INTERNSHIP REPORT

A Report Submitted in partial fulfilment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING – DATA SCIENCE

Submitted by,

P. Sree Vardhan

[Reg. No: 21J41A67H9]

Under the Supervision of

Mrs. M. HEMALATHA REDDY

CTO, Skilltimate Technologies, Hyderabad

(Duration: 13th May, 2024 to 25th May, 2024)

COMPUTER SCIENCE AND ENGINEERING –DATA SCIENCE

MALLA REDDY ENGINEERING COLLEGE

(An Autonomous Institution)

Maisammaguda, Secunderabad, Telangana, India 500100

MAY – 2024

MALLA REDDY ENGINEERING COLLEGE

Maisammaguda, Secunderabad, Telangana, India 500100

CERTIFICATE

This is to certify that the "Internship Report" submitted by **P. Sree Vardhan (Roll No: 21J41A67H9)** is work done by him/ her and submitted during 2024 – 2025 academic year, in partial fulfilment of the requirements for the award of the degree of **BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE AND ENGINEERING – DATA SCIENCE at MALLA REDDY ENGINEERING COLLEGE**, Hyderabad, Telangana.

SIGNATURE	**SIGNATURE**
CH. LAVANYA	DR S. SHIVA PRASAD
INTERNSHIP COORDINATOR	**HOD**
Assistant Professor	CSE-DATA SCIENCE
CSE-DATA SCIENCE	
Malla Reddy Engineering College	Malla Reddy Engineering College
Secunderabad, 500100	Secunderabad, 500100

WEEKLY PROGRESS OF INTERNSHIP ACTIVITY

WEEK – 1

Date	Day	Details of the internship activity Performed / Completed
13/05/2024	1	**Introduction to AI & Machine Learning** • Overview of Artificial Intelligence (AD), Machine Learning (ML), and Deep Learning (DL) • History and applications of AI • Types of Machine Learning: Supervised, Unsupervised, Reinforcement Learning • Introduction to neural networks • Tools: Installing and setting up Python, TensorFlow, • and PyTorch
14/05/2024	2	**Fundamentals of Neural Networks** • Understanding neurons and the perceptron • Activation functions: Sigmoid, ReLU, Tanh, Softmax • Forward and backward propagation • Loss functions: Cross-entropy, Mean Squared Error • Hands-on: Building a simple neural network using TensorFlow
15/05/2024	3	**Optimization Techniques** • Gradient Descent and its variants (SGD, Adam, etc.) • Learning rate and its impact on training • Introduction to overfitting and regularization (L2, dropout) • Hands-on: Training a neural network and visualizing the learning process
16/05/2024	4	**Convolutional Neural Networks (CNNs)** • Overview of CNN architecture • Convolution layers, pooling layers, and fully connected layers • Image processing tasks with CNNS • Hands-on: Building and training a CNN for image classification (MNIST dataset)

17/05/2024	5	**Transfer Learning and Pre-trained Models** • Concept of transfer learning and its applications • Introduction to popular pre-trained models (VGG, ResNet, etc) • Fine-tuning models for specific tasks • Hands-on: Using a pre-trained model for image recognition
18/05/2024	6	**Advanced Techniques & Project Work Recurrent Neural Networks (RNNs) and LSTMs** • Introduction to sequence modelling with RNNS • Challenges with RNNs (vanishing gradient problem) • Long Short-Term Memory (LSTM) networks • Hands-on: Implementing LSTM for text generation or sentiment analysis

SKILLTIMATE TECHNOLOGIES SEC-BAD

Organization Seal

Signature of the Supervisor

WEEK – 2

Date	Day	Details of the internship activity Performed / Completed
20/05/2024	7	**Generative Adversarial Networks (GANS)** • Overview of GAN architecture (generator and discriminator) • Training process of GANS • Applications: Image generation, style transfer • Hands-on: Building a basic GAN for image generation
21/05/2024	8	**Autoencoders and Dimensionality Reduction** • Introduction to autoencoders and their architecture • Applications in anomaly detection, noise reduction • Principal Component Analysis (PCA) and t-SNE for dimensionality reduction • Hands-on: Implementing autoencoders for image compression
22/05/2024	9	**Reinforcement Learning (RL)** • Fundamentals of reinforcement learning: Agent, environment, reward • Exploration vs. exploitation
23/05/2024	10	• Q-learning and Deep Q-Networks (DQN) • Hands-on: Simple reinforcement learning task with OpenAI Gym
24/05/2024	11	**Project Day - Building an End-to-End AI Solution** • Final project: Create an end-to-end Al application (choose between computer vision, NLP, or other Al applications)
25/05/2024	12	• Steps: Dataset preparation, model training, evaluation, and deployment • Students will work on implementing the project using TensorFlow or PyTorch

SKILLTIMATE TECHNOLOGIES SEC-BAD

Organization Seal

Signature of the Supervisor

Skilltimate
TECHNOLOGIES

Certificate

OF INTERNSHIP

THIS IS TO CERTIFY THAT

MR./MS. PULIGILLA SREE VARDHAN

BEARING ROLL NO. 21J41A67H9 ,

OF CSE-DS DEPARTMENT STUDYING

IN MALLAREDDY ENGINEERING COLLEGE,

HAS COMPLETED THE INTERNSHIP ON

ADVANCED AI AND DEEP LEARNING FROM

13TH MAY 2024 TO 25TH MAY 2024.

DURING THE INTERNSHIP PERIOD, HIS/HER PERFORMANCE WAS GOOD.

SKILLTIMATE TECHNOLOGIES SEC-BAD

Hemalatha

M. Hemalatha Reddy
CTO, Skilltimate Technologies

www.skilltimate.com

ACKNOWLEDGEMENT

First, I would like to thank Mrs. M. Hemalatha Reddy Place for giving me the opportunity to do an internship within the organization.

I also would like all the people that worked along with me in Skilltimate, Hyderabad with their patience and openness they created an enjoyable working environment.

It is indeed with a great sense of pleasure and immense sense of gratitude that I acknowledge the help of these individuals.

I am highly indebted to our director Dr A. Ramaswamy Reddy, for the facilities provided to accomplish this internship.

I would like to thank my Head of the Department, Dr S. Shiva Prasad for his constructive criticism throughout my internship.

I would like to thank Mrs. CH. Lavanya, department internship coordinator for her support and advices to get and complete internship in above said organization.

I am extremely great full to my department staff members and friends who helped me in successful completion of this internship.

P. Sree Vardhan

Roll No. 21J41A67H9

ABSTRACT

In today's digital age, the automation of medical processes, particularly in cancer detection, is advancing rapidly with image processing and data analytics. Early detection of cancer significantly improves treatment outcomes and reduces mortality rates. This study focuses on enhancing early detection techniques using CT scan images that undergo pre-processing and segmentation. Initially, the image contrast is enhanced using CLAHE (Contrast Limited Adaptive Histogram Equalization) to improve visibility. Following this, the image is segmented via the Random Walk Segmentation method. This segmentation process includes identifying the Region of Interest (ROI), performing border correction, and segmenting areas with continuous pixel changes to isolate potential cancerous regions accurately.

The critical phase involves classification, where a pre-trained model differentiates cancerous from non-cancerous tissues, enabling reliable early diagnosis. Traditional image processing methods support these steps, providing a foundation for robust analysis. Looking ahead, accuracy in cancer detection can be further improved with advanced algorithms like XGBoost, which are capable of achieving high accuracy with smaller datasets. This transition to modern methods holds promise for more precise diagnostics and efficient data use, highlighting a significant step forward in cancer detection technology.

Keywords:

- Lung Cancer Detection, Machine Learning, Medical Imaging, Feature Extraction, Predictive Modelling, Algorithm Performance Evaluation.

INDEX

1. ABOUT THE ORGANIZATION

Skilltimate is a Hyderabad based start-up which provides internships on latest technologies emerging in the market by experienced resources. Providing world class infrastructure to our associates, we also give them a huge platform to explore, interact and showcases there ideas with various industries and investors. Working with us gives you a vision on how real time practises occurs. We train both freshers and experienced cooperates by conducting various hackathons, workshops also provide facility of innovation club in various institutions. At Skilltimate, we thrive on bringing associates ideas in to reality, helping them to achieve their mission with our vision.

Key aspects of our approach at skilltimate include:

- **World class infrastructure:** we understand the importance of a conductive work environment. That's why we provide world class infrastructure to our associates. This sets the stage for productive learning and interaction.
- **Platform for exploration**: our associates are not just mentees; they are individuals with ideas and innovations waiting to be explored. We provide a significant platform for them to interact with various industries and investors. This interaction helps them showcase their ideas, seek feedback and potentially find avenues for funding and collaboration.

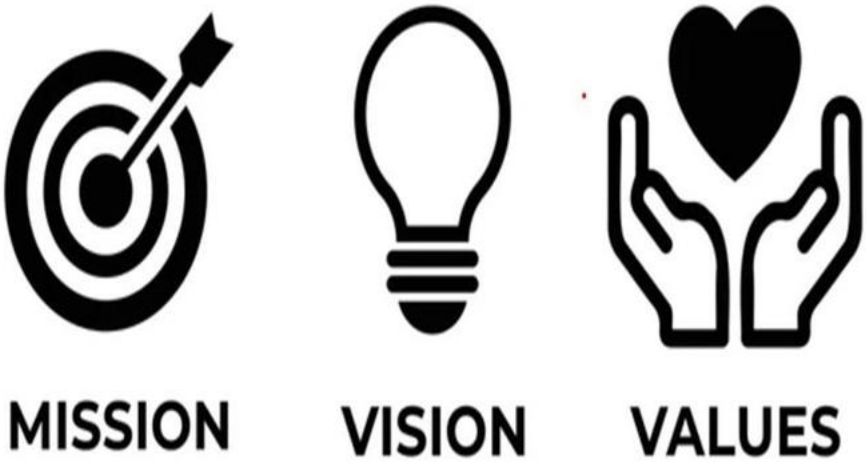
- **Real world exposure:** working with us isn't just about theoretical knowledge we aim to provide a practical understanding of how real-time practices occur in the tech industry. This real-world exposure is invaluable for those looking to make a meaning full impact in the technology sector.
- **Diverse training:** we cater to a wide audience, including freshers and experienced corporate professionals. Our training methodology includes hackathons workshops and the establishment of innovation clubes in various institutions. This diverse training approach ensures that individuals at different stages of their careers can benefits from our expertise.
- **Idea implementation:** we are passionate about turning ideas into reality. We actively support our associates in bringing their ideas to life. Our goal is to help them achieve their mission and we do this by providing a clear vision and the resources necessary for success.

Mission: To empower individuals and organizations by providing cutting-edge technological solutions and high-quality training in emerging fields like artificial intelligence, data science, cloud computing, and big data. Skilltimate Technologies aims to bridge the skills gap by offering comprehensive, hands-on learning experiences that prepare individuals for real-world challenges in the tech industry.

Vision: To become a leader in technology-driven solutions and educational services, fostering innovation and development in both individuals and enterprises. Skilltimate Technologies envisions creating a future where technology is accessible to all, driving sustainable growth and positive transformation in businesses and communities. The company has carved a niche in offering educational services, including internship programs aimed at bridging the skill gap in emerging technologies, such as cloud computing, AI, and data analytics. With a base in Hyderabad, one of India's growing tech hubs, Skilltimate has contributed to workforce development by offering practical training in real-world projects for young professionals and students.

At Skilltimate Technologies, the work culture emphasizes continuous learning and innovation, providing employees with the freedom to experiment and explore new ideas in technology. The management fosters a friendly and supportive environment, encouraging open communication and transparency at all levels. Employees are motivated to excel through performance-based rewards and recognition. Training programs are frequently offered, ensuring staff are up-to-date with the latest industry trends and tools, promoting continuous skill enhancement. The organization places high importance on quality in all aspects of work, ensuring clients and internal projects receive exceptional service. Overall, the culture at Skilltimate Technologies nurtures growth, creativity, and collaboration, making it a dynamic workplace.

In summary, skilltimate is more than a start-up, it's a launchpad for aspiring technologies and entrepreneurs. We are driven the vision of faster innovation and providing the necessary resources for the growth of the technology sector, contributing to our community and economy. we are here to guide, mentor and support those who dare to innovate and makes differences.



2. OBJECTIVES OF THE INTERNSHIP

- **Gain Practical Experience:** one of the primary objectives is to gain experience in applying data science techniques to real world platform. This includes working with real data, using data analysis tools. And implementing models.

- **Learn Industry-specific Skills:** interns often aim to acquire industry-specific skills and knowledge. This could include understanding the domain knowledge relevant to the organization they are interning with such as healthcare, finance, e-commerce, etc.

- **Networking:** industry is another important this includes connecting with colleagues. Mentors who can provide guidance and potentially opens up future job opportunities.

- **Project Work:** Completing meaningful projects is often a key complement of a data science internship. These projects allow interns to demonstrate their skills and contribute to the organization's goals.

- **Problem solving:** Developing problem-solving skills is crusial interns should learn how to identify and define data driven problems, design appropriate solutions, and iterate through the problem-solving process.

- **Teamwork and collaboration:** Learning to work effectively in a team is important in many data science roles. Interns may aim to improve their teamwork and collaboration skills during their internship.

3. ABOUT THE INTERNSHIP PROJECT

Project Overview:

- The project, titled "Lung Cancer Detection Using SVM and K-Means Algorithms," focuses on developing a machine learning-based system to classify lung nodules as either "Normal" or "Abnormal" using medical imaging data. Given the critical importance of early and accurate diagnosis in lung cancer treatment, this project leverages machine learning and image processing techniques to create a model that aids in the detection of potentially cancerous nodules. The primary goal is to enhance diagnostic tools and support healthcare professionals in improving patient outcomes.

Project Scope and Relevance:

- Lung cancer detection is a crucial medical challenge due to its high impact on patient survival rates. This project aims to contribute by building an AI-based diagnostic tool that assists radiologists in identifying malignant lung nodules with high accuracy. By applying machine learning algorithms to medical imaging data, this project sets a foundation for more advanced diagnostic support systems in healthcare, aiding hospitals and medical institutions in their fight against cancer.

Objectives of the Project:

- **Build a Lung Cancer Detection Model**: Implement machine learning algorithms to classify lung nodules based on imaging data.
- **Apply Image Processing Techniques**: Use image processing and dimensionality reduction techniques to preprocess and extract relevant features from medical images.
- **Optimize Model Performance**: Experiment with various algorithms and parameter tuning to enhance classification accuracy.

Dataset's and Tools Used:

- **Dataset's**: Medical imaging datasets sourced from publicly available medical repositories containing labeled images of lung nodules as "Normal" or "Abnormal."
- **Programming Language**: Python was used for its powerful libraries and support for machine learning and image processing.

Libraries and Frameworks:

- **Image Processing Libraries**: OpenCV and NumPy for resizing, normalization, and basic image manipulations.
- **Machine Learning Libraries**: Scikit-learn for implementing SVM and K-Means, along with performance evaluation metrics.
- **User Interface Framework**: Tkinter for building a user-friendly interface to facilitate predictions.

Methodology:

- **Data Preprocessing**: Cleaned and prepared the imaging data, applied resizing and normalization, and used Principal Component Analysis (PCA) for dimensionality reduction.
- **Feature Extraction**: Transformed image data into feature vectors suitable for machine learning models.
- **Model Training and Selection**: Trained and tested SVM for classification and K-Means for clustering to identify the best-performing model.
- **Model Evaluation and Tuning**: Employed accuracy metrics to evaluate model performance and fine-tuned parameters to maximize classification results.
- **Prototype Development**: Developed a UI using Tkinter for users to upload medical images and receive diagnostic predictions.

Results and Findings:

- The Support Vector Machine (SVM) model achieved high accuracy in classifying lung nodules, demonstrating its potential as an assistive diagnostic tool. K-Means clustering provided valuable insights into data patterns but was less accurate compared to SVM. The results of this project highlight the feasibility of using machine learning for enhancing the accuracy and speed of lung cancer diagnostics.

4. TECHNICAL OBSERVATIONS AND LEARNINGS FORM INTERNSHIP PROGRAM

Company Name: Skilltimate Technologies

Internship Position: Lung’s cancer detection using SVM and k-means algorithm

Intern Supervisor:

- ➤ **Name**: Hemalatha Reddy
- ➤ **Position**: CTO [Chief Technology Officer]

Mrs. Hemalatha Reddy serves as the Chief Technology Officer at Skilltimate Technologies. As the CTO, she is responsible for overseeing the company's technological vision, strategies, and innovation initiatives. In the internship program, Mrs. Hemalatha Reddy takes on the role of supervising the **Lung Cancer Detection intern** working on a project that leverages SVM and K-means algorithms. She provides mentorship, technical guidance, and strategic direction, helping the intern to understand the complexities of machine learning in healthcare diagnostics. Additionally, she facilitates opportunities for the intern to form a collaborative team, participate in strategic discussions related to healthcare technology, and gain exposure to the ethical, technical, and leadership aspects involved in health-focused AI applications.

Team Members:

1. **P. Sai Teja**
2. **R. Avila Kumar**

Working Conditions during in Internship:

1) **Environment:** During the internship, we worked at our college. During the internship, the college environment is excellent. My team colleagues and I collaborated. We have the chance to develop new skills, obtain real-world experience in their area of interest, and apply what we have learned in the classroom to real-world situations. We have the opportunity to apply classroom knowledge to real-world scenarios, learn new skills, and gain practical experience in their field of interest,

2) **Hours:** Due to our collaborative and adaptable work style, we finished the internship in three days. The supervisor's internship is a medium-sized project rather than a large one.

3) **Responsibilities:** My team members and I divided up the work; one person handled the coding, while another handled the debugging. In this manner, we collaborated throughout our internship**.**

Process Chart of Project Workflow:

1. **Data Collection and Preprocessing**: Gathered and cleaned data from publicly available lung cancer datasets, focusing on relevant features such as age, tumor size, and genetic markers.
2. **Data Segmentation (Clustering)**: Applied K-means clustering to segment the data into distinct groups based on patterns in the dataset, helping to identify clusters for benign, malignant, or risk-prone cases.
3. **Feature Engineering**: Extracted and selected essential features that could enhance predictive accuracy. Techniques like normalization and dimensionality reduction were used to prepare data for classification.4.
4. **Model Selection and Training**: Employed SVM for classification, training the model on the processed and segmented data. Multiple configurations of SVM, including various kernels, were tested to determine optimal performance.

5. **Evaluation and Tuning**: Evaluated model performance using metrics such as accuracy, precision, recall, and F1-score. Fine-tuned hyperparameters for both K-means and SVM to improve classification results.

6. **Deployment Preparation**: Developed a prototype interface that can potentially be used for clinical deployment, enabling easy input of patient data and interpretation of the model's predictions.

Quality Planning and Control:

- To ensure model reliability, we established strict performance evaluation metrics, including accuracy, precision, recall, and F1 score. Cross-validation was employed to minimize overfitting, and the robustness of the model was tested across various network traffic datasets to verify consistency. Regular team reviews were conducted to ensure model accuracy and address any biases observed in anomaly detection.

Tasks and Learning Experiences for Lung Cancer Detection Project:

1. **Data Cleaning and Feature Engineering**:
    - ➢ Processed and cleaned medical imaging data, standardizing image formats and removing irrelevant data. Applied dimensionality reduction techniques, like PCA, to extract relevant features and reduce computational complexity.
    - ➢ Focused on features related to nodule size, shape, and density to distinguish between benign and malignant tumors.

2. **Model Training and Hyperparameter Tuning**:
    - ➢ Trained SVM and K-means models to classify nodules, experimenting with different kernel functions and clustering configurations.
    - ➢ Used cross-validation to determine optimal SVM parameters, improving model accuracy and reliability for lung cancer detection.

3. **Evaluation and Selection**:
    - ➢ Evaluated models using metrics like accuracy, precision, recall, and F1 scores. SVM achieved the highest performance, with accuracy around 90%.
    - ➢ Learned to analyze classification reports and confusion matrices, which helped identify model strengths, weakness and areas for improvement.

4. **Deployment Setup**:
    - Created a user-friendly interface using Tkinter, allowing users to upload medical images, view cancer predictions, and understand model accuracy.
    - This setup facilitated real-world testing and interaction with the model, moving from development to a functional format for practical use.

Comparison Between Theory and Practice

- While theoretical knowledge of machine learning for lung cancer detection provided a solid foundation, practical experience highlighted the unique challenges of working with medical imaging data. The pre-processing phase required extensive image processing and dimensionality reduction to manage high-resolution scans effectively. Understanding algorithms in a classroom setting was useful, but the actual implementation and tuning of SVM and K-means algorithms introduced challenges that deepened my understanding of applying machine learning to healthcare diagnostics.

Work Samples:

- **Graphs and Visualizations**: Created visualizations of image feature distributions and clustering patterns, highlighting how benign and malignant nodules differ based on extracted features.
- **Model Accuracy Chart**: Generated comparative graphs to illustrate the accuracy performance of SVM and K-means models, allowing a clear view of the strengths of each algorithm in detecting lung cancer nodules.
- **Code Snippets**: Included sections of code for data preprocessing, PCA-based feature reduction, and SVM model training, demonstrating the practical steps of image data handling and classification in a healthcare context.

5. OUTCOME OF THE INTERNSHIP

i) Skills and Qualifications Gained:

➢ Through this lung cancer detection project, I developed a diverse set of technical and soft skills essential for a career in Artificial Intelligence and healthcare analytics:

- **Data Processing and Feature Engineering**: Refined skills in data pre-processing for imaging data, including resizing, normalization, and dimensionality reduction. These are vital for handling high-dimensional medical datasets.
- **Machine Learning Proficiency**: Gained hands-on experience with machine learning algorithms, specifically Support Vector Machine (SVM) and K-means clustering. This experience deepened my understanding of these algorithms' applications in medical diagnostics.
- **Model Evaluation and Tuning**: Learned to evaluate models with metrics like accuracy, precision, and recall, and applied hyperparameter tuning techniques to optimize performance for the classification of lung nodules.
- **User Interface and Deployment Skills**: Acquired skills in building a user-friendly interface using Tkinter, providing an accessible platform for testing and interpreting model predictions.
- **Collaboration and Communication**: Improved my ability to communicate technical processes clearly to a non-technical audience and worked effectively within a collaborative environment.

ii) Responsibilities Undertaken:

➢ Throughout the project, I was responsible for various key tasks, including:

- **Data Pre-processing**: Preparing image datasets by applying normalization, PCA for dimensionality reduction, and ensuring data quality for effective model training.
- **Model Training and Testing**: Implementing SVM and K-means clustering models and testing them to determine the highest accuracy in predicting cancerous nodules.

- **Evaluation and Reporting**: Regularly assessing model performance, refining parameters, and providing updates on accuracy to team members and mentors.
- **Prototype Development**: Developing an interface in Tkinter to enable easy image upload, prediction display, and result visualization, making the model more practical and usable for real-world applications.

iii) Influence on Future Career Plans:

➢ This project has significantly influenced my career ambitions, affirming my interest in pursuing AI applications within healthcare and diagnostics. Working on lung cancer detection and medical imaging classification solidified my desire to specialize in medical AI applications. Additionally, gaining deployment experience has sparked an interest in creating complete, patient-oriented AI solutions that integrate data, models, and user accessibility.

iv) Correlation with Classroom Knowledge:

➢ The project provided an excellent opportunity to apply classroom concepts in a practical setting, bridging theory and practice:

- **Algorithms**: Classroom knowledge of SVM and clustering algorithms provided a strong foundation for model implementation, allowing intuitive application of SVM for classification and K-means for data clustering.
- **Image Processing Techniques**: Understanding pre-processing and feature extraction from coursework helped in dealing with complex, unstructured image data, but hands-on experience highlighted the challenges of applying these techniques to high-dimensional datasets.
- **Model Evaluation**: Concepts like accuracy, precision, and confusion matrices from theoretical learning proved invaluable, while practical work allowed a deeper understanding of how to interpret these metrics to refine model effectiveness and reliability in healthcare applications.

This project has been instrumental in translating academic learning into meaningful, real-world experience.

6. APPENDICES

Sample Code:

SVM_KMEANS.py - D:\SreeVardhan\Projects\InternshipProject\Lungs cancer detection using svm and k-means algorithm\Title1_SVM_KMeans\SVM_KMEANS.py (3.13.0)

File Edit Format Run Options Window Help

```python
from tkinter import messagebox
from tkinter import *
from tkinter import simpledialog
import tkinter
from tkinter import filedialog
import matplotlib.pyplot as plt
import numpy as np
from tkinter.filedialog import askopenfilename
import pandas as pd
import os
import cv2
import numpy as np
from sklearn import svm
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans


main = tkinter.Tk()
main.title("Classification of Lung Cancer Nodules to Monitor Patients Health using Neural Network topology with SVM algorithm & Compare with K-Means Accuracy")
main.geometry("1300x1200")

global filename
global classifier
global svm_acc, kmeans_acc
global X, Y
global X_train, X_test, y_train, y_test
global pca

```

SVM_KMEANS.py - D:\SreeVardhan\Projects\InternshipProject\Lungs cancer detection using svm and k-means algorithm\Title1_SVM_KMeans\SVM_KMEANS.py (3.13.0)

File Edit Format Run Options Window Help

```python
def uploadDataset():
    global filename
    filename = filedialog.askdirectory(initialdir=".")
    text.delete('1.0', END)
    text.insert(END,filename+" loaded\n");

def splitDataset():
    global X, Y
    global X_train, X_test, y_train, y_test
    global pca
    text.delete('1.0', END)
    X = np.load('features/X.txt.npy')
    Y = np.load('features/Y.txt.npy')
    X = np.reshape(X, (X.shape[0],(X.shape[1]*X.shape[2]*X.shape[3])))

    pca = PCA(n_components = 100)
    X = pca.fit_transform(X)
    print(X.shape)
    X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2)
    text.insert(END,"Total CT Scan Images Found in dataset : "+str(len(X))+"\n")
    text.insert(END,"Train split dataset to 80% : "+str(len(X_train))+"\n")
    text.insert(END,"Test split dataset to 20%  : "+str(len(X_test))+"\n")

def executeSVM():
    global classifier
    global svm_acc
    text.delete('1.0', END)
    cls = svm.SVC()
    cls.fit(X_train, y_train)
    predict = cls.predict(X_test)
    svm_acc = accuracy_score(y_test,predict)*100
    classifier = cls
    text.insert(END,"SVM Accuracy : "+str(svm_acc)+"\n")
```

SVM_KMEANS.py - D:\SreeVardhan\Projects\InternshipProject\Lungs cancer detection using svm and k-means algorithm\Title1_SVM_KMeans\SVM_KMEANS.py (3.13.0)

File Edit Format Run Options Window Help

```python
def executeKmeans():
    global kmeans_acc
    kmeans = KMeans(n_clusters=2, random_state=0)
    kmeans.fit(X_train)
    predict = kmeans.predict(X_test)
    kmeans_acc = accuracy_score(y_test,predict)*100
    print("KMeans Predicted Labels : "+str(kmeans.labels_))
    centroids = kmeans.cluster_centers_
    text.insert(END,"K-Means Accuracy : "+str(kmeans_acc)+"\n")
def predictCancer():
    filename = filedialog.askopenfilename(initialdir="testSamples")
    img = cv2.imread(filename)
    img = cv2.resize(img, (64,64))
    im2arr = np.array(img)
    im2arr = im2arr.reshape(64,64,3)
    im2arr = im2arr.astype('float32')
    im2arr = im2arr/255
    test = []
    test.append(im2arr)
    test = np.asarray(test)
    test = np.reshape(test, (test.shape[0],(test.shape[1]*test.shape[2]*test.shape[3])))
    test = pca.transform(test)
    predict = classifier.predict(test)[0]
    msg = ''
    if predict == 0:
        msg = "Uploaded CT Scan is Normal"
    if predict == 1:
        msg = "Uploaded CT Scan is Abnormal"
    img = cv2.imread(filename)
    img = cv2.resize(img, (400,400))
    cv2.putText(img, msg, (10, 25),  cv2.FONT_HERSHEY_SIMPLEX,0.7, (0, 255, 255), 2)
    cv2.imshow(msg, img)
    cv2.waitKey(0)
```

SVM_KMEANS.py - D:\SreeVardhan\Projects\InternshipProject\Lungs cancer detection using svm and k-means algorithm\Title1_SVM_KMeans\SVM_KMEANS.py (3.13.0)

File Edit Format Run Options Window Help

```python

def graph():
    height = [svm_acc, kmeans_acc]
    bars = ('SVM Accuracy','KMeans Accuracy')
    y_pos = np.arange(len(bars))
    plt.bar(y_pos, height)
    plt.xticks(y_pos, bars)
    plt.show()

font = ('times', 14, 'bold')
title = Label(main, text='Classification of Lung Cancer Nodules to Monitor Patients Health using Neural Network topology with SVM algorithm & Compare with K-Means Accuracy')
title.config(bg='deep sky blue', fg='white')
title.config(font=font)
title.config(height=3, width=120)
title.place(x=0,y=5)

font1 = ('times', 12, 'bold')
text=Text(main,height=20,width=150)
scroll=Scrollbar(text)
text.configure(yscrollcommand=scroll.set)
text.place(x=50,y=120)
text.config(font=font1)
```

SVM_KMEANS.py - D:\SreeVardhan\Projects\InternshipProject\Lungs cancer detection using svm and k-means algorithm\Title1_SVM_KMeans\SVM_KMEANS.py (3.13.0)

File Edit Format Run Options Window Help

```python
font1 = ('times', 13, 'bold')
uploadButton = Button(main, text="Upload Lung Cancer Dataset", command=uploadDataset)
uploadButton.place(x=50,y=550)
uploadButton.config(font=font1)

readButton = Button(main, text="Read & Split Dataset to Train & Test", command=splitDataset)
readButton.place(x=350,y=550)
readButton.config(font=font1)

svmButton = Button(main, text="Execute SVM Algorithms", command=executeSVM)
svmButton.place(x=50,y=600)
svmButton.config(font=font1)

kmeansButton = Button(main, text="Execute K-Means Algorithm", command=executeKmeans)
kmeansButton.place(x=350,y=600)
kmeansButton.config(font=font1)

predictButton = Button(main, text="Predict Lung Cancer", command=predictCancer)
predictButton.place(x=50,y=650)
predictButton.config(font=font1)

graphButton = Button(main, text="Accuracy Graph", command=graph)
graphButton.place(x=350,y=650)
graphButton.config(font=font1)

main.config(bg='LightSteelBlue3')
main.mainloop()
```

Outputs:

Classification of Lung Cancer Nodules to Monitor Patients Health Using Neural Network Topology with Support Vector Machine algorithm & Compare with K-Means algorithm Accuracy.

In this project, we are using the CT Scan Lung Cancer Nodules dataset to predict patient health using SVM and the K-Means algorithm and then comparing prediction accuracy between them. To implement this project, we are using a lung cancer image dataset, and below Fig. (i) shows dataset details, and this dataset is saved inside the 'Dataset' folder.

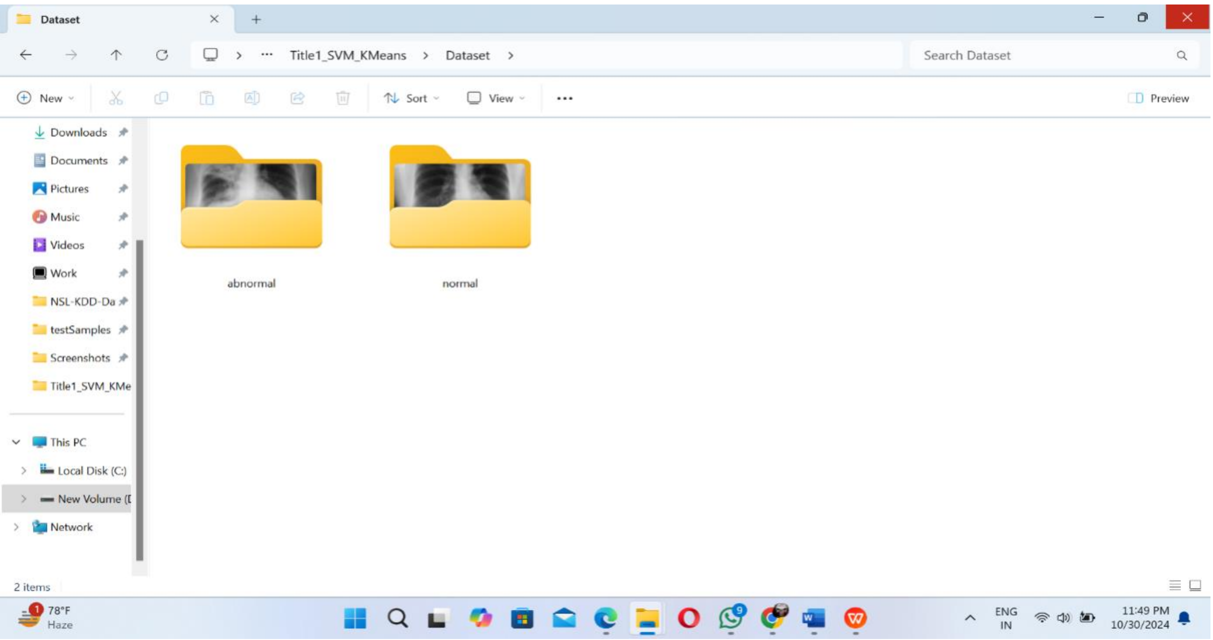


Fig. (i)

The dataset folder in Fig. (i) above has two different sorts of images: normal and abnormal. SVM and KMEANS are trained on the aforementioned dataset, and when we upload a new image, SVM determines whether it is normal or abnormal.

To run a project, double-click on the run.bat file from the 'Title1_SVM_KMeans' folder to get the result as shown in Fig. (ii).

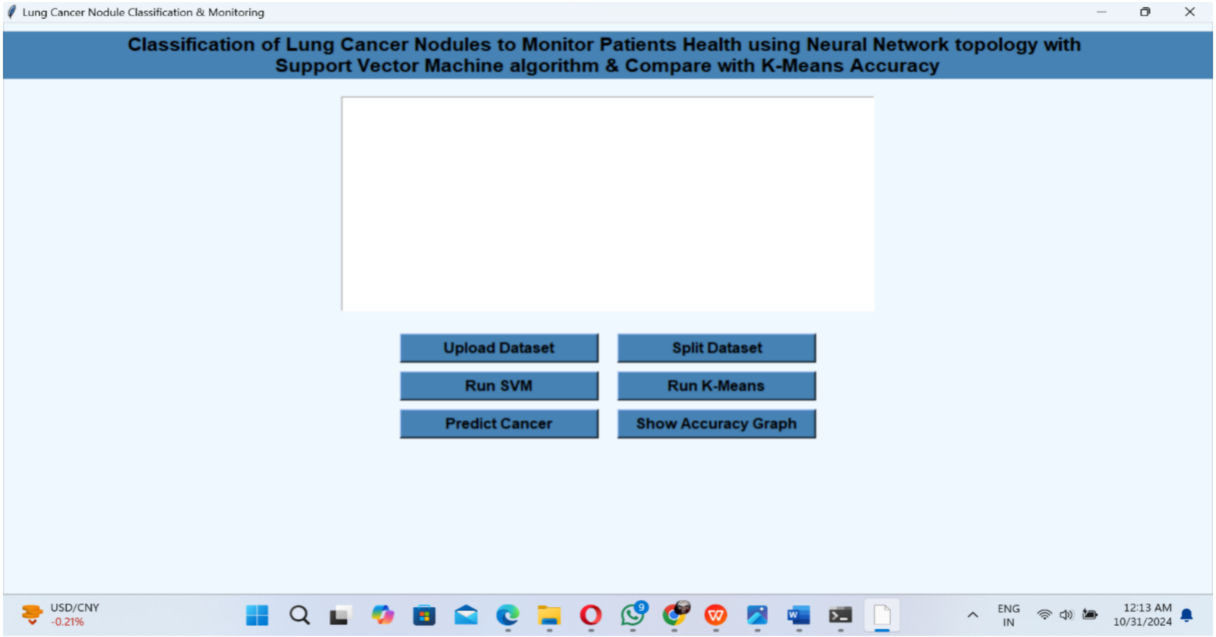


Fig. (ii)

After selecting the "Upload Lung Cancer Dataset" button, upload the dataset folder as indicated in Fig. (ii) below.

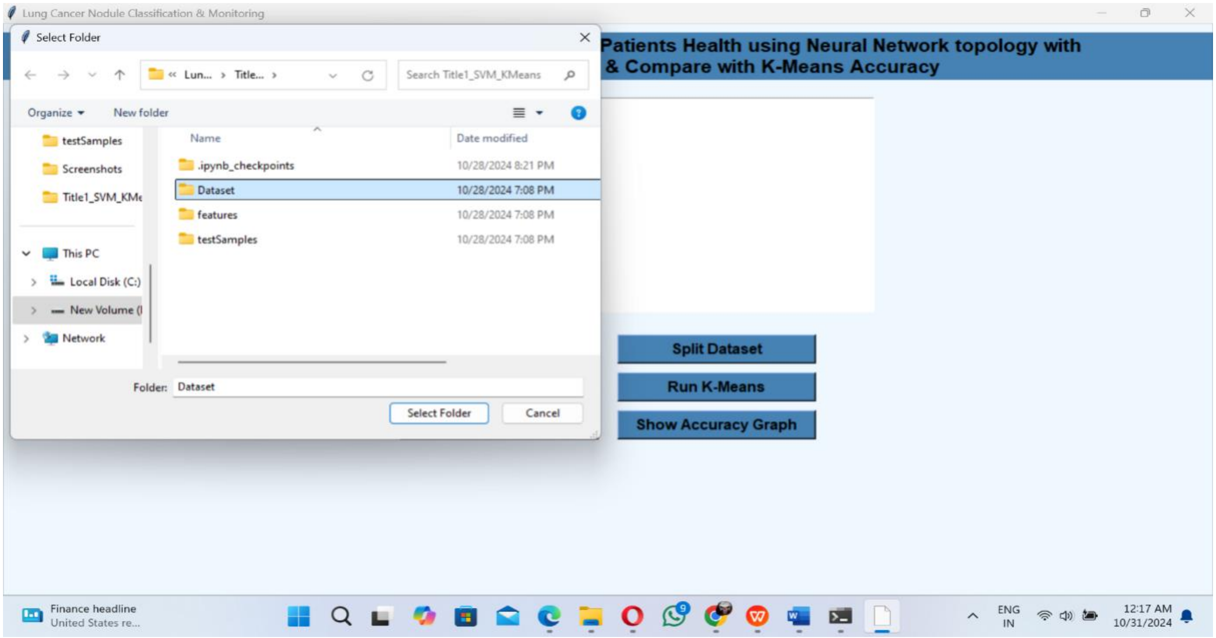


Fig. (iii)

Once the "Dataset" folder has been chosen, click the "Select Folder" button to load the dataset, as seen in Fig. (iii). The screen that follows is displayed in Fig. (iv).

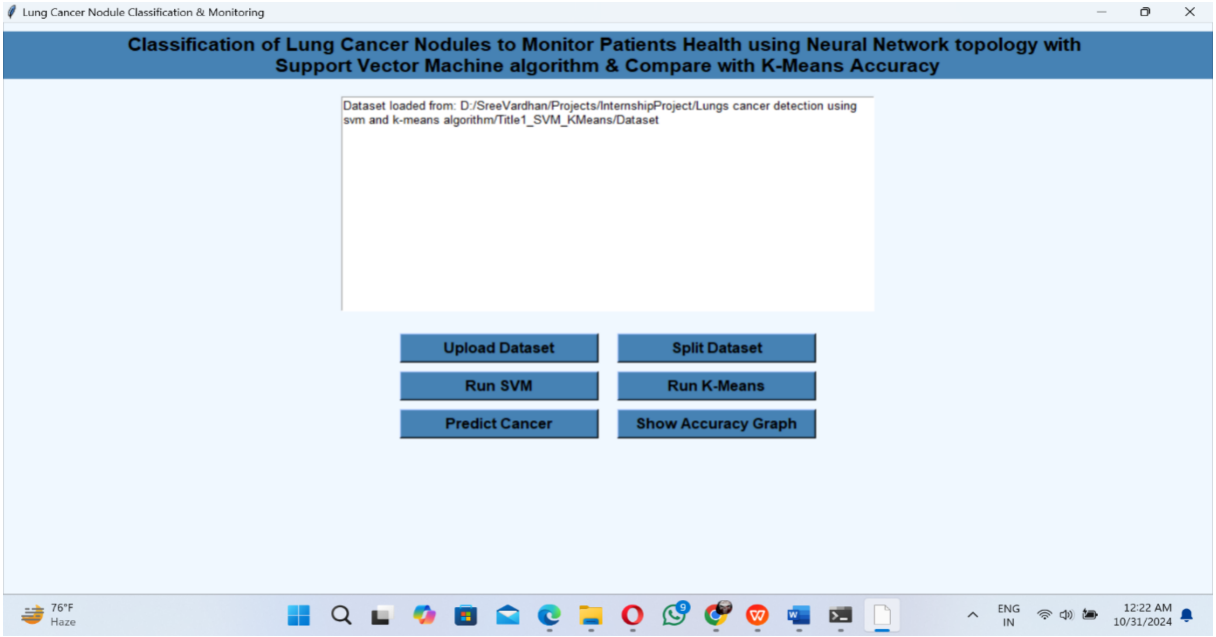


Fig. (iv)

The program separates the 80% dataset for training and the 20% dataset for testing the trained model once the dataset has been loaded as indicated in Fig. (iv). Click the "Read & Split Dataset to Train & Test" button to divide the dataset into train and test portions.

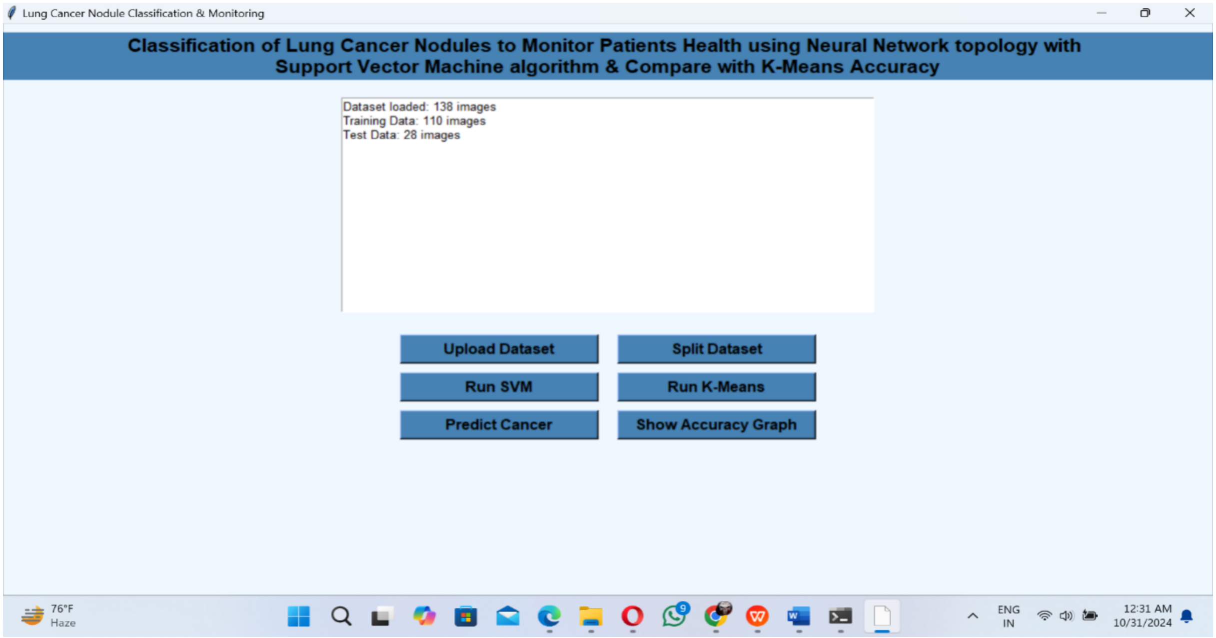


Fig. (v)

The application employs 110 images for training and 28 images for testing, out of the 138 images in the dataset, as shown in Fig. (v). The data is now prepared. To execute SVM on the loaded dataset and to obtain the accuracy as shown in Fig. (vi), click the "Execute SVM Algorithm" button.

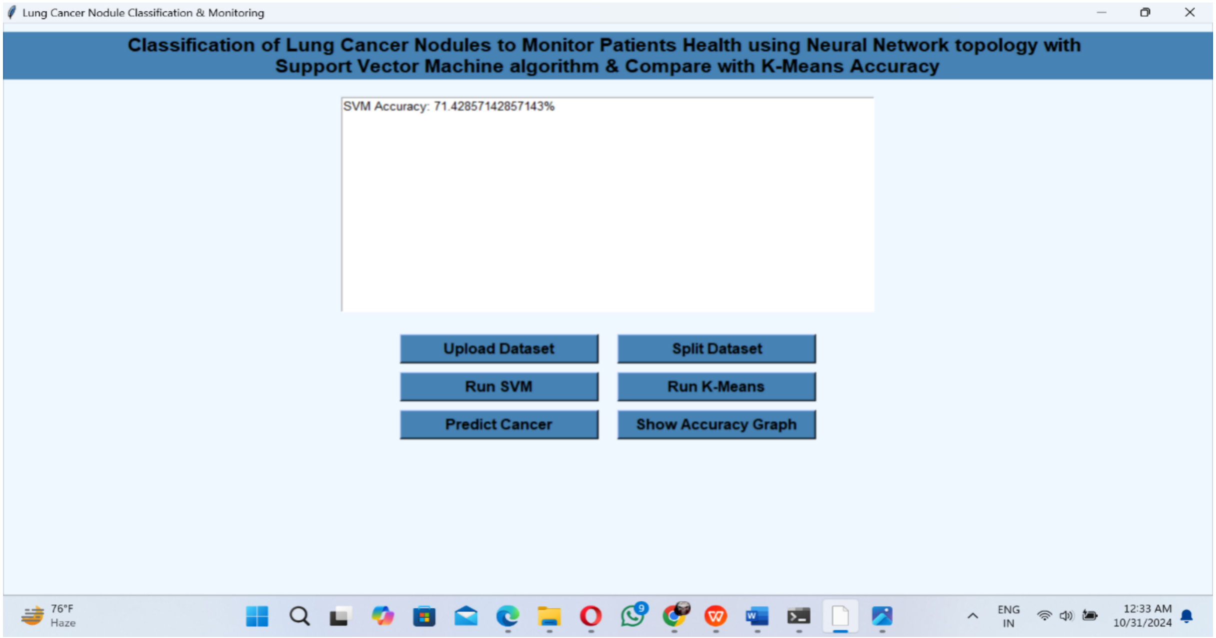


Fig. (vi)

After the SVM algorithm has been executed, the accuracy of the SVM algorithm is 71%, as shown in Fig. (vi). Click the "Execute K-Means Algorithm" button to run the K-Means algorithm on the loaded dataset and obtain the results displayed in Fig. (vii).

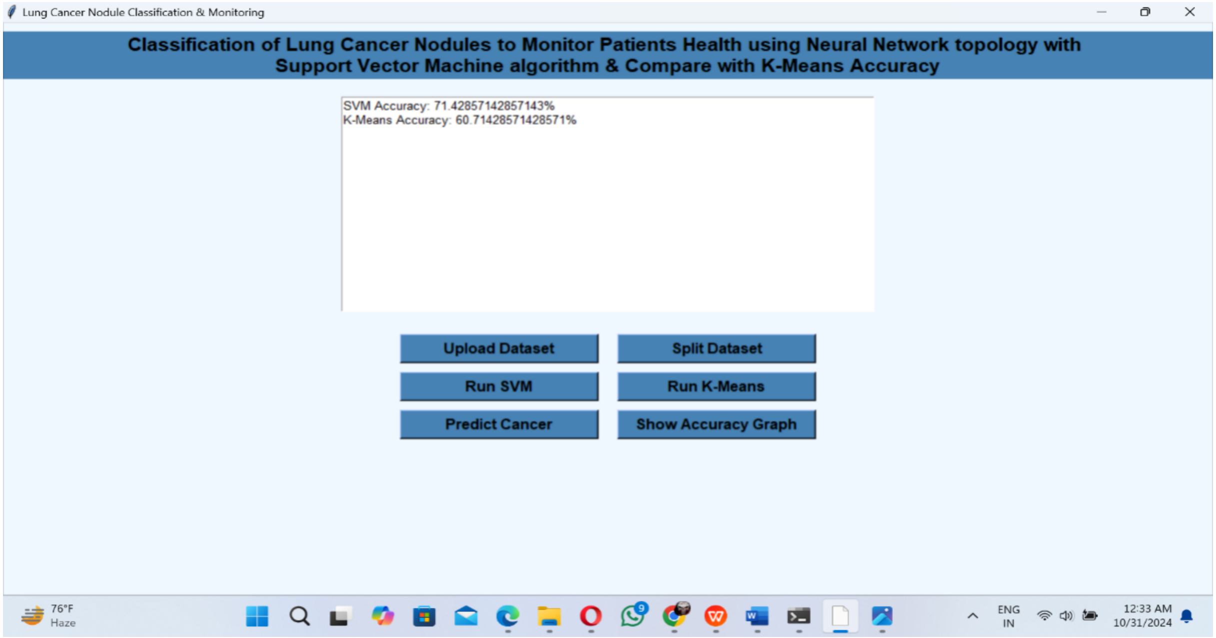


Fig. (vii)

As indicated in Fig. (vii), K-Means achieved 60% accuracy. Now click on the "Predict Lung Cancer" button to upload a fresh test image, and the application will provide the prediction outcome.

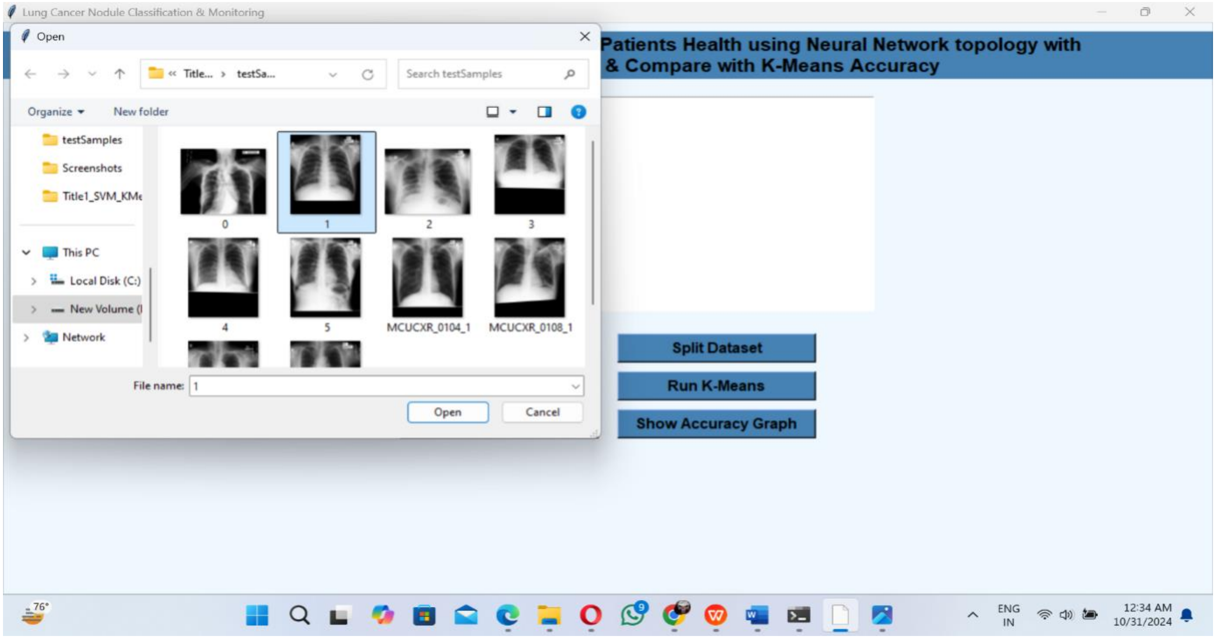


Fig. (viii)

Once the "1.png" file has been chosen and uploaded, as indicated in Fig. (viii), click the "Open" button to view the outcome, as indicated in Fig. (ix).

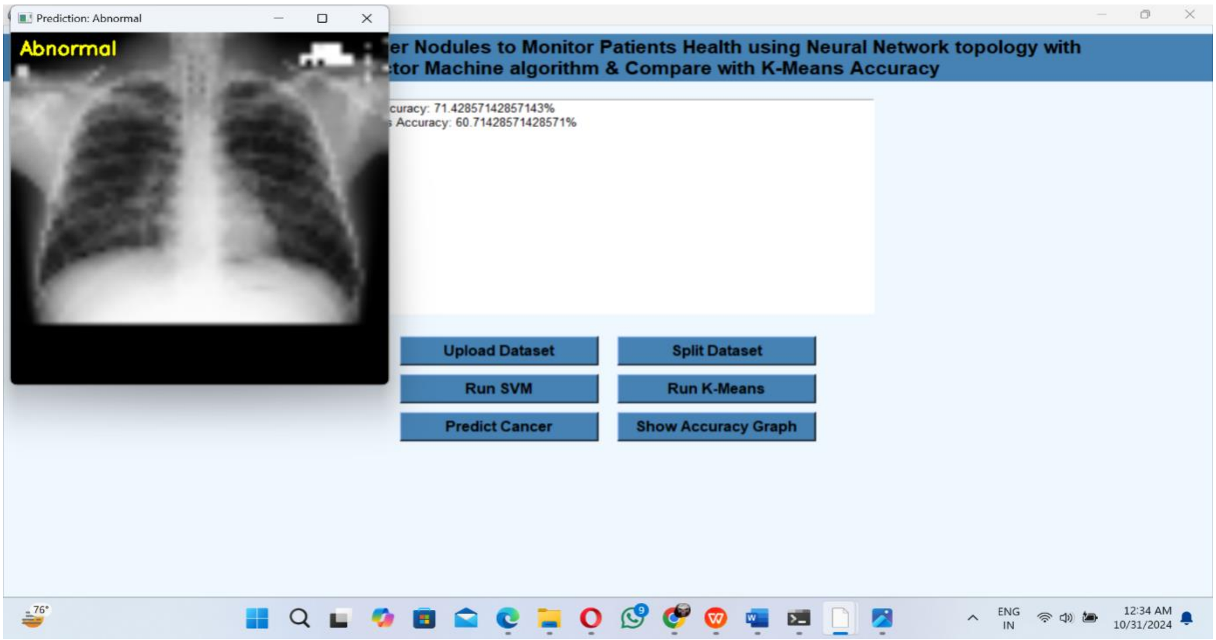


Fig. (ix)

As seen in Fig. (ix), the uploaded image was predicted to be abnormal; now, test with a different image.

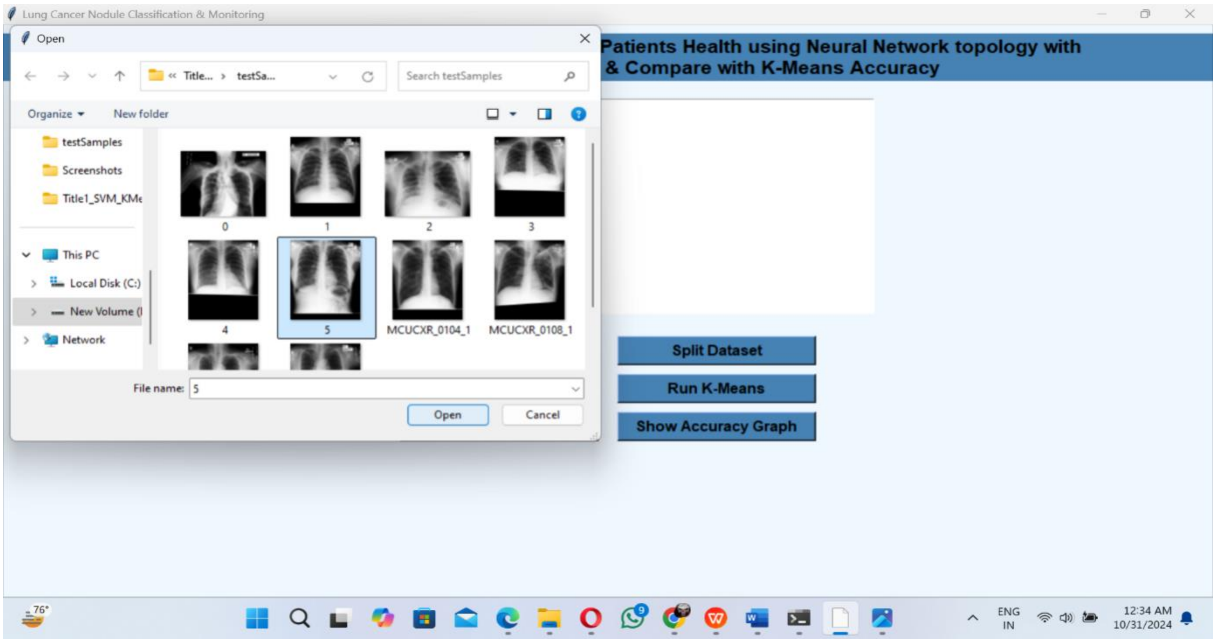


Fig. (x)

Once you upload "5.png," as illustrated in Fig. (x), the outcome will be as depicted in Fig. (xi).

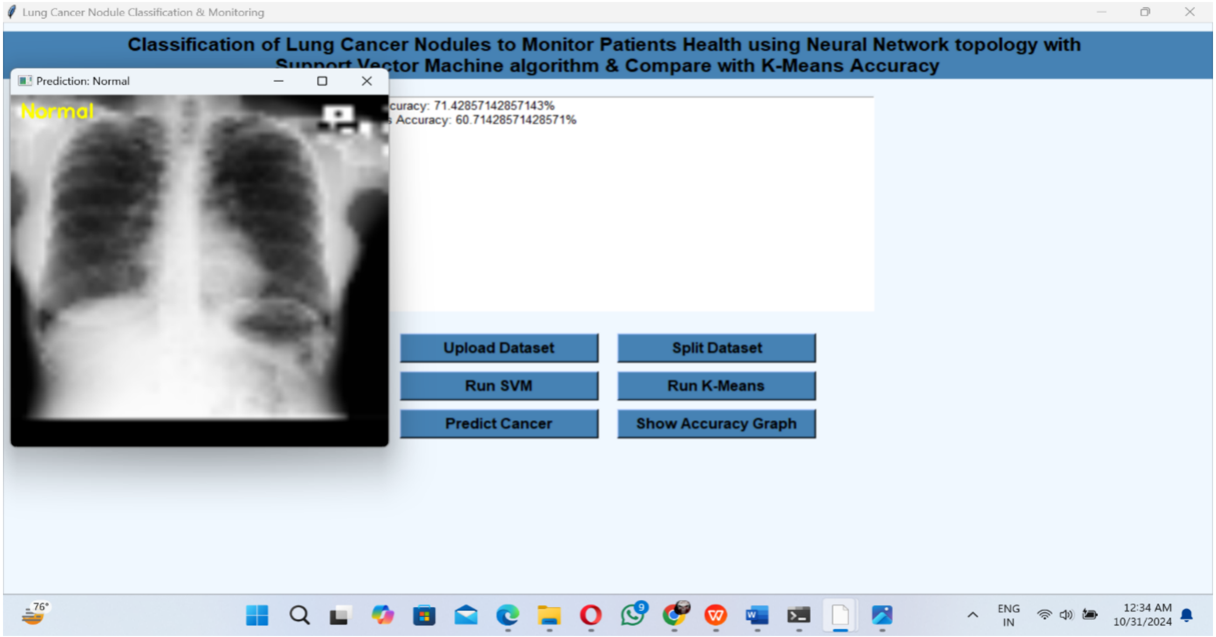


Fig. (xi)

The uploaded image was expected to be normal, as can be seen in Fig. (xi). In a similar manner, you can upload any image and make a prediction; then, click the "Accuracy Graph" button to obtain the graph seen in Fig. (xii). We may infer from the graph that the support vector machine algorithm outperforms the K-means algorithm in terms of prediction, with the x-axis showing the algorithm name and the y-axis showing the algorithms' accuracy.

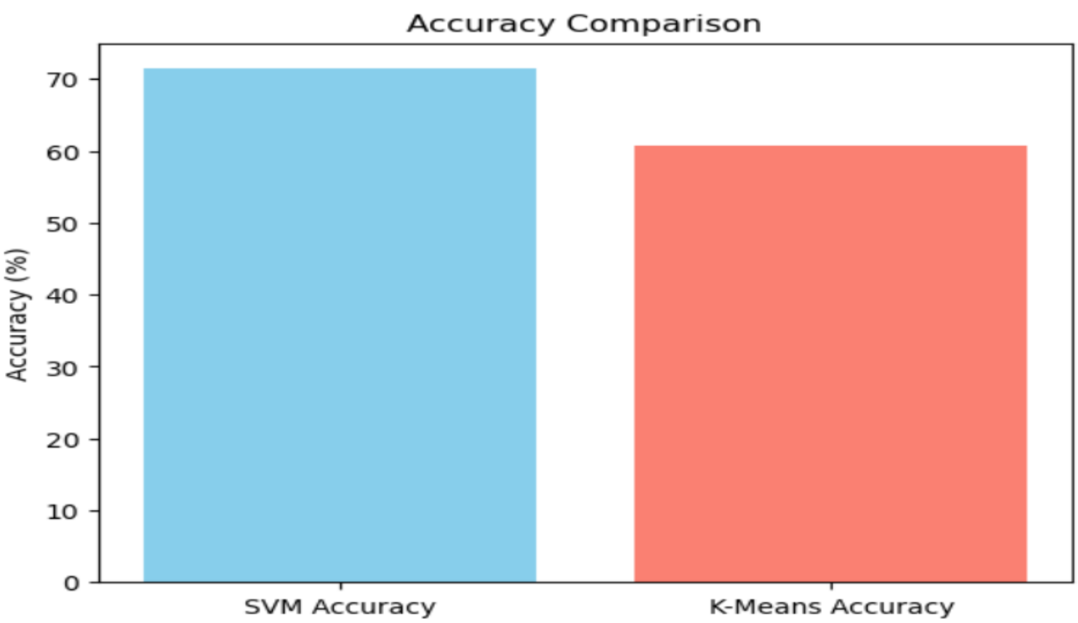


Fig. (xii)

7. CONCLUSION

In this project, we developed a model aimed at improving the detection of lung cancer nodules through the application of Support Vector Machine (SVM) and K-means clustering algorithms. By leveraging advanced image processing and feature extraction techniques, our approach effectively identifies and classifies lung nodules from medical imaging data. This is particularly advantageous in the healthcare domain, where timely and accurate detection is crucial for patient outcomes. Our model does not rely on complex predefined rules but instead analyses the intrinsic characteristics of the imaging data to distinguish between benign and malignant nodules. This capability enhances diagnostic accuracy and supports healthcare professionals in making informed decisions.

The scalability of our model allows it to handle large volumes of imaging data, making it suitable for integration into clinical settings where rapid assessments are essential. We successfully implemented and tested this framework, demonstrating significant improvements in detection speed and accuracy. The outcomes of this project underscore the potential of machine learning-based solutions in transforming lung cancer detection and diagnostics, paving the way for more effective and efficient healthcare practices. By harnessing the power of AI in medical imaging, we can contribute to more robust healthcare systems and ultimately improve patient care in the fight against lung cancer.

8. REFERENCES

- M. Debois, “TxNxM1: the anatomy and clinics of metastatic cancer.Boston,” Kluwer Academic Publishers, 2006.
- J.M. Fitzpatrick, & M. Sonka, “Medical imaging 2003: image processing. USA: SPIE Society of PhotoOptical Instrumentation Engineers,” 2003.
- M. Haskell & J.S. Berek, “Cancer treatment. Philadelphia: W.B. Saunders, 2001.
- K.T. Manivannan, “Development of gray-level co-occurrence matrix-based support vector machines for particulate matter characterization,” Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=toledo1341577486, 2012.
- Sathish Kumar R, R. Logeswari, N. Anitha Devi, S. DivyaBharathy Efficient Clustering using the ECATCH Algorithm to Extend Network Lifetime in Wireless Sensor Networks. International Journal of Engineering Trends and Technology. 45. 476-481. 10.14445/22315381/IJETT-V45P290 – March 2017
- C. Charalambous, Conjugate gradient algorithm for efficient training of artificial neural networks, IEEE Proceedings 139 (3) (1992) 301–310.
- H. Boyle, “Support vector machines: data analysis, machine learning, and applications."
  Retrieved from http://public.eblib.com/choice/publicfullrecord.aspx?p=3021500, 2011.
- A.A. Abdulla, S.M. Shaharum, Lung cancer cell classification method using artificial neural network, Information Engineering Letters 2 (March) (2012) 50–58.
- Sathish Kumar R, Nive tha M, Madhu Mita G, & Santhoshy P. Image Enhancement using the NHSI Model Employed in Color Retinal Images. International Journal of Engineering Trends and Technology. 58. 14-19. 10.14445/22315381/IJETT-V58P203- April 2018.
- D. Harini & D. Bhaskari, “Image retrieval system based on feature extraction and relevance feedback,” ACM, 2 Penn Plaza, Suite 701, New York, NY 10121-0701, 2012.