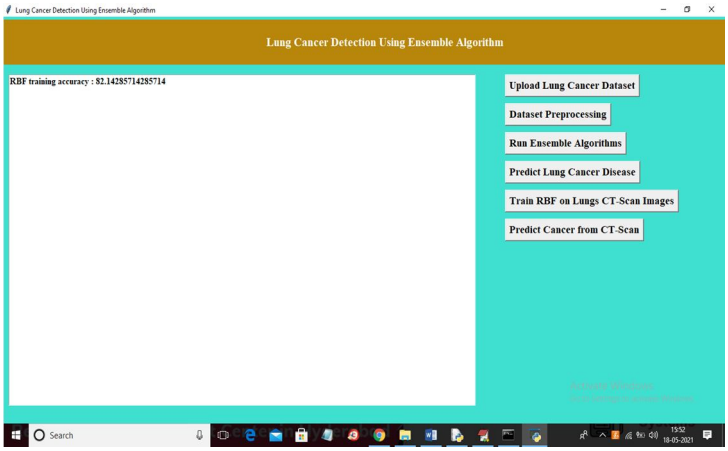


Fig 5:Accuracy

In above screen RBF trained on images and its got prediction accuracy as 82 and always this accuracy may vary as Radial Basis Function calculate accuracy on random test data. Now click on 'Predict Cancer from CT-Scan' button to upload test image and then get prediction result.

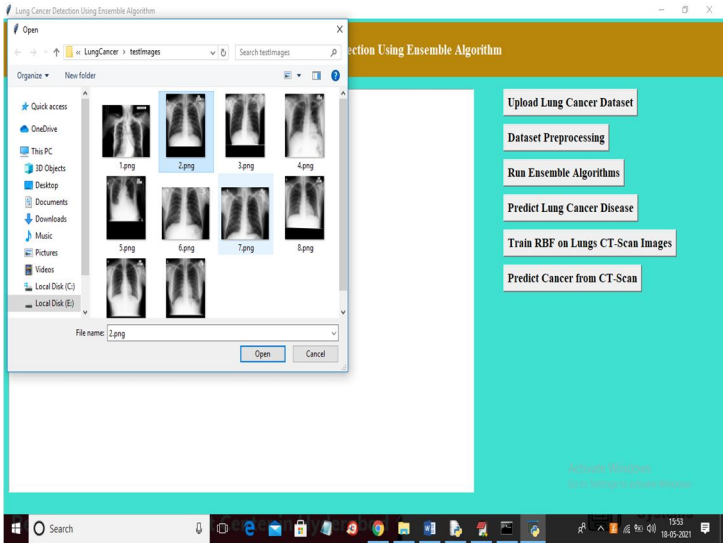


Fig 6:Upload the CT-scan images

In above screen selecting and uploading image and then click on 'Open' button to get below result.

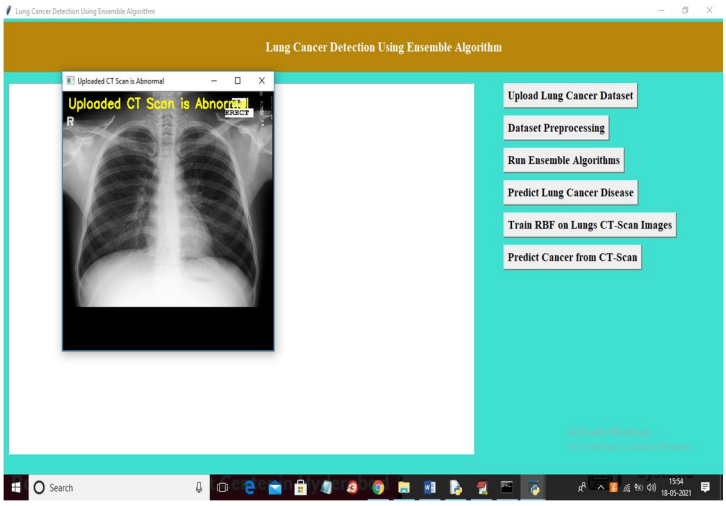


Fig 7: Final output

In above screen in image title bar or yellow colour text you can see prediction result as 'ABNORMAL'.

VI. CONCLUSION AND FUTURE SCOPE

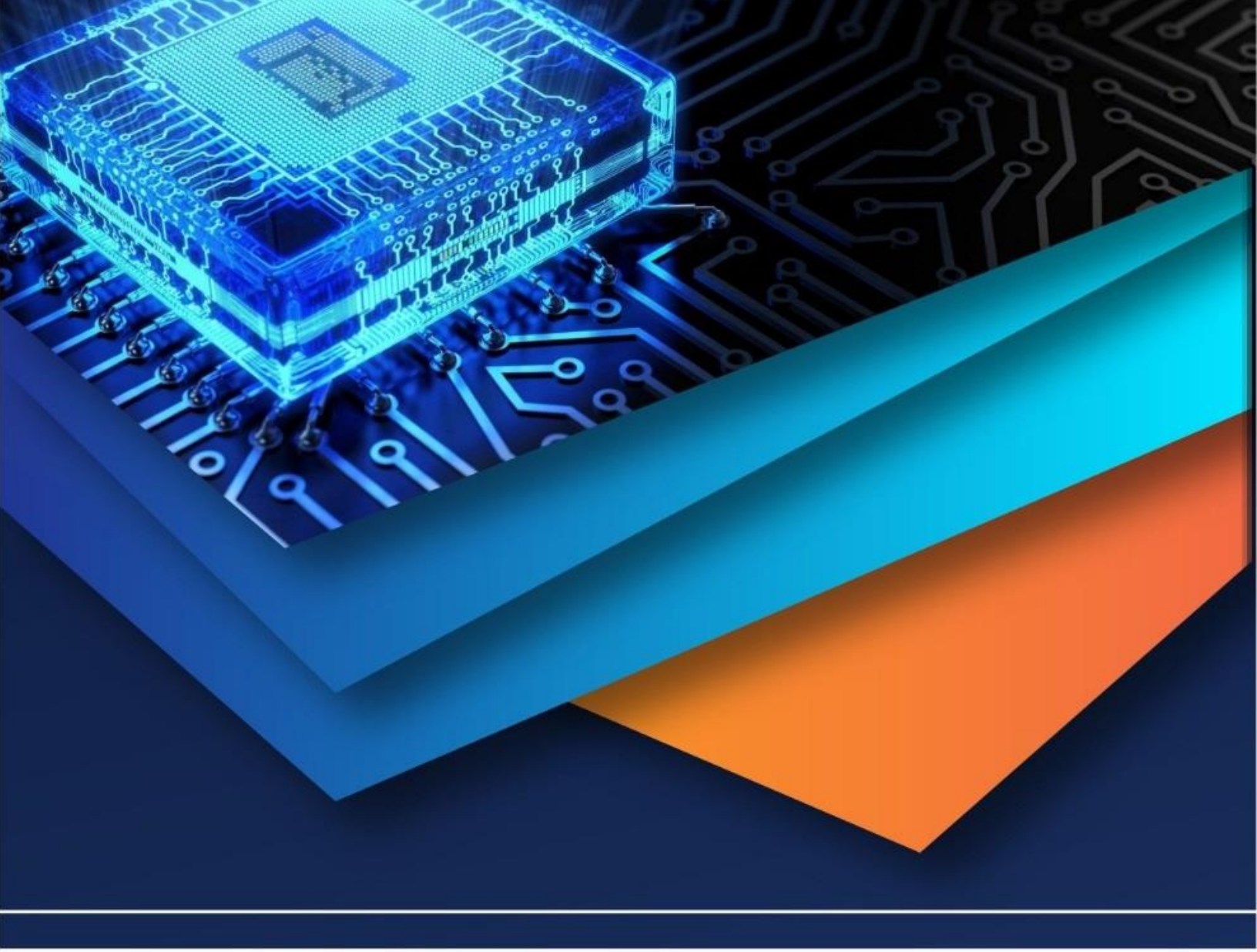
In this project we are using Lung Cancer dataset to train ensemble algorithm by combining AdaBoost, Multilayer Perceptron and Decision Tree and then after training when we upload test data then this algorithm will predict lung cancer stage as HIGH, LOW and MEDIUM and if HIGH detected then application will ask 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 Radial Basis Function algorithm and then after training when user upload test image then application will predict whether uploaded CT-SCAN is normal or abnormal.

In the future, Deep-Learning techniques like CNN can be used for the prediction of carcinoma. More range of pictures can be considered from different scanning techniques such as MRI, CT, PET, Xray, which can evoke additional accuracy, thereby serving to the medical field to supply fast prevention at low value. Continuous information can even be used rather than simply categorical information.

REFERENCES

[1] Sumathipala, Yohan, "Machine learning to predict lung nodule biopsy method using CT image features: A pilot study." Computerized Medical Imaging and Graphics 71 (2019).
[2] Fenwa, Olusayo D., Funmilola A. Ajala, and A. Adigun. "Classification of cancer of the lungs using SVM and ANN." Int. J. Comput. Technol. 15.1 (2016): 6418-6426.
[3] Daoud, Maisa, and Michael Mayo. "A survey of neural network-based cancer prediction models from microarray data." Artificial intelligence in medicine (2019).
[4] M.Siddardha Kumar, Prof.Dr.K.Venkata Rao. "Prediction Of Lung Cancer Using Machine Learning Technique: A Survey" International Conference on Computer Communication and Informatics (ICCCI -2021), Jan. 27 – 29, 2021, Coimbatore, India.
[5] Swati Mukherjee, Prof. S. U. Bohra. "Lung Cancer Disease Diagnosis Using Machine Learning Approach" Third International Conference on Intelligent Sustainable Systems [ICISS 2020].
[6] Wasudeo Rahane, Himali Dalvi. "Lung Cancer Detection Using Image Processing and Machine Learning HealthCare" IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India, IEEE 2018.
[7] Ozge Gunaydin, Melike Gunay, Oznur Sengel. "Comparison of Lung Cancer Detection Algorithms" Scientific Meeting on Electrical-Electronics & Biomedical Engineering, Computer Science (EBBT) IEEE 2019.
[8] Anita chaudhary, Sonit Sukhraj Singh,"Lung Cancer Detection on CT Images Using Image Processing ", International transaction on Computing Sciences,VOL 4, 2012.
[9] Mr.Vijay A.Gajdhane, Deshpande, "Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques"IOSR Journal of Computer Engineering (IOSR-JCE),e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 5, Ver. III , Sep – Oct. 2014.
[10] DendiGayathri Reddy, Emmidi Naga Hemanth Kumar, Desireddy Lohith Sai Charan Reddy, Monika P "Integrated Machine Learning Model for Prediction of Lung Cancer Stages from Textual data using Ensemble Method". 1st International Conference on Advances in Information Technology IEEE 2019.
[11] Amrit Sreekumar, Karthika Rajan Nair, Sneha Sudheer, Ganesh Nayar H, and Jyothisha J Nair. "Malignant Lung Nodule Detection using Deep Learning" International Conference on Communication and Signal Processing, July 28 - 30, 2020, India.
[12] Krishnaiah, V., G. Narsimha, and Dr N. Subhash Chandra. "Diagnosis of lung cancer prediction system using data mining classification techniques." International Journal of Computer Science and Information Technologies 4.1 (2013): 39-45.
[13] Radhika P R, Rakhi.A.S.Nair. "A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms". International Conference on Electrical, Computer and Communication Technologies (ICECCT) IEEE 2019.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)