Lung Cancer Identification Using CT-Scan with NCA-XG Boosting & KNN Algorithm

Uday kumar Kamasani - 700738157

Email: uxk81570 @ucmo.edu

University Of Central Missouri, MO, USA.

Department Of Computer Science.

Triveni Kummetha - 700739716

Email: txk97160 @ucmo.edu

University Of Central Missouri, MO, USA.

Department Of Computer Science.

Navya Bandla - 700745181

Email: nxb51810@ucmo.edu

University Of Central Missouri, MO, USA.

Department Of Computer Science.

ABSTRACT:

Lung cancer happens when abnormal cells start to grow in the walls of the bronchi or alveoli. Cancer-causing chemicals like those found in.

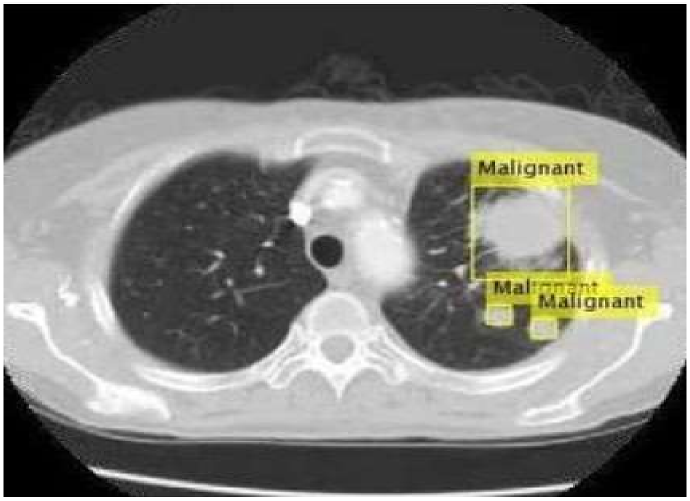
This growth and increase of cells are caused by tobacco, radiation, and asbestos. The main way to divide lung cancer into different types is by where the cancerous cells come from. Thanks to improvements in medicine, the condition can now also be identified at the molecular level. Scientists use this method to look for mistakes in the DNA and proteins that cancer cells make. There are hundreds of molecular findings, and mutations in EGFR, ALK, KRAS, and ROS1 are just a few of them. The molecular subtype of an illness may affect how quickly it develops and spreads. They can also predict how chemo, targeted therapy, and immunotherapy will affect the illness. Molecular cancer detection lets doctors make personalized treatment plans for each patient's cancer that have the best chance of working. In its early, treatable stages, lung cancer often doesn't show any signs.

GITHUB REPOSITORY URL: https://github.com/uday2909/FINALPROJECT.git

But as the disease gets worse, it can damage the tissue around it. This can make it hard for the lungs to work normally and cause signs like coughing up blood, shortness of breath, or pain.

Lung cancer is frequently disseminated via lymphatic metastases. Lymph, a transparent fluid that drains from our tissues, carries immune cells that assist in the body's defense against disease. It is transported throughout the body by lymphatic vessels. The lymph nodes are small bean-shaped connective tissue structures between lymph veins. They frequently trap cancer cells that have spread to the lymphatic system. The circulatory system is an alternative route by which cancer cells may reach other organs. In stage IV lung cancer, also known as metastatic lung cancer, the disease has spread to other organs.

It is still common to refer to cancer that has migrated to other organs as lung cancer. The treatment for lung cancer varies significantly based on whether the disease has spread to lymph nodes or other organs.



KEYWORDS:

Lung cancer, Computerized Tomography, Machine

Learning, Datasets, Algorithm, KNN Classifier,

AdaBoost Classifier.

INTRODUCTION:

Lung cancer is defined as the uncontrolled and abnormal multiplication of cells that begins in one or both lungs and then extends to the rest of the body. Lung cancer can originate from either lung. Lung cancer could originate in either lung. As a result of smoking, it is conceivable for cancer to develop in either lung.

GITHUB REPOSITORY URL: https://github.com/uday2909/FINALPROJECT.git

In healthy tissues, abnormal cells do not proliferate; however, when present in diseased tissues, they rapidly proliferate and form tumors. Normal cells do not proliferate. In healthy tissues, the proliferation of abnormal cells is negligible. Secondary lung cancer, in contrast to primary lung cancer, which begins in one section of the lungs and does not extend to other parts of the body, originates in a different part of the body and spreads to the lungs. Lung cancer primarily originates in a localized area of the lungs and does not spread. The term "primary lung cancer" alludes to the most prevalent form of lung cancer. Primary lung cancer is diagnosed approximately two to three times more frequently in men than in women. Early symptoms of lung cancer, when present in a patient's body, are frequently indicative of the onset of the disease in the patient's body. This is because lung cancer tends to progress slowly over time. As the number of individuals afflicted with lung diseases in today's industrialized cities continues to rise, there is a growing need for diagnostic techniques that are both precise and prompt. This is due to the growing demand for advanced medical technology.

According to medical personnel, the smoking habit, which influences lung cells, is the leading cause of lung cancer. Because cigarette smoke contains potentially dangerous substances such as carcinogens, a smoker will almost immediately experience the effects of this on the lung tissues. Due to the inhaled smoke, cigarette smoking is associated with an increased risk of developing lung cancer. Cigarette smoking is associated with an increased risk of lung cancer.

As a conceivable solution to these issues, deep learning is a possibility. The researchers at Northwestern University and the National Institutes of Health constructed a model using de-identified chest CT screening data from 45,856 participants in the National Lung Screening Trial.

After that, the performance of the model was compared to that of six board-certified radiologists.

When a single CT scan was utilized as a diagnostic instrument, the model performed as well as or better than human radiologists. At 94.4 percent AUC, the algorithm's performance attained a state-of-the-art level. In addition, the model decreased the number of mistaken positives by 11% and the number of false negatives by 5%.

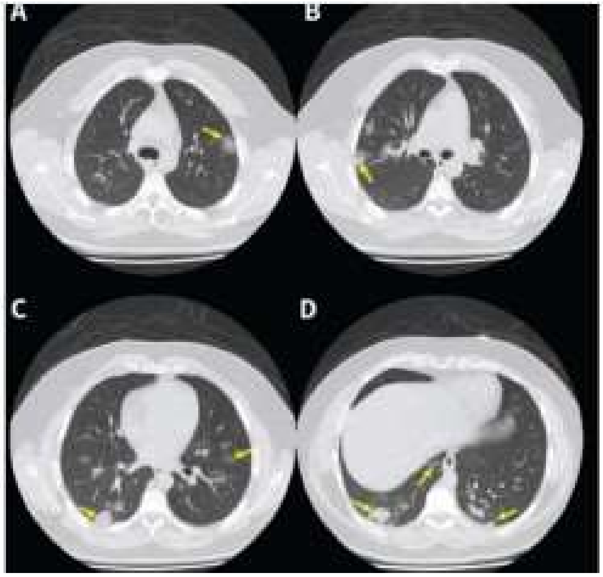
The model has the capability to detect moderately malignant tissue in a patient's lungs in addition to determining the patient's overall lung cancer stage. Given that the progression of suspicious tissue may indicate the presence of cancer, the fact that the deep learning system may be able to incorporate information from earlier imaging is advantageous.

The findings of this study, according to its authors, demonstrate that AI and deep learning have the potential to considerably enhance lung cancer screenings.

Even though lung cancer screenings are extremely essential, only 2% to 4% of eligible individuals in the United States receive them.

Researchers demonstrated that machine learning algorithms can detect breast cancer cells that have migrated to lymph nodes nearby. This discovery is essential for determining the most effective method of patient treatment. The machine learning models outperformed prior automated techniques and produced results comparable to those of human medical experts.

GITHUB REPOSITORY URL: https://github.com/uday2909/FINALPROJECT.git



***MOTIVATION*:**

Lung cancer is the most common type of cancer in men, but it is only the third most common type in women. Men are most likely to get lung cancer. Type of sickness that is most common. Lung cancer is the most common type of cancer in men. It happens when abnormal cells in the lungs turn cancerous and spread, forming tumors. Screening for lung cancer must start earlier than it does now if we want to cut down on the number of lives around the world that are caused by this disease. Most lung cancer symptoms don't show up until the disease is in an advanced state. Because of this, it is important to find the disease early by using all medical imaging technologies that are currently available. The goal of this study is to come up with a way to classify lung cancer that can automatically do diagnosis tasks in the early stages of the disease. For the review, computed tomography (CT) images of the lungs are used. The NCA-XG Boosting and KNN algorithms were used to sort the data from this study into different groups. After the pictures of the lung were preprocessed, the VGG19 algorithm was used to figure out which pictures were of the lung. This classifier uses an adaptive boosting strategy, and the pretrained approach is the basis of that strategy.

Main contributions and objectives:

- Easy to change - The software is easy to make, and its accuracy will get better as we get more pictures to look at.
- Database administration - Convenient data management methods in a single library format.
- Application: A CT scan, or computerized tomography, is the best way to find out if someone is sick at the tissue level. With the help of current machine learning algorithms, it is also the easiest and fastest way to do this.
- CT scans are now available at all primary and secondary health care centers.
- Adaptability: You don't have to be an expert in machine learning to use CT scans. The technology is already out there and has a flexible user interface that needs some training.

GITHUB REPOSITORY URL: https://github.com/uday2909/FINALPROJECT.git

Related work:

AI gets better when machine learning lets its parts learn from experience or extrapolate data. The software makes difficult decisions as it grows and learns from what it has already done. Here is a summary of the published research on how machine learning methods can be used to find lung cancer.

This study looks at how Decision trees, Naive Bayes, and Artificial neural network techniques can be used to predict the life expectancy of lung cancer patients after surgery. The above-mentioned methods were put through a stratified 10-fold cross-validation comparison study, and the accuracy of each predictor was found.

This paper looks at how different classification methods can be used to find brain tumors. Using information about volume and location, the total accuracy rate was calculated using 2 classification classes, such as logistic regression and quadratic discriminant, and 3 classification classes, such as Linear SVM, Coarse Gaussian SVM, Cosine KNN, and Complex and middle tree.

In this piece, different results are shown for each classifier on the lung cancer dataset. The KNN, SVM, NN, and Logistic Regression classifiers were used, and the correct success rates were found.

Support Vector Machine is the most accurate, with a 99.3% accuracy rate. When the suggested method was used on medical information, it helped doctors make more accurate decisions.

Several segmentation methods were talked about, such as Naive Bayes and Hidden Markov Model. There is a detailed explanation of how and why different segmentation methods are used to find lung cancer.

Instructions were given on how to make a simple flowchart for a program that could help find brain tumors. We talked about classification methods for two different kinds of data mining approaches.

1. Statistics: Naive Bayes and Support Vector Machine

2. Compression techniques for decision tree and neural networks

3. Numerous data sets were the topic of conversation.

The BRATS Dataset, the OASIS Dataset, and the NBTR Dataset are available.

Alternately, the authors conducted an experiment to investigate the impact of referral path and side effects on delays in a rapid outpatient diagnostic program for patients with suspected lung cancer, as well as the relationship between delays and disease stage and prognosis. There has been an exhaustive investigation into the characteristics of tumors, their structure, and the numerous delays that have occurred. A total of 565 patient restoration schematics were gathered for this investigation. 51.0% of the participants had lung growths, 8.5% had various injuries, and 19.6% had non-life-threatening radiological abnormalities. In the case of hemoptysis, first-line wait times were significantly shorter than in other instances. During the rulemaking process, an RODP was developed to facilitate the analysis. Estimates indicate that most patient postponements are caused by delays in the first and second treatment lines.

GITHUB REPOSITORY URL: https://github.com/uday2909/FINALPROJECT.git

They investigated numerous methods for measuring lung growth. The use of artificial neural networks, image processing, linear dependency analysis (LDA), and self-organizing maps (SOM) was among these. In conclusion, it is suggested that support vector machines be used as a characterization instrument. Support vector machines can be used to examine data and identify patterns during machine learning. At the outset of their study, [10] devised a technique for identifying lung development. In this manner, data pre-processing is performed to initiate the image enhancement procedure. At this juncture, datasets that have been prepared for testing under information mining and neural systems, which are both essential for differentiating between rehabilitative treatments, can be evaluated. Using back-propagation neural networks (BPNN) to classify images of information as malignant or innocuous, researchers were able to achieve the desired result. When making a diagnosis, medical professionals determine which stage of cancer will benefit them the most.

This study utilized network-based biomarker discovery and quality set enhancement methodologies to identify and validate traits associated with the development of lung cancer and related pathways. In addition to the characteristics predicted by previous research in these areas, they discovered a vast array of novel and unexpected characteristics associated with the hypothesized physiological capacity in smoking. Developed a network-based technique for dealing with observable smoking confirmation, classifying the qualities associated with lung tumor survival from those associated with non-smoking groups, and identifying all the qualities associated with lung tumor survival and non-smoking groups. A six-quality smoking score has been shown to predict the risk of lung enlargement and the probability of survival. This quality indicator may allow smokers to observe and identify lung expansion.

To investigate lung expansion, they employed information mining and streamlining techniques to extract insights from a vast quantity of datasets. It can be used to recognize and exploit malignant patterns in datasets. These patterns, which are identified in databases, can then be used to predict the prognosis of a disease based on the exact therapy cases stored in databases. Using computed tomography images and a previously described computer-aided diagnostic (CAD) order procedure, the authors demonstrated the identification of neural system enlargement.

To reconstruct the lung, CT scan highlights were reconstructed after being spliced together. The mean, standard deviation, skewness, and kurtosis, along with the fifth and sixth central moments, were used to determine if the data contained malignant cells. Objects are organized using forward- and backward-feeding neural networks to improve segmentation.

There's no question that the authors have been working for a long time on how to use different algorithms for artificial intelligence to find illnesses and give people medicine. An artificial neural network (ANN) can be used to look at data about breast cancer. Using data from microarrays and the UCI machine learning library, multilayer feedforward neural networks can be used in the same way as ANNs to find out when lung cancer starts.

The backpropagation method is used to set up the system. With cross-approval, you can compare datasets that have different numbers of secret layers and hubs that link to the same dataset. If an event from the UCI dataset (like breast cancer) happens, the different combinations of masked layers and linked hubs are expected to make this structure more accurate. The research gets more accurate as the number of hubs and hidden layers in the NCBI dataset keeps going up. Using a similar system in the brain, it is possible to predict what will happen to a patient. This is possible with the help of a system that makes decisions on its own.

Proposed framework:

Before starting to build a model from scratch, most machine learning engineers spend a lot of time pre-processing or cleaning the data. Some examples of data preprocessing processes are finding outliers and figuring out what to do with them, dealing with missing values, and getting rid of unwanted or noisy data.

"Image pre-processing" and "image processing" are both terms for the same thing: taking a picture down to its most basic level of abstraction. Entropy is a way to measure information, and this method does not add to the amount of picture information that is stored in the image. Instead, it lowers the amount of picture information that is stored in the image. The goal of preprocessing is to improve the quality of the image data by getting rid of unwanted distortions and improving certain visual properties that are needed for the picture to be further processed and analyzed. Before the picture is processed and analyzed, it is first pre-processed. The steps that makeup pre-processing can be put into two groups, which are explained below. The steps of preparation can be put into two different groups:

1. Filtering and cutting up images.

2. The Fourier transform and restoring a picture.

Pre-processing is needed for CT scans that are used as input. This is done to cut down on the noise that is already there and to get the pictures ready to be used in later steps, like image segmentation.

As a direct result, the input pictures will have less distortion, and the important parts of the inputs will stand out.

In the input database, there are four different types of cancer nodules: the well-circumscribed type, the juxta-pleural type, the vascularized type, and the pleural-tail type. Nodules from the first and second stages of cancer are also considered. Figure 2 shows the CT picture of the lung with cancer that was used as a starting point.

NCA-XG boosting:

The decision-tree-based ensemble Machine Learning technique known as XG Boost employs a gradient-boosting framework to improve both the precision and efficiency of its predictions. XG Boost was created by Microsoft Research and was given its present name in honor of the company's founder. Artificial neural networks typically outperform other algorithms and frameworks (images, text, etc.) when it comes to resolving prediction problems involving unstructured data. It is widely acknowledged that decision tree-based algorithms are the most effective method for the administration of relatively small to relatively medium-sized structured or tabular data sets.

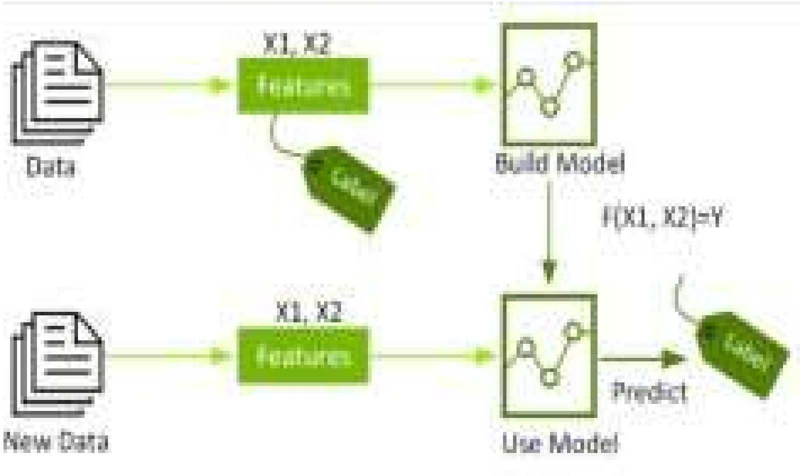
XG Boost and Gradient Boosting Machines (GBM) are two typical varieties of ensemble tree algorithms. The gradient descent architecture is utilized by both of this ensemble tree approaches to enhance the performance of feeble learners (CARTs in general).

Since it has helped individuals and teams win virtually every structured data competition on Kaggle, individuals and teams have developed a special fondness for the XG Boost tool. Participants in these competitions are required to submit data, after which statisticians and data miners compete to see who can develop the most reliable models for analyzing and interpreting the data. Initially, Python was used to develop XB Boost before R took over. Due to the high demand for its services, XG Boost has begun providing package implementations for a variety of languages, including Java, Scala, Julia, and Perl, among others. The increased popularity of XB Boost among the Kaggle community is a direct consequence of the new implementations that have been made possible by these.

XB Boost has been incorporated with a variety of additional tools and packages, such as scikit-learn for Python users and caret for R users. Due to its integration, distributed processing frameworks like Apache Spark and Task are also compatible with XB Boost. This year, InfoWorld gave XB Boost its Technology of the Year award, which it won with glowing colors.

A pattern-finding algorithm is applied to a labeled data set to train a model, which is then applied to a new data set to make label predictions.

While machine learning algorithms can be developed to process unprocessed data, the initial step is featuring selection. Neighborhood Component Analysis (NCA) is an efficient method for selecting significant feature elements when working with massive amounts of high-dimensional.



This algorithm is based on the closest neighbor feature weighting technique, which will be discussed in greater detail below. As an instrument for feature selection, the NCA method has proven effective on multiple microarray datasets for malignancies such as colon cancer, brain tumor, leukemia, lung cancer, and prostate cancer. For automated subtype categorization of cancer, an efficient classifier is required.

As a classifier for machine learning, we use the ensemble based XB Boosting algorithm. XB Boost is a rapid and effective implementation of gradient-boosted decision trees with performance optimization. The method was intended to be implemented as efficiently as possible in terms of processing time and memory consumption. One of the design goals was to optimize the utilization of available resources for model training.

KNN classifier:

Because it is sometimes advantageous to include more than one neighbor, this technique is also known as k-Closest Neighbor (k-NN) Classification, in which k closest neighbors are used to determine the class. Since training examples are required at runtime, i.e., they must be in memory, Memory-Based Classification is another name for this technique. This method is categorized as Lazy Learning because induction is deferred until execution time.

Classification is also referred to as Example-Dependent Classification or Case-Dependent Classification because it relies on training examples.

Following this distance evaluation, the k nearest neighbors are selected. The k nearest neighbors can then be utilized in a variety of methods to determine the category of q. Assigning the majority class to the query's closest neighbors is the quickest and smoothest method.

When determining the query's class, it is frequently prudent to assign a greater weight to the query's closest neighbors. Distance-weighted voting, in which neighbors vote on the class of the query case with ballots weighted by the inverse of their distance from the query, is a standard method.

ADA-Booster:

AdaBoost (Adaptive Boosting) is a popular boosting method that combines many weak classifiers to make a single strong classifier. By putting together several weak models and learning from the things they got wrong, we might be able to make a strong model. Decision Trees, Logistic Regression, and other methods can be used as classifiers. What does it mean to have bad classifiers? A weak classifier does better than guessing at chance, but it can't put things into groups. A bad predictor might say that people over 40 can't run a marathon but that people under 40 can. Even if you got more than 60% of the facts right, you would still get some wrong.

AdaBoost can learn from any way of grouping things and suggest a more accurate model. It is called the "best out-of-the-box classifier" because of this. AdaBoost and Decision Stumps are two to look at. In the Random Forest, Decision Stumps are "young" trees. There are two nodes on a leaf.

Instead, stumps are used by AdaBoost. Making choices on a whim is a bad idea. A mature tree predicts the target value by putting together all the possible variable choices. A stump can only decide what to do based on one factor. To understand how the AdaBoost algorithm works, we'll look at several ways to figure out if a person is "fit" (healthy).

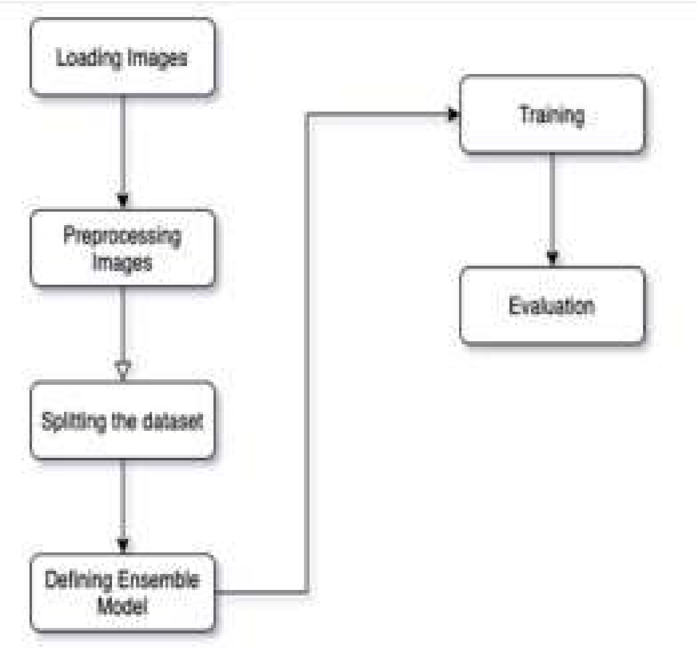
Data description:

Before building a model from inception, the vast majority of Machine Learning Engineers devote a significant amount of time to data pre-processing and cleaning. Examples of data preprocessing techniques include locating and handling outliers, handling absent values, and removing unwanted or noisy data.

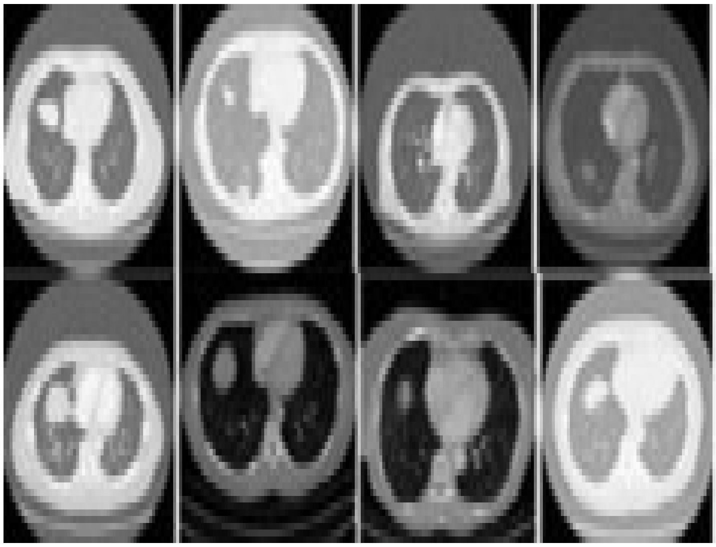
Image pre-processing, which is synonymous with image processing, refers to the most fundamental level of abstraction for images. According to entropy as a measure of information, this process does not contribute to the image's information; rather, it removes it. Pre-processing aims to improve the quality of the image data by removing unwanted distortions and enhancing certain visual qualities that are essential for subsequent image processing and analysis.

The procedures for pre-processing can be divided into the two categories enumerated below.

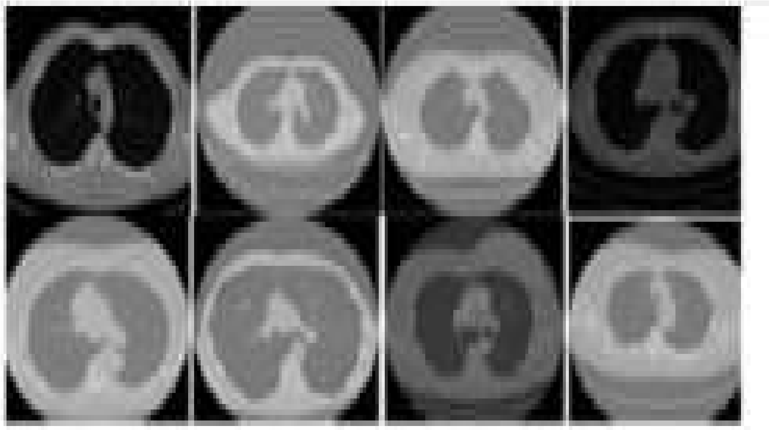
CT images must be preprocessed to remove noise and prepare them for subsequent stages such as image segmentation. As a result, input images will be less distorted and more of their correct components will be highlighted. CT images are prepared using MATLAB before being utilized. The study examines both primary and secondary-phase cancer nodules in the input database. Examine the well-circumscribed, juxta-pleural, vascularized, and pleural-tail nodules. The CT image of the lung with carcinoma that was used to train the model is depicted in Figure 2.



GITHUB REPOSITORY URL: https://github.com/uday2909/FINALPROJECT.git



(a)



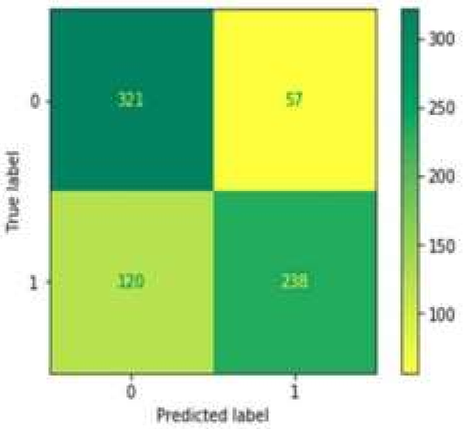
(b)

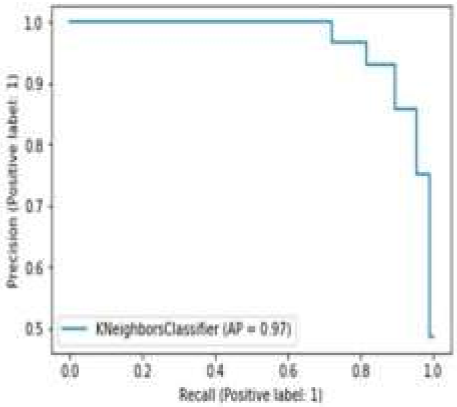
***Results*:**

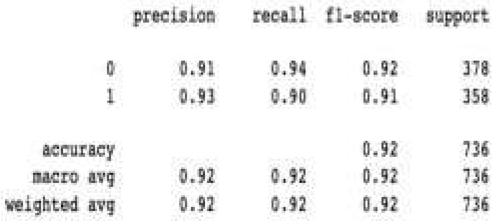
The success measures (accuracy, sensitivity, and specificity) are used to figure out the best way to teach (accuracy, sensitivity, and specificity). This system did the best, with average rates of 92 percent for accuracy, 92 percent for sensitivity, and 92 percent for specificity across all three performance categories. The KNN model has a 76 percent success rate, while NCA-XB Boosting

has the lowest success rate of the three at 76
percent.

GITHUB REPOSITORY URL: https://github.com/uday2909/FINALPROJECT.git

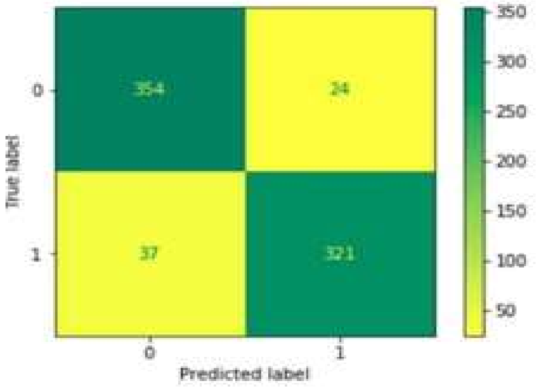


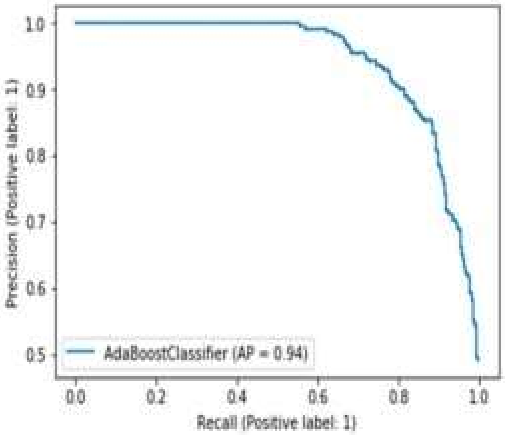


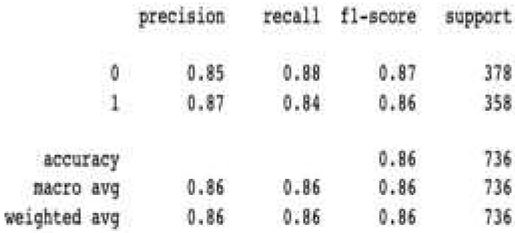


```
              precision    recall  f1-score   support

           0       0.91      0.94      0.92       378
           1       0.93      0.90      0.91       358

    accuracy                           0.92       736
   macro avg       0.92      0.92      0.92       736
weighted avg       0.92      0.92      0.92       736
```

Results of KNN Boosting



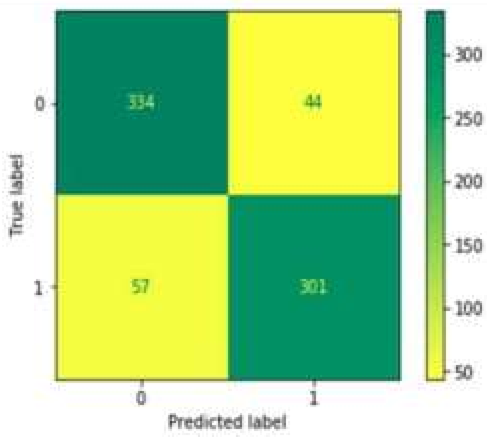




```
              precision    recall  f1-score   support

           0       0.85      0.88      0.87       378
           1       0.87      0.84      0.86       358

    accuracy                           0.86       736
   macro avg       0.86      0.86      0.86       736
weighted avg       0.86      0.86      0.86       736
```

Results of AdaBoost

GITHUB REPOSITORY URL: https://github.com/uday2909/FINALPROJECT.git



Conclusion:

Since the beginning of the 20th century, cancer rates have gone up a lot. This is because people are becoming less active, eating poorly, and smoking more. So, researchers and experts have come up with ways to fight this deadly disease. The results of a scientific study show that finding this condition early makes it easier to treat and lowers the risk of death that comes with it. This study suggests using an independent method based on probabilistic neural networks to make the most accurate diagnosis possible from CT images of the lungs. The suggested method was good at classifying and diagnosing because it used deep neural networks to pull out high-level traits. The KNN model is better than its competitors when it comes to precision and accuracy. The accuracy of the model can be improved by adjusting it with feature selection and a stacking-based approach, which can be used together. If we need to, we could add more photos to the collection in order to improve how well the model works.

GITHUB REPOSITORY URL: https://github.com/uday2909/FINALPROJECT.git

REFERENCES:

- [1] R. Navid, A. Mohsen, K. Maryam et al., “Computer-aided diagnosis of skin cancer: a review,” Current Medical Imaging, vol. 16, no. 7, pp. 781–793, 2020.
- [2] L. Hussain, W. Aziz, A. A. Alshdadi, M. S. Ahmed Nadeem, I. R. Khan, and Q.-U.-A. Chaudhry, “Analyzing the dynamics of lung cancer imaging data using refined fuzzy entropy methods by extracting different features,” IEEE Access, vol. 7, pp. 64704– 64721, 2019.
- [3] S. Lakshmanaprabu, S. N. Mohanty, K. Shankar, N. Arunkumar, and G. Ramirez, “Optimal deep learning model for classification of lung cancer on CT images,” Future Generation Computer Systems, vol. 92, pp. 374–382, 2019.
- [4] Armato SG, McLennan G, Bidaut L, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Med Phys. 2011;38:915–31.
- [5] Askarzadeh A. A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm. Comput Struct. 2016;169:1–12.
- [6] Cascio D, Magro R, Fauci F, Iacomi M, Raso G. Automatic detection of lung nodules in CT datasets based on stable 3D mass– spring models. Comput Biol Med. 2012;42:1098–109.
- [7] Chen H, Zhang J, Xu Y, Chen B, Zhang K. Performance comparison of artificial neural network and logistic regression model for differentiating lung nodules on CT scans. Exp Sys Appl. 2012;39:11503–9.
- [8] Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature. 2018;553:446.
- [9] Cancer. Accessed: Apr. 30, 2021. [Online]. Available: https:// en.wikipedia.org/wiki/Cancer
- [10] P. Chaudhari, H. Agarwal, and V. Bhateja, “Data augmentation for cancer classification in oncogenomics: an improved KNN based approach,” Evol. Intell., pp. 1–10, 2019.
- [11] S. F. Khorshid and A. M. Abdulazeez, “BREAST CANCER DIAGNOSIS BASED ON K-NEAREST NEIGHBORS: A REVIEW,” PalArch’s J. Archaeol. Egypt/Egyptology, vol. 18, no. 4, pp. 1927–1951, 2021.
- [12] F. Q. Kareem and A. M. Abdulazeez, “Ultrasound Medical Images Classification Based on Deep Learning Algorithms: A Review.”
- [13] D. Q. Zeebaree, A. M. Abdulazeez, D. A. Zebari, H. Haron, and H. N. A. Hamed, “Multi-Level Fusion in Ultrasound for Cancer Detection Based on Uniform LBP Features.”
- [14] J. R. F. Junior, M. Koenigkam-Santos, F. E. G. Cipriano, A. T. Fabro, and P. M. de Azevedo-Marques, “Radiomics-based features for pattern recognition of lung cancer histopathology and metastases,” Comput. Methods Programs Biomed., vol. 159, pp. 23–30, 2018.
- [15] I. Ibrahim and A. Abdulazeez, “The Role of Machine Learning Algorithms for Diagnosing Diseases,” J. Appl. Sci. Technol. Trends, vol. 2, no. 01, pp. 10–19, 2021.
- [16] P. Das, B. Das, and H. S. Dutta, “Prediction of Lungs Cancer Using Machine Learning,” EasyChair, 2020.

GITHUB REPOSITORY URL: https://github.com/uday2909/FINALPROJECT.git

- [17] G. A. P. Singh and P. K. Gupta, “Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans,” Neural Comput. Appl., vol. 31, no. 10, pp. 6863–6877, 2019.
- [18] B. Charbuty and A. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” J. Appl. Sci. Technol. Trends, vol. 2, no. 01, pp. 20–28, 2021.
- [19] H. A. Hussein and A. M. Abdulazeez, “COVID-19 PANDEMIC DATASETS BASED ON MACHINE LEARNING CLUSTERING ALGORITHMS: A REVIEW,” PalArch’s J. Archaeol. Egypt/Egyptology, vol. 18, no. 4, pp. 2672–2700, 2021.
- [20] D. M. Abdullah and N. S. Ahmed, “A Review of most Recent Lung Cancer Detection Techniques using Machine Learning,” Int. J. Sci. Bus., vol. 5, no. 3, pp. 159–173, 2021.
- [21] M. I. Faisal, S. Bashir, Z. S. Khan, and F. H. Khan, “An evaluation of machine learning classifiers and ensembles for earlystage prediction of lung cancer,” in 2018 3rd International Conference on Emerging Trends in Engineering, Sciences and Technology (ICEEST), 2018, pp. 1–4.
- [22] D. Q. Zeebaree, H. Haron, and A. M. Abdulazeez, “Gene selection and classification of microarray data using convolutional neural network,” in 2018 International Conference on Advanced Science and Engineering (ICOASE), 2018, pp. 145–150.
- [23] D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and D. A. Zebari, “Trainable model based on new uniform LBP feature to identify the risk of the breast cancer,” in 2019 International Conference on Advanced Science and Engineering (ICOASE), 2019, pp. 106–111.
- [24] H. Tang, J. Zhao, and X. Yang, “Explore machine learning for analysis and prediction of lung cancer related risk factors,” in Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence, 2018, pp. 41–45.

GITHUB REPOSITORY URL: https://github.com/uday2909/FINALPROJECT.git