**ABSTRACT**

Lung cancer is one of the most serious and challenging cancer type to detect among other types of cancers, with highest mortality due to cancer. Through the time Machine learning techniques are being used in medical researches and many significant developments were made in cancer research for decades.

There have been a lot of Machine Learning algorithms which could significantly help in feature extraction and other radiological analysis. With help of available datasets over the research organizations it is made possible easily. Even after all these findings and developments, problem in finding datasets used for training and testing the models persists.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO**  **1** | **TITLE ABSTRACT**  **LIST OF ABBREVIATIONS**  **INTRODUCTION** | **PAGE NO** |
|  | 1.1 Overview |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **2** | **LITERATURE SURVEY** |  |
| **3** | **PROBLEM DEFINITION** |  |
|  | 3.1 Existing System |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

1. DESIGN OF PROPOSED SYSTEM
2. IMPLEMENTATION OF PROPOSED SYSTEM
3. CONCLUSION AND FUTURE ENHANCEMENT
   1. [Conclusion](#_TOC_250005)
   2. [Future Enhancement](#_TOC_250004)

[REFERENCES](#_TOC_250003)

[APPENDIX](#_TOC_250002)

[Source Code](#_TOC_250001)

# 7 Screenshots

**LIST OF ABBREVIATIONS**

Tr Training data

Ts Testing data

ANN Artificial Neural Network

DCNN DenseNet Convolution Neural Network

SVM Support Vector Machine

AdaBoost Adaptive Boost

SNMV Self Normalized Multi-View

CNN Convolution Neural Network

# INTRODUCTION:

* 1. **OVERVIEW**

Lung cancer is a type of cancer that starts when abnormal cells grow in an uncontrolled way in the lungs. It is a serious health issue that can cause severe harm and death.Symptoms of lung cancer include a cough that does not go away, chest pain and shortness of breath.It is important to seek medical care early to avoid serious health effects. Treatments depend on the person’s medical history and the stage of the disease.

The most common types of lung cancer are non-small cell carcinoma (NSCLC) and small cell carcinoma (SCLC). NSCLC is more common and grows slowly, while SCLC is less common but often grows quickly.Lung cancer is a significant public health concern, causing a considerable number of deaths globally.

GLOBOCAN 2020 estimates of cancer incidence and mortality produced by the International Agency for Research on Cancer (IARC) show as lung cancer remains the leading cause of cancer death, with an estimated 1.8 million deaths (18%) in 2020.Smoking tobacco (including cigarettes, cigars, and pipes) is the primary risk factor for lung cancer but it can also affect non-smokers. Other risk factors include exposure to secondhand smoke, occupational hazards (such as asbestos, radon and certain chemicals), air pollution, hereditary cancer syndromes, and previous chronic lung diseases.

**Early stage disease**: The primary treatment for early stage lung cancer (i.e. tumour limited to the lung, with no metastatic dissemination to distant organs or lymph nodes) is surgical removal of the tumour through procedures such as lobectomy, segmentectomy, or wedge resection. Neoadjuvant therapy (chemotherapy and/or radiation therapy before surgery) can help reduce tumour size, making it more manageable for surgical removal. Adjuvant treatment (chemotherapy and/or radiation therapy) is very often recommended after surgery to reduce the risk of cancer recurrence. In cases where surgery is not feasible, radiation therapy or stereotactic body radiation therapy (SBRT) may be used as the primary treatment. Targeted therapy and immunotherapy may also be considered based on specific tumour characteristics. Individualized treatment plans should be discussed with healthcare professionals.

**Advanced disease:**The treatment for metastatic stage lung cancer, where the cancer has spread to distant organs or lymph nodes, is based on various factors, including the patient's overall health, the extent and location of metastases, histology, genetic profile, and individual preferences. The primary goal is to prolong survival, alleviate symptoms, and improve quality of life.Systemic therapies, such as chemotherapy, targeted therapy, and immunotherapy, play a crucial role in the treatment of metastatic lung cancer.Chemotherapy is often the first-line treatment for the majority of patients around the world and involves the use of drugs that circulate throughout the body to kill cancer cells. Combination chemotherapy regimens are commonly used, and the choice of drugs depends on factors such as the histological type of the cancer and the patient's general health conditions.

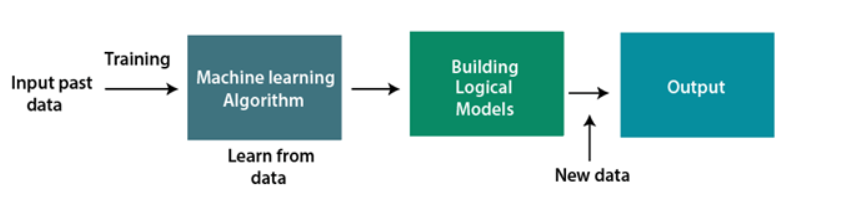
## **Machine learning**

Machine learning is a subfield of artificial intelligence that uses algorithms trained on data sets to create self-learning models that are capable of predicting outcomes and classifying information without human intervention. Machine learning is used today for a wide range of commercial purposes, including suggesting products to consumers based on their past purchases, predicting stock market fluctuations, and translating text from one language to another.

In common usage, the terms “machine learning” and “artificial intelligence” are often used interchangeably with one another due to the prevalence of machine learning for AI purposes in the world today. While AI refers to the general attempt to create machines capable of human-like cognitive abilities, machine learning specifically refers to the use of algorithms and data sets to do so.

## **How does machine learning work?**

At its core, the method simply uses algorithms – essentially lists of rules – adjusted and refined using past data sets to make predictions and categorizations when confronted with new data. For example, a machine learning algorithm may be “trained” on a data set consisting of thousands of images of flowers that are labeled with each of their different flower types so that it can then correctly identify a flower in a new photograph based on the differentiating characteristics it learned from other pictures.



**Various Machine Learning Algorithm**

In general cancer cells are cells that multiply abnormally without any control, which results in tumour growth and it can mitigate from one part to other part of the body or a lymph node through the bloodstream in a process called metastasis. It contains mutated genes that transforms normal cells into cancer cells, making it a genetic disease. It has been found that Lung cancer is the most challenging to detect as the symptoms starts to develop and occur only in advanced stages, while showing little to no signs in the early stages, but early predictions and diagnosis were found to be one of the best ways to survive with it.

Over the period, many different approaches have been discovered in predicting Lung Cancer using machine learning, many of those seems to give a desirable output. Some of the most commonly used Machine Learning and Deep Learning algorithms are,

* SVM
* RF
* CNN
* KNN
* Naïve Bayes
* Hybrid/fusion

## **Types of machine learning:**

### **Unsupervised machine learning**

Unsupervised machine learning uses *unlabeled* data sets to train algorithms. In this process, the algorithm is fed data that doesn't include tags, which requires it to uncover patterns on its own without any outside guidance.

### **Supervised machine learning**

In supervised machine learning, algorithms are trained on*labeled* data sets that include tags describing each piece of data. In other words, the algorithms are fed data that includes an “answer key” describing how the data should be interpreted. Supervised machine learning is often used to create machine learning models used for prediction and classification purposes.

Supervised learning can be grouped further in two categories of algorithms:

* Classification
* Regression

**Classification algorithm**

The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data. Classes can be called as targets/labels or categories.

Unlike regression, the output variable of Classification is a category, not a value, Since the Classification algorithm is a Supervised learning technique, hence it takes labeled input data, which means it contains input with the corresponding output.

f(x), where y = categorical output

**Learners in classification problems**

In the classification problems, there are two types of learners:

1. **Lazy Learners:** Lazy Learner firstly stores the training dataset and wait until it receives the test dataset. In Lazy learner case, classification is done on the basis of the most related data stored in the training dataset. It takes less time in training but more time for predictions.  
   **Example:** K-NN algorithm, Case-based reasoning
2. **Eager Learners:** Eager Learners develop a classification model based on a training dataset before receiving a test dataset. Opposite to Lazy learners,. **Example:** Decision Trees, Naïve Bayes, ANN.

**Regression**

Regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1.

Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems.

**K-Nearest Neighbour Algorithm**

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. algorithm stores all the available data and classifies a new data point based on the similarity.

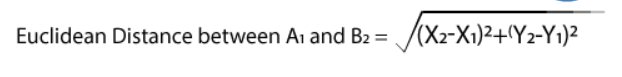
KNN can be used to identify patients with similar characteristics. It is applied to patient data, considering features such as markers or clinical variables. Patients with similar feature patterns are more likely to share similar lung cancer outcomes.

**How does K-NN Algorithm Work?**

The K-NN working can be explained on the basis of the below algorithm:

* **Step-1:** Select the number K of the neighbors
* **Step-2:** Calculate the Euclidean distance of **K number of neighbors**
* **Step-3:** Take the K nearest neighbors as per the calculated Euclidean distance.
* **Step-4:** Among these k neighbors, count the number of the data points in each category.
* **Step-5:** Assign the new data points to that category for which the number of the neighbor is maximum.
* **Step-6:** Our model is ready.

Euclidean Distance:



**Support Vector Machine**

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.  It can be applied to classify patients into cancerous and non-cancerous based on extracted features. With reference to nature of decision boundary the algorithm is divided into two main parts namely, the linear and non-linear SVMs. It is utilized with medical imaging data like CT scans, it aims in find optimal boundaries that separate the two classifications of patients.

**Types of SVM**

**Linear SVM:** Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

**Non-linear SVM:** Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

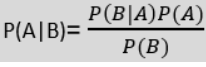
**Naïve Bayes algorithm**

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. .It is mainly used in *text classification* that includes a high-dimensional training dataset.Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.

It assumes independence between features, it can be employed to calculate the probability of a patient having cancer with the help of observed feature values. It is used in dealing with categorical and textual data in Lung Cancer studies.

**Bayes' theorem**

Bayes' theorem is also known as Bayes' Rule or Bayes' law, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.



**P(A|B) is Posterior probability**: Probability of hypothesis A on the observed event B.

**P(B|A) is Likelihood probability**: Probability of the evidence given that the probability of a hypothesis is true.

**P(A) is Prior Probability**: Probability of hypothesis before observing the evidence.

**P(B) is Marginal Probability**: Probability of Evidence.

**Decision Tree**

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.The decisions or the test are performed on the basis of features of the given dataset. It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.

**How does the Decision Tree algorithm Work?**

In a decision tree, for predicting the class of the given dataset, the algorithm starts from the root node of the tree. This algorithm compares the values of root attribute with the record (real dataset) attribute and, based on the comparison, follows the branch and jumps to the next node.

* Begin the tree with the root node, says S, which contains the complete dataset.
* Find the best attribute in the dataset using Attribute Selection Measure (ASM).
* Divide the S into subsets that contains possible values for the best attributes.
* Generate the decision tree node, which contains the best attribute.
* Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

**Random Forest**

It is an ensemble learning method that builds multiple decision trees and combine their predictions which are more accurate and stable. This can also be used in both classification and regression. Each of the d-trees are trained on a random subset of the data and features. It is effective in handling heterogeneous datasets and can capture complex relationships between different types of information, making it valuable for integrative lung cancer prediction models.

**Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.**Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

**Hybrid/Fusion**

These involves combining multiple models or techniques to address the limitation of individual methods and enhance the performance. Even though these seems pretty good as one is used to overcome other’s limitation, they are still prone to disadvantages in times these would be sensitive towards noise of the datasets. Some of these algorithms are, RF-SVM, GA-NN, ENS-NNSVM, ANN EA, …

**Convolution Neural Network (CNN)**

Convolutional Neural Networks is a deep learning neural network architecture used in computer vision which enables computers for processing structured grid data, such as images. It uses convolutional layers to automatically learn hierarchical features. It can analyse medical images (e.g., CT scans) to identify patterns associated with cancerous lesions. CNNs have shown remarkable success in image-based lung cancer prediction. They can automatically extract features from medical images, enabling accurate classification of cancerous and non-cancerous regions.

Convolutional Neural Networks, like neural networks, are made up of neurons with learnable weights and biases. Each neuron receives several inputs, takes a weighted sum over them, pass it through an activation function and responds with an output. The whole network has a loss function and all the tips and tricks that we developed for neural networks still apply on Convolutional Neural Networks. Neural networks, as its name suggests, is a machine learning technique which is modelled after the brain structure. It comprises of a network of learning units called neurons. The more labelled images the neurons are exposed to, the better it learns how to recognize other unlabelled images. We call this the process of training neurons.

There are four layered concepts we should understand in Convolutional Neural Networks:

1. Convolution,
2. ReLu,
3. Pooling and
4. Full Connectedness (Fully Connected Layer).

**LITERATURE SURVEY**

Radiomics can quantify tumours phenotypic characteristics non-invasively by applying feature algorithms to medical imaging data. In the study performed on lung cancer patients, radiomic features and the tumour histologic subtypes are investigated. Machine Learning methods have been used to predict histologic subtypes and their performance are valuated. In this study 440 features were extracted from 350 patients segmented tumour volumes of pretreatment CT images. With the multivariate analysis 24 feature selection methods and 3 classifier methods for histology prediction was performed and it is found that Naïve Bayes had highest accuracy than others.

An efficient lung cancer detection and prediction algorithm was built using MATLAB, where image processing technique is utilized. Multi-stage classification was used to detect lung cancer. Prediction of the lung cancer has been done with the proposed algorithm, if there is no cancer affected cell in the input image, the algorithm then checks the probability of lung cancer. If cancer affected cell has been found, the algorithm then checks the corresponding stage of the cancer such as initial, middle and final stage. Before every stage of classification image enhancement and segmentation has been done using several techniques. Image scaling, colour space transformation and contrast enhancement has been used for image enhancement. Threshold and marker-controlled watershed-based segmentation have been used for segmentation. For feature extraction GLCM (Gray Level Cooccurrence Method) technique is utilized which can arrange large combination of pixel brightness values. Acquired results have been compared with other preexisting works to find the performance.

# PROBLEM DEFINITION

# EXISTING SYSTEM

# In the proposed system few different smaller datasets are taken, the X ray images are then pre-processed which include,

# • Increasing contrast of the image using histogram equalization.

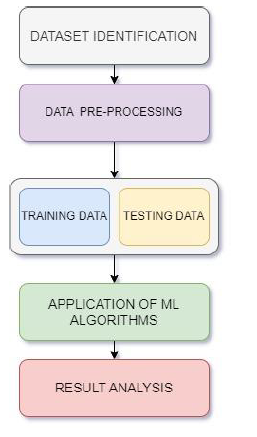
# • Removal of noise using median filtering

# .• Image resizing

# • Image normalization

# based on mean standard deviation. DenseNet-121 of CNN architecture is used for the proposed system, with this a base model is trained and subdivided into two modes further which are trained with two datasets and are evaluated individually, further the model is re-trained with another dataset too.

**General working flow of cancer prediction systems**



# 

# COMPARISION CHART

# 

# PROPOSED SYSTEM

Proposed a system which reveals that logistic regression algorithm gives high accuracy than other classifiers and provides maximum classification accuracy.

**IMPLEMENTATION OF PROPOSED SYSTEM**

**CONCLUSION AND FUTURE ENHANCEMENTS**

**CONCLUSION:**

Naïve Bayes outperformed other classifiers on non-invasive and cost-effective radiomics for precision medicine.

# FUTURE ENHANCEMENT:

Real-time object detection is the goal of the majority of robotics and computer vision systems. Because of the early research in this area, it is creating output in a range of directions with significant advancement. In many scenarios, the Yolo algorithm's object detection is underutilized, a problem that may be resolved in the future. Subsequent studies ought to concentrate on improving precision, handling complex scenarios, tracking many objects,and identifying objects particular to a certain domain.With the day-to-day growth in technology, it led the algorithm to get enhanced by overcoming the shortfalls from all the 7 previous versions. The prediction accuracies can further be enhanced by the quality of dataset in the future by preprocessing the dataset by enhancing the contrast, reducing the noise and normalizing the image the quality can further be improved which might be used to receive even higher accuracy, with the higher computational power, larger amount of data can be utilized to improve training the models with multiple varieties of data across different terrains of the world. With the standardizing and reducing the difference in Lung CTs a common model can be applied for any kind of data across different regions which further reduces training, testing and validating of different models for every type of data. With the increase in the number of classes of datasets, it might be more precise in prediction. It might seem to be a bit over looked to use these in real-time identifications with involving considerably lesser involvement of oncologists with the current technological advancements, but in the near future it might be possible to have a lesser involvement of humans which makes a significant effectiveness in time and man power consumption. Although it might help reducing the work of Oncologists, the matter privacy and patient's concern in using their data should be handled securedly.

# REFERENCES

[1] G. Sruthi, C. L. Ram, M. K. Sai, B. P. Singh, N.Majhotra and N. Sharma, "Cancer Prediction using Machine Learning," 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha

Nagar, India, 2022, pp. 217-221, doi: 10.1109/ICIPTM54933.2022.9754059.

[2] D. Manivannan., N. K. Manikandan and M.Kavitha, "Discovering Lung Cancer Cell UsingMachine Learning," 2022 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), Chennai, India, 2022, pp. 1-3, doi:

10.1109/ICDSAAI55433.2022.10028961.

[3] Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, Mak R, Aerts HJ. Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology. Front Oncol. 2016 Mar 30;6:71. doi: 10.3389/fonc.2016.00071.

[4] Hazra, Animesh & Bera, Nanigopal & Mandal, Avijit. (2017). Predicting Lung Cancer Survivability using SVM and Logistic Regression Algorithms. International Journal of Computer Applications. 174. 19-24. 10.5120/ijca2017915325.

[5] Akram, Sheeraz & Javed, Muhammad & Qamar, Usman & Khanum, Aasia & Hassan, Ali. (2014). Artificial Neural Network based Classification of Lungs Nodule using Hybrid Features from Computerized Tomographic Images. Applied Mathematics & Information Sciences. 9.183-195. 10.12785/amis/010124.

[6] J. Alam, S. Alam and A. Hossan, "Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifie," 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering(IC4ME2), Rajshahi, Bangladesh, 2018, pp. 1-4,

doi: 10.1109/IC4ME2.2018.8465593.

[7] D. Rawat, Meenakshi, L. Pawar, G. Bathla and R. Kant, "Optimized Deep Learning Model for Lung Cancer Prediction Using ANN Algorithm," 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC),Coimbatore,India,2022,pp.889-894, doi:10.1109/ICESC54411.2022.9885607.

[8] R. P.R., R. A. S. Nair and V. G., "A Comparative Study of Lung Cancer Detection

using Machine Learning Algorithms," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2019, pp. 1-4, doi: 10.1109/ICECCT.2019.8869001.

[9] Khan A, Tariq I, Khan H, Khan SU, He N, Zhiyang L, Raza F. Lung Cancer Nodules Detection via an Adaptive Boosting Algorithm Based on Self-Normalized Multiview Convolutional Neural Network. J Oncol. 2022 Sep 26;2022:5682451. doi: 10.1155/2022/5682451.

[10] Ausawalaithong, Worawate & Thirach, Arjaree & Marukatat, Sanparith & Wilaiprasitporn,Theerawit. (2018). Automatic Lung Cancer Prediction from Chest X-ray Images Using the Deep Learning Approach. 1-5.

10.1109/BMEiCON.2018.8609997.

[11] V. Nisha Jenipher and S. Radhika, "A Study on Early Prediction of Lung Cancer Using Machine Learning Techniques," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 911-916,

doi: 10.1109/ICISS49785.2020.9316064.

[12] J. Al-Tawalbeh, B. Alshargawi, H. Alquran, W.Al-Azzawi, W. A. Mustafa and A. Alkhayyat, "Classification of Lung Cancer by Using Machine

Learning Algorithms," 2022 5th International Conference on Engineering Technology and its Applications (IICETA), Al-Najaf, Iraq, 2022, pp.528-531,

doi:

10.1109/IICETA54559.2022.9888332.

[13] J. Oentoro, R. Prahastya, R. Pratama, M. S. Kom and M. Fajar, "Machine Learning Implementation in Lung Cancer Prediction – A Systematic Literature Review," 2023 International Conference on Artificial Intelligence in Information

and Communication (ICAIIC), Bali, Indonesia,

2023, pp. 435-439, doi: 10.1109/ICAIIC57133.2023.10067128.

[14] Huang S, Yang J, Shen N, Xu Q, Zhao Q.Artificial intelligence in lung cancer diagnosis and prognosis: Current application and future

perspective. Semin Cancer Biol. 2023 Feb;89:30-

37. doi: 10.1016/j.semcancer.2023.01.006.

[15] J. Chen, "Comparative Analysis of Machine Learning Models for Lung Cancer Prediction," 2023 IEEE International Conference on Image Processing and Computer Applications (ICIPCA), Changchun, China, 2023, pp. 242-246, doi:

10.1109/ICIPCA59209.2023.102577

# APPENDIX

# SOURCE CODE

LUNG CANCER DETECTION USING YOLOV8 (Classification)

import os

import shutil

import random

from sklearn.model\_selection import train\_test\_split

# Define paths to your folders

data\_folders = ['Bengin cases', 'Malignant cases', 'Normal cases']

source\_directory = r'/content/drive/MyDrive/The IQ-OTHNCCD lung cancer dataset'

destination\_directory = r'/content/drive/MyDrive/The IQ-OTHNCCD lung cancer dataset'

# Define ratios for train, validation, and test sets

train\_ratio = 0.7

val\_ratio = 0.15

test\_ratio = 0.15

# Create destination directories for train, validation, and test sets

for folder in ['train', 'val', 'test']:

    for subfolder in data\_folders:

        os.makedirs(os.path.join(destination\_directory, folder, subfolder), exist\_ok=True)

# Split data into train, validation, and test sets for each folder

for folder in data\_folders:

    files = os.listdir(os.path.normpath(os.path.join(source\_directory, folder)))

    random.shuffle(files)

    train\_files, test\_val\_files = train\_test\_split(files, test\_size=(val\_ratio + test\_ratio), random\_state=42)

    val\_files, test\_files = train\_test\_split(test\_val\_files, test\_size=test\_ratio/(test\_ratio + val\_ratio), random\_state=42)

    # Move files to their respective directories

    for file in train\_files:

        shutil.move(os.path.join(source\_directory, folder, file), os.path.join(destination\_directory, 'train', folder))

    for file in val\_files:

        shutil.move(os.path.join(source\_directory, folder, file), os.path.join(destination\_directory, 'val', folder))

    for file in test\_files:

        shutil.move(os.path.join(source\_directory, folder, file), os.path.join(destination\_directory, 'test', folder))

print("Data split successfully.")

!nvidia-smi

import os

HOME = os.getcwd()

print(HOME)

!pip install ultralytics==8.0.196

from IPython import display

display.clear\_output()

import ultralytics

ultralytics.checks()

from ultralytics import YOLO

from IPython.display import display, Image

from google.colab import drive

drive.mount('/content/drive')

!yolo task=classify mode=train model='yolov8n-cls.pt' data='/content/drive/MyDrive/gdrive' epochs=100

!yolo task=classify mode=val split=test batch=1 model=runs/classify/train/weights/best.pt data='/content/drive/MyDrive/gdrive'

!yolo task=classify mode=predict save\_txt=True model=runs/classify/train/weights/best.pt conf=0.25 save=True source=/content/Bengin-case-29-\_jpg.rf.f9afac9e3fc492fc975e6bca98a9ebae.jpg

with open('/content/runs/classify/predict/labels/Bengin-case-29-\_jpg.rf.f9afac9e3fc492fc975e6bca98a9ebae.txt', 'r') as f:

    text = f.read()

print(text)

# Define the path to your text file

file\_path = '/content/runs/classify/predict3/labels/Normal-case-106-\_jpg.rf.6d8694cf4f7f25734e421f073cad92d1.txt'

# Initialize an empty list to store the lines

lines = []

# Open the file and read its content

with open(file\_path, 'r') as file:

    # Read each line and append it to the list

    for line in file:

        lines.append(line.strip())  # strip() removes leading and trailing whitespace

for i in lines:

  myString = i

  myList = myString.split(' ')

  length=len(myList)

  for j in range(length):

    if j==0:

      val=float(myList[j])

      if val>0.00:

        val\_1=myList[j+1]

        if val\_1=='Bengin' or val\_1=='Malignant':

          print('cancerous')

        elif val\_1=='Normal':

          print('Non cancerous')

    break

# Define the path to your text file

file\_path = '/content/runs/classify/predict2/labels/Malignant-case-154-\_jpg.rf.3cbfa978555e05727a27099627b41331.txt'

# Initialize an empty list to store the lines

lines = []

# Open the file and read its content

with open(file\_path, 'r') as file:

    # Read each line and append it to the list

    for line in file:

        lines.append(line.strip())  # strip() removes leading and trailing whitespace

for i in lines:

  myString = i

  myList = myString.split(' ')

  length=len(myList)

  for j in range(length):

    if j==0:

      val=float(myList[j])

      if val>0.00:

        val\_1=myList[j+1]

        if val\_1=='Bengin' or val\_1=='Malignant':

          print('cancerous')

        elif val\_1=='Normal':

          print('Non cancerous')

    break

# Define the path to your text file

file\_path = '/content/runs/classify/predict/labels/Bengin-case-29-\_jpg.rf.f9afac9e3fc492fc975e6bca98a9ebae.txt'

# Initialize an empty list to store the lines

lines = []

# Open the file and read its content

with open(file\_path, 'r') as file:

    # Read each line and append it to the list

    for line in file:

        lines.append(line.strip())  # strip() removes leading and trailing whitespace

for i in lines:

  myString = i

  myList = myString.split(' ')

  length=len(myList)

  for j in range(length):

    if j==0:

      val=float(myList[j])

      if val>0.00:

        val\_1=myList[j+1]

        if val\_1=='Bengin' or val\_1=='Malignant':

          print('cancerous')

        elif val\_1=='Normal':

          print('Non cancerous')

    break

!yolo task=classify mode=predict save\_txt=True show\_labels=True model=runs/classify/train/weights/best.pt conf=0.25 save=True source=/content/Malignant-case-154-\_jpg.rf.3cbfa978555e05727a27099627b41331.jpg

!yolo task=classify mode=predict save\_txt=True model=runs/classify/train/weights/best.pt conf=0.25 save=True source=/content/Normal-case-106-\_jpg.rf.6d8694cf4f7f25734e421f073cad92d1.jpg

Image(filename =f'/content/runs/classify/train/results.png',width=640)

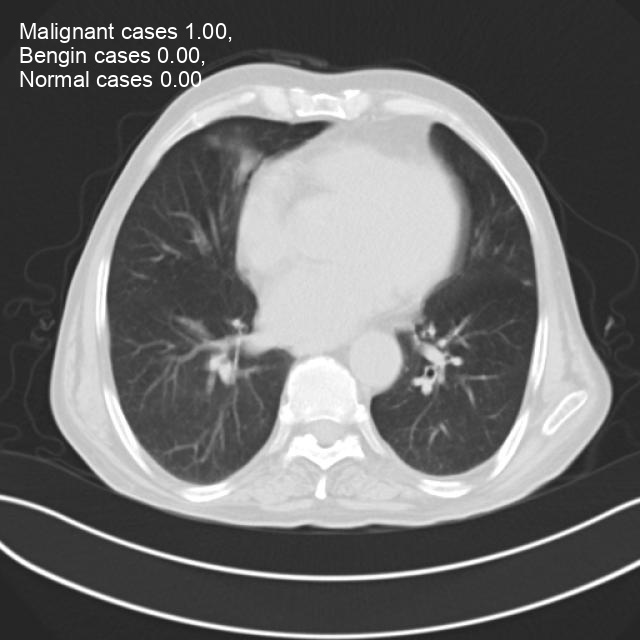
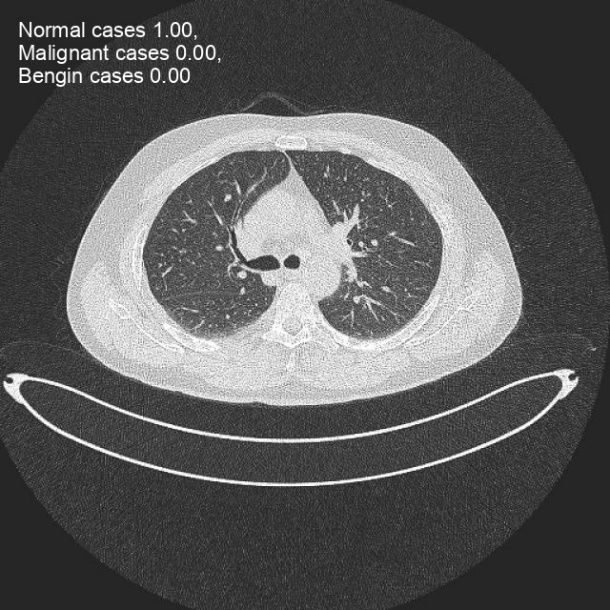
Image(filename =f'/content/runs/classify/train/confusion\_matrix.png',width=640)

Image(filename =f'/content/runs/classify/val/confusion\_matrix.png',width=640)

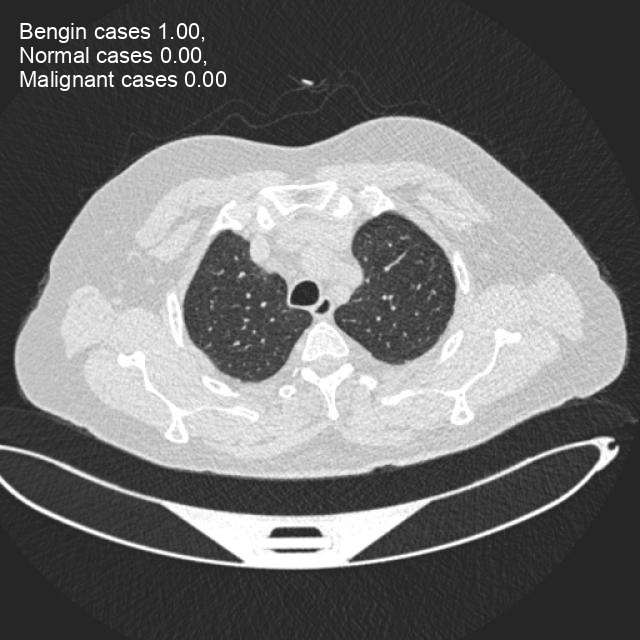
Image(filename =f'/content/runs/classify/predict3/Normal-case-106-\_jpg.rf.6d8694cf4f7f25734e421f073cad92d1.jpg',width=640)

Image(filename =f'/content/runs/classify/predict/Bengin-case-29-\_jpg.rf.f9afac9e3fc492fc975e6bca98a9ebae.jpg',width=640)

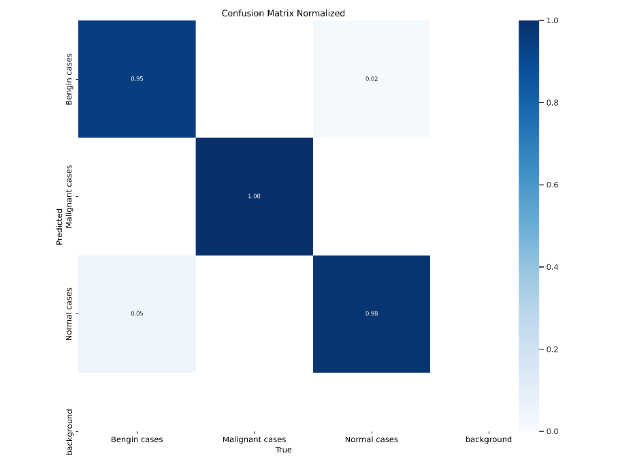
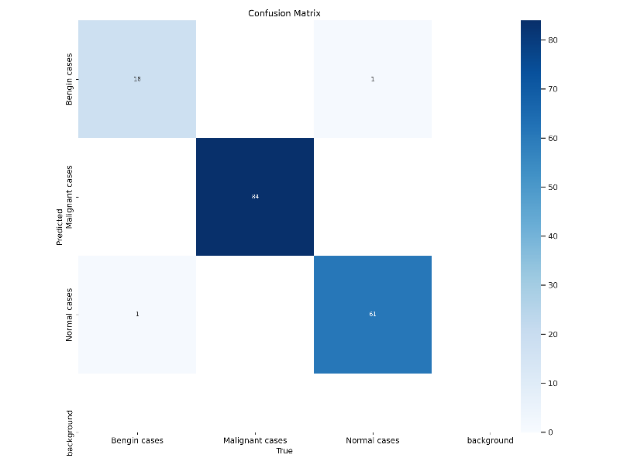
**SCREEN SHOTS**

\



**OUTPUT**

 ****

