International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7694
***Article Received:** 07 June 2023 **Revised:** 30 July 2023 **Accepted:** 16 August 2023*

Diagnosis and Stage Determination of CT Scanned Images of Lung Cancer using Hybrid model

Sandhya Waghere[1], Akshata Saptasagar[2], Rajesh Phursule[3], Rahul Badgujar[4], Atharva Misal[5], Omkar Raskar[6],
[1] Assistant Professor, Department of Information Technology, Pimpri Chinchwad College of Engineering, Pune, India
sandhya.shindhe@pccoepune.org
[2] Student, Department of Information Technology, Pimpri Chinchwad College of Engineering, Pune, India
akshata.saptasagar@gmail.com
[3] Associate Professor, Department of Information Technology, Pimpri Chinchwad College of Engineering, Pune, India
rajesh.phursule@pccoepune.org
[4] Student, Department of Information Technology, Pimpri Chinchwad College of Engineering, Pune, India
rahulbadgujar1508@gmail.com
[5] Student, Department of Information Technology, Pimpri Chinchwad College of Engineering, Pune, India
asmisal26@gmail.com
[6] Student, Department of Information Technology, Pimpri Chinchwad College of Engineering, Pune, India
omkarmraskar@gmail.com

Abstract—Varied Machine learning models are developed and are being developed for detecting and diagnosing diseases present globally. One of the unique machine learning models is proposed by the paper where the authors concentrate on combining and providing a single-stop solution as the model will be helping the medical facilities in diagnosing lung Cancer as well as determining the stage severity of the same. An amalgamation of pertinent algorithms will boost the medical processes and will help generate highly accurate results with greater precision.

Keywords- Diagnosis, Stage Determination, Lung Cancer, Support Vector Machines, Support Vector Regression, advanced analysis.

I. INTRODUCTION

It is now known that harmful drugs like alcohol, tobacco, and others increase the risk of acquiring cancer and other diseases. Unfavourable food habits, insufficient exercise, and the presence of polluted air have all been identified as significant contributors to a variety of health problems. Cancer is characterized by uncontrolled cell growth, and lung cancer specifically refers to this abnormal proliferation in the lungs. It can also occur when cells from other organs migrate to the lungs. Recent studies indicate that lung cancer is the second most common malignancy worldwide, with approximately 2.2 million cases reported in 2020. Risk factors for lung cancer include a history of lung disease, occupational exposure, and indoor air pollution. Early detection of lung cancer is challenging as it can be mistaken for flu-like symptoms, but employing multiple machine learning models can aid in its early identification. In 2022, around 236,740 cases of lung cancer were diagnosed, resulting in nearly 130,000 fatalities, according to the American Cancer Society.

[5][7]There are benign and malignant stages of lung cancer. It has been found that benign lung tumors are unneeded, aberrant tissue growths that are not malignant. These tumors may develop from different pulmonary structures. In cases where a scan suggests a potentially malignant nodule or its growth, a biopsy or complete removal of the nodule may be advised. Malignant lung cancer, on the other hand, arises from lung cells and has the potential to spread to other organs. Unfortunately, the outlook for malignant lung cancer is poor as it is considered the advanced stage of the disease. The five-year survival rate is low, but early detection and appropriate treatment can significantly improve the chances of recovery and cure. Various machine-learning models can be utilized for disease diagnosis.

[9][10]One way entails dividing the dataset into various classes using the Support Vector Machine algorithm. Support vector classifiers help with data classification and produce high-accuracy results, especially for small datasets. A regression model using Support Vector Regression (SVR) is used to further predict the severity of lung cancer. The severity is predicted on a scale of 1-10.

A. Principle of Diagnosis and Stage Determination of Disease

The model is built by employing two principles namely Diagnosis and Determination. The diagnosis principle concentrates on decisions framed by a skilled and knowledgeable person and delivering to society. This knowledge revolves around the historic life of the person along with their interests, attitude towards life and their experiences.

The determination principle constitutes of anatomical stretches of a particular disease. The disease augmentation along with the scanning of the cells and its biopsy-based tests assist the medical facilities in determining the cell's stage. Tumors, nodes and metastasis analysis quoted as TNM assists in showcasing the severity of the cells on a particular numbered scale.

II. LITERATURE REVIEW

[1]A patient's chance of survival can be dramatically impacted by lung cancer if it is discovered early. Clinical decision-making is difficult for doctors when interpreting CT scan pictures to determine the stages of lung cancer. Thankfully, computer-aided techniques using the Statistical parametric method and Grey level co-occurrence matrix have been developed to assist in this process. These models leverage advancements in machine learning and image processing in medicine. The method involves acquiring and preprocessing the images, extracting features using statistical parametric technique and GLCM, and classifying the stages. Two methods, the statistical parametric technique and the Gray-level co-occurrence matrix, are used to extract features. Among the four classifiers tested, the SVM achieved the highest accuracy of 78.95%, with a precision of 0.77 and recall of 0.83, for diagnosing lung cancer stages.

[2] From babies to the elderly, lung cancer affects people of all ages, and identifying and treating lung cancer patients results in large annual costs. Current healthcare practices rely on pricey, high-tech tools like X-rays and other imaging procedures Therefore, ensuring accurate predictions and utilizing reliable and cost-effective approaches becomes crucial. This underscores the importance of employing machine learning models with medical datasets that are relatively more efficient and affordable for medical diagnostics. While long-term cigarette use accounts for 85% of lung cancer cases, approximately 10 to 15% of cases involve non-smokers. Various techniques and resources are available today for data computation and analysis, which can be utilized in research to develop prediction models for identifying lung cancer in patients. The study compares and evaluates the performance accuracy of multiple ensemble and classification models, including SVM, KNN, Random Forest, Voting classifier, and a combination of artificial neural networks. By leveraging advanced technology, early detection of lung cancer can be easily achieved in patients.

[3] Research on the registration requirements for cancer patients is included in the paper. The TNM system's most recent ninth version makes it possible to identify the stages of skin and breast cancer. The study presents its findings in a publication, which utilized regression models developed through machine learning techniques to identify the stages of breast, skin, bronchial, and lung cancer. The experiments involved three machine learning methods: ensemble algorithms, SVM and SVR. Model comparisons were conducted, and evaluation metrics such as coefficient of determination, MSE, MAE, and RMSE were used.

[4] Rising lung cancer cases have resulted in greater death rates for both men and women. There is presently no treatment for lung cancer, which includes unchecked cell growth in the lungs. However, there are ways to lower the risk of lung cancer, highlighting the value of early identification for better patient outcomes. The number of smokers is inversely correlated with lung cancer incidence. Several classification techniques were used to estimate the chance of acquiring lung cancer, including Naive Bayes, SVM, and Logistic Regression. The primary objective of this study is to diagnose lung cancer at an early stage by evaluating the effectiveness of classification algorithms.

[6]The incidence of lung cancer has been increasing, leading to higher mortality rates among both men and women. Lung cancer involves uncontrolled cell proliferation in the lungs and currently lacks a cure. However, the risk of lung cancer can be reduced, emphasizing the importance of early detection for improved patient outcomes. The prevalence of lung cancer is inversely related to the number of individuals who smoke. To assess the likelihood of developing lung cancer, various classification methods, including Naive Bayes, SVM, Decision Trees, and Logistic Regression, were employed. The primary objective of this study is to diagnose lung cancer at an early stage by evaluating the effectiveness of classification algorithms.

[8]Detecting cancer at an early stage is crucial for improving a patient's chances of survival. This research investigates the effectiveness of logistic regression (LR) and support vector machine (SVM) algorithms in predicting the survival rate of lung cancer patients. The study compares the performance of these two algorithms using metrics such as accuracy, precision, recall, F1 score, and confusion matrix. By employing these methods, healthcare professionals can assess the probability of survival for lung cancer patients and make predictions about the disease's progression.

[11]Early identification of lung nodules is crucial for the effective diagnosis and treatment of lung cancer. Various methods have been explored, such as thresholding, pattern recognition, computer-aided diagnosis, and backpropagation computation. To address the limitations of existing classifiers, a deep hybrid learning-based technique is proposed, aiming to assess the performance of lightweight, low-memory deep neural network architectures. The suggested approach achieves an accuracy of 85.21 percent by striking a suitable balance between specificity and sensitivity. The study employs binary classification networks like vanilla 2D CNN, 2D MobileNet, and 2D SqueezeNet to distinguish between early-stage and later-stage lung cancer in patient CT scans.

[12]A computer-based detection method was developed to recognize lung nodules and identify cancer cells within them. The project's primary goal is to create data and CPU infrastructure capable of automatic image processing and malignancy computation. The proposed approach includes a

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7694
***Article Received:** 07 June 2023 **Revised:** 30 July 2023 **Accepted:** 16 August 2023*

neural classifier and improvement filter to enhance imported images, assist in nodule selection, and reduce false positives. The model is tested on both inner and outer nodules, and the results are presented as a response characteristic curve.

III. PROPOSED WORK

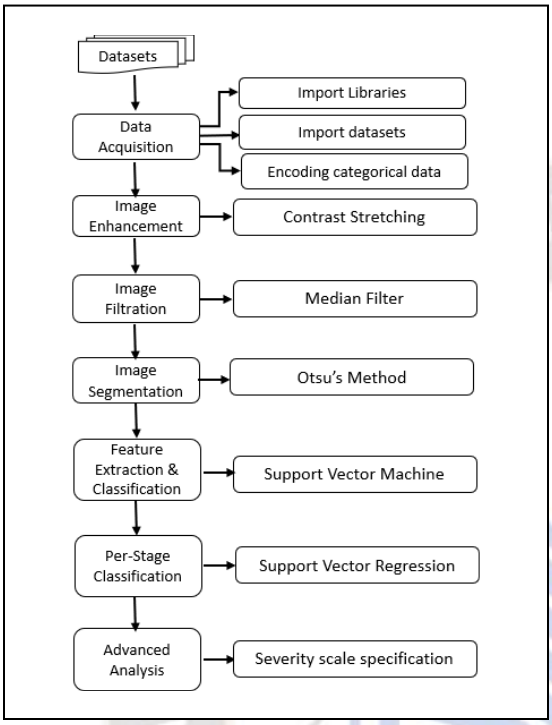

Figure 1. Workflow of the proposed model.

The model proposed in this research paper constitutes the amalgamation of ML algorithms.

The flow described in the diagram depicts the procedure of the model builds using various ML algorithms. Also, all the significant steps are portrayed along with the respective algorithms used to employ those steps from importing all the necessary entities to procuring the result with accuracy and precision.

A. Data Pre-processing

Pre-processing the data is a prime step for all the ML models as it constitutes the quality of the model being built. The paper's proposed model is also built keeping in mind, the model's quality and significance, henceforth a set of the below-listed techniques are implemented to build a qualitative model.

1) Importing the Libraries required for Implementation:

For the development of the complete project, PyCharm IDE is used as it is user-friendly and intuitive to use. For including SVM and SVR we are using the sklearn python library. Sklearn is a popular Python library that provides a range of tools for machine learning, including algorithms for classification, regression, clustering, and dimensionality reduction, as well as tools for data preprocessing, model selection, and evaluation.

CV2 is a Python library that is utilized for performing image processing operations and is an integral component of the OpenCV project. The cv2 module provides a broad range of functions that can be used for tasks such as image and video processing, machine learning, feature detection, and object recognition. It can handle various image and video file formats and enables users to capture video from webcams or other video capture devices. We are leveraging CV2 for performing image manipulation and preprocessing functions.

2) Importing the Datasets required for Implementation:

Importing the necessary modules and libraries into the workspace is a critical requirement for building and enhancing the performance of a model. These modules and libraries enable the creation of various plots and facilitate straightforward numerical computations. Additionally, this process aids in generating data frames and provides a streamlined interface between clients and frameworks, all while taking necessary precautions to mitigate risks.

3) Importing the Libraries required for Implementation:

The process of label encoding involves converting categorical data into an integer format specific to the model so that it can be easily utilized by the model.

In this particular case, three label encodings are used:

- 0: Benign - Refers to non-cancerous growths that do not invade other areas of the body. For instance, a benign lung tumor may cause symptoms such as chest pain, shortness of breath, and coughing if it exerts pressure on surrounding tissue or airways.
- 1: Refers to lung cancer that begins in the lungs and has the potential to spread to other organs of the body. It is a severe and potentially life-threatening illness that is frequently caused by smoking or exposure to certain chemicals or pollutants.
- 2: Normal - Refers to healthy lung tissue that is not affected by cancer or other abnormalities. Normal lung tissue is essential for proper breathing and oxygen exchange within the body.

B. Image Enhancement

We need to increase the clarity and contrast of the image. For this, we are using a technique called Image Enhancement. This process provides the machine with better input for processing tasks. The process of image enhancement involves reducing image noise, sharpening, and brightening the image. For this process, the proposed model uses the Contrast Stretching technique.

1) Contrast Stretching:

This technique enhances contrast by expanding the range of intensity values to cover the entire dynamic range of the image. It is a monotonically increasing linear transformation function that continuously enhances the contrast of the image. This is

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7694
***Article Received:** 07 June 2023 **Revised:** 30 July 2023 **Accepted:** 16 August 2023*

done by digitally mapping the pixels present in the image over a wider range. This improves the visibility of details and features.

```python
image-enhancer.py

def image_enhancement(self, ip_img_array):
    # Get the minimum and maximum pixel values in the image
    min_val, max_val, _, _ = cv2.minMaxLoc(ip_img_array)
    # Compute the output image using contrast stretching
    out_img = np.uint8((ip_img_array - min_val) * (255 / (max_val - min_val)))
    return out_img
```

Figure 2. Image Enhancement Code Snippet.

C. *Image Filtration*

After the enhancement of the image, the further step is to filter the image. For further processing, we need a sharp image. For this image filtration technique is used. The proposed model uses a median filter for the image filtration process.

1) *Median Filter*:

The Median Filter is a nonlinear digital signal processing method used to eliminate noise in an image or signal. It works by replacing the value of each pixel with the median value of the pixels surrounding it. In image denoising, the median filter is often preferred over mean and Gaussian filters because it is effective in removing impulse noise while preserving edges and fine details in the image. This is because mean filters can blur the edges while filtering the image, while Gaussian filters preserve the edges, resulting in a blurry image that appears soft.

```python
image-filtrator.py

def image_filtration(self, ip_img_array):
    filtered_img = cv2.medianBlur(ip_img_array, ksize=3)
    return filtered_img
```

Figure 3. Image Filtration Code Snippet.

D. *Image Segmentation*

After Filtration, we proceed with a technique called image segmentation. In Image segmentation, the image is divided into multiple segments or regions, each representing a distinct object or location within the image. The proposed model uses Otsu's method (global thresholding method) for image segmentation.

1) *Otsu's method (global thresholding method)*:

Otsu's method, also known as global thresholding, is a technique used to determine the optimal threshold value for segmenting an image into foreground and background regions. The method is based on the principle that an image can be divided into foreground and background pixels, and the ideal threshold value is the one that minimizes the variance within each class (foreground or background) while maximizing the variance between the classes. The intra-class variance represents the average deviation of pixel intensities within a class, while the inter-class variance represents the separation between the two classes. Otsu's method is commonly used in computer vision and image processing to automate image segmentation.

```python
image-segmentor.py

def image_segmentation(self, ip_img_array):
    _, segmented_img = cv2.threshold(ip_img_array, 0, 255, cv2.THRESH_BINARY + cv2.THRESH_OTSU)
    return segmented_img
```

Figure 4. Image Segmentatioin Code Snippet.

E. *Feature Extarction:*

Define abbreviations and acronyms the first time they are used in the text, even with a technique called image segmentation. In Image With the data preprocessing done, the processed images are passed into the model for the feature extraction process. It is the process of obtaining relevant information or patterns from unprocessed data such as images, audio signals or text. In image processing, feature extraction involves identifying and extracting relevant patterns or characteristics from an image that can be used to represent the image in a compact and meaningful way.

Two commonly used feature extraction techniques in image processing are Hu Moments and Histogram of Oriented Gradients (HOG):

- Hu Moments: Hu Moments is a set of mathematical properties that define the contour or shape of an object in an image, also known as Hu invariant moments. These properties are used to represent an object's shape regardless of its size, orientation, or position in the image.
- Histogram of Oriented Gradients: HOG is a feature descriptor that is used to extract useful information from an image by representing the distribution of gradient orientations in the image. It is often used for object detection and recognition tasks in computer vision. HOG descriptors provide information about the local structure or texture of an image and can be used to identify patterns such as edges and corners.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7694
***Article Received:** 07 June 2023 **Revised:** 30 July 2023 **Accepted:** 16 August 2023*

```python
# feature-extractor.py
def fd_hu_moments(image):
    return cv2.HuMoments(cv2.moments(image)).flatten()
def quantify_image(image):
    # For Speed and pressure of signature image
    # compute the histogram of oriented gradients feature vector for
    # the input image
    features = feature.hog(image, orientations=9,
                           pixels_per_cell=(10, 10), cells_per_block=(2, 2),
                           transform_sqrt=True, block_norm="L1")
    # return the feature vector
    return features
```

Figure 5. Feature Extraction using SVM - Code Snippet.

F. Classification:

Once we have obtained the required features from the previous process, the next step is the classification of the image. It is a supervised machine-learning task that involves predicting the category or class of an input instance by learning from labelled training examples. This task can be applied to a wide range of data types, including images, audio signals, and text documents, as long as they can be represented numerically.

The proposed model uses a unified approach where both feature extraction and classification are combined using a single algorithm. The commonly used algorithm for this purpose is Support-Vector Machine (SVM). within the image. The proposed model uses Otsu's method (global thresholding method) for image segmentation.

1) Support Vector Machines:

In classification, the SVM algorithm aims to find the optimal hyperplane that separates the input instances into different classes based on their features. The SVM algorithm works by mapping the input instances into a high-dimensional space and then finding the hyperplane that maximally separates the classes.

In the proposed model, the extracted features from the image are fed into an SVM model for classification. The SVM model uses the extracted features to predict the class of the input instance. The SVM algorithm has been widely used in image classification tasks because of its ability to handle high-dimensional feature spaces and its robustness to noise and outliers.

```python
# feature-extraction.py
def get_input_feature(self, image_path, training_config: SvmModelTrainingConfig):
    image = cv2.imread(image_path)
    image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
    image = cv2.resize(image, (200, 200))
    if training_config.pretraining_preprocessing_enabled:
        image = PreprocessingUtils.apply_all_preprocessors(image)
    ft1 = quantify_image(image)
    ft2 = fd_hu_moments(image)
    return np.hstack([ft1, ft2])
```

Figure 6. Classification using SVM - Code Snippet.

Figure 7. Training Dataset Classification using SVM - Code Snippet.

Figure 8. Testing Dataset Classification using SVM - Code Snippet.

G. Regression

Regression refers to a statistical method used to analyze and model the relationship between a dependent variable (also called the outcome or response variable) and one or more independent variables (also called predictors, explanatory variables or features). The goal of regression analysis is to estimate the relationship between the variables and make predictions about the value of the dependent variable based on the values of the independent variables. Regression models can be used for both descriptive and predictive purposes.

1) Support Vector Regression:

Support Vector Regression (SVR) is a machine learning algorithm that is used for regression analysis. SVR is a type of regression analysis that uses the Support Vector Machine (SVM) algorithm to perform regression tasks. It involves finding a hyperplane in a high-dimensional space that maximizes the margin between the predicted values and the actual values of the dependent variable. The SVR model creates a hyperplane that best fits the data points, where the goal is to minimize the error between the predicted and actual values of the dependent variable. This is achieved by identifying a narrow band or margin around the hyperplane that contains a maximum number

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7694
***Article Received:** 07 June 2023 **Revised:** 30 July 2023 **Accepted:** 16 August 2023*

of data points, while at the same time minimizing the distance between the data points and the hyperplane.

```
svr.py

from sklearn.svm import SVR


# Train the SVR model
model = SVR()
model.fit(x_train, y_train)
```

Figure 9. Support Vector Regression Code Snippet.

H. Providing the severity of the lung cancer:

Using SVR and SVM we are implementing a severity meter that measures the severity of lung cancer on a scale of 1 to 10.

IV. RESULTS: FIGURES AND TABLES

All the necessary pre-processing steps are implemented which include importing all the necessary entities like libraries and datasets. All the successive steps following the importing steps are noted below.

A. Image Enhancement: Contrast Stretching:

Post-encoding the data into categorical format, image enhancement is implemented via contrast stretching which leads to improving the clarity of visualizing models.

B. Image Filtration: Median Filter

Image filtration is performed after enhancing the image where all the images are smoothened and sharpened using Median Filter.

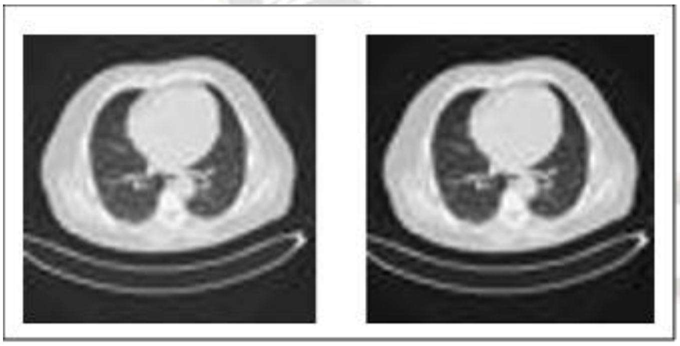

Figure 10. Original Image (left) and Filtered Image(Right).

C. Image Segmentation: Otsu's Method

Via Otsu's method, the filtered images are segmented and converted into pixelated images.

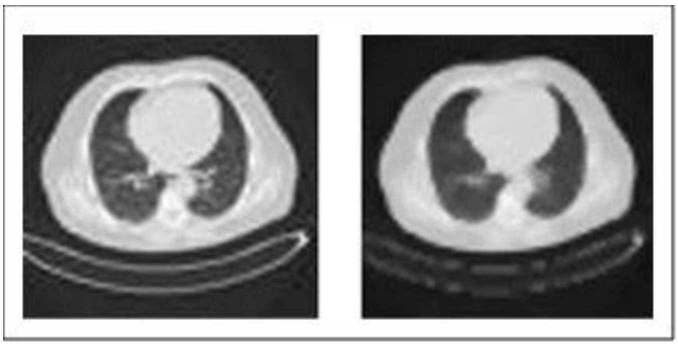

Figure 11. Original Image (left) and Segmented Image(Right).

D. Original Dataset Classification

The unedited raw dataset is divided into 3 categories namely Normal, Benign and Malignant. 2277 images of the entire dataset constitute to be training dataset and 750 encompass as testing dataset.

TABLE I. CANCER DATASET

Cancer Type	Training Dataset	Testing Dataset
Normal	936	294
Benign	435	132
Malignant	906	324

E. Distribution of dataset for: Normal Case

The dataset has been distributed across the model and the statics depict the region in which the normal cases are most likely to be found.

TABLE II. DISTRIBUTION OF NORMAL CASES ON A SCALE OF 1-10

Scale	Percentage Samples
1	11.22
2	12.24
3	27.55
4	4.76
5	11.22
6	16.33
7	13.61
8	3.06
9	0
10	0

F. Distribution of dataset for: Benign Case

Benign cases are most to be found from 5-8 on a scale of 1-10 when the dataset is dispersed across the model for training as well as testing purposes found.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7694
***Article Received:** 07 June 2023 **Revised:** 30 July 2023 **Accepted:** 16 August 2023*

TABLE III. DISTRIBUTION OF BENIGN CASES ON A SCALE OF 1-10

Scale	Percentage Samples
1	0
2	6.06
3	18.18
4	30.3
5	39.39
6	6.06
7	0
8	0
9	0
10	0

G. Distribution of dataset for: Malignant Case

The malignant cases are densely populated from 8-10 on the range of 1-10 scale. The distribution depicts the region and provides a hint of severity as well.

TABLE IV. DISTRIBUTION OF MALIGNANT CASES ON A SCALE OF 1-10

Scale	Percentage Samples
1	0
2	0
3	3.7
4	8.33
5	0.93
6	6.48
7	4.63
8	53.7
9	12.96
10	9.26

H. Distribution Heatmap

The heatmap constitutes a scale of 1-10 where it showcases the distribution of the dataset. Via having a sight at the heatmap, the conclusion is derived where the region from 1-5 belongs to the normal case, 5-8 for benign cases and 8-10 for malignant cases. A gauge-similar metric scale can be used as a visualization tool to visualize the severity of lung cancer.

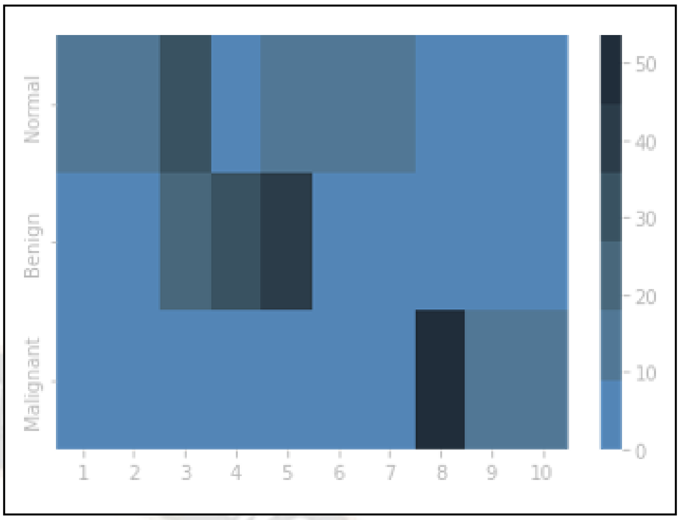

Figure 12. Distribution Heatmap.

V. MODEL SUMMARY

The model is implemented using a basic flow wherein each step is designated to be followed using specific algorithms. All these algorithms constitute for increasing the accuracies and precisions of the model. Also, the model is a combi- nation of 2 parts namely diagnosis and stage determination. Diagnosis is done using a Support Vector machine coupled with pre-processing techniques which provides an accuracy of 77.2%. The stage determination is implemented using the support vector regression and thus a heatmap is generated for illustrating the distribution of the dataset. Also, this heatmap assisted in providing the severity of the lung cancer which is plotted on a scale of 1-10. Thus, the model can be used as a whole for generating a complete statement about lung cancer for human society.

TABLE V. COMPARING BOTH THE MODEL SUMMARIES

Model Specifications	Non-Pre-processed Model	Pre-Processed Model
Accuracy	71.86666666666667	77.2
Training Time (in nanoseconds)	8852864129	9931449479

VI. CONCLUSION

The proposed model focuses on classifying the collected dataset, employing Support Vector Machine (SVM) as the main technique along with data pre-processing methods. The model achieves an overall accuracy rate of 77.2%. It is a unique and comprehensive solution that not only classifies the tumor but also provides advanced analysis and evaluates the curability status of lung cancer. Additionally, the paper emphasizes the significance of pre-processing techniques when constructing a machine-learning model. By incorporating SVR into the model, there is an enhancement in the accuracy and reliability of the predictions and it also assists in making informed decisions regarding patient outcomes.

VII. FUTURE WORK

The described model in the paper revolves completely around diagnosing lung cancer as well as determining the severity of the stages procured as a result via the Hybrid ML model. The model includes a 100% implementation of all the algorithms and provides a complete solution with notable accuracy and precision rates. The model's accuracy can be increased and along with this, the model can also be paired with new algorithms leading to a gradual increase of accuracy as well as precision values eventually dominating the medical fields in technical aspects. Additionally, the models can be trained using a variety of datasets making them flexible for implementation in larger domains thus increasing the model's use-case:

REFERENCES

[1] M. Islam, A. H. Mahamud and R. Rab, "Analysis of CT Scan Images to Predict Lung Cancer Stages Using Image Processing Techniques," 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2019, pp. 0961-0967, doi: 10.1109/IEMCON.2019.8936175.

[2] A. Hoque, A. K. M. A. Farabi, F. Ahmed and M. Z. Islam, "Automated Detection of Lung Cancer Using CT Scan Images," 2020 IEEE Region 10 Symposium (TENSYMP), 2020, pp. 1030-1033, doi: 10.1109/TENSYMP50017.2020.9230861.

[3] C. Thallam, A. Peruboyina, S. S. T. Raju and N. Sampath, "Early Stage Lung Cancer Prediction Using Various Machine Learning Techniques," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2020, pp. 1285-1292, doi: 10.1109/ICECA49313.2020.9297576.

[4] Bhavani M., & M. Durgadevi. (2023). Streamlined Classification of Microscopic Blood Cell Images. International Journal of Intelligent Systems and Applications in Engineering, 11(1s), 57–62. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2477

[5] M. Todorova, "Application of Machine Learning Methods for Determining the Stage of Cancer," 2020 International Conference Automatics and Informatics (ICAI), 2020, pp. 1-4, doi: 10.1109/ICAI50593.2020.9311355.

[6] Juan Cui, Fan Li, Guoqing Wang, Xuedong Fang, J David Puett, and Ying Xu. Gene-expression signatures can distinguish gastric cancer grades and stages. PloS one, 6(3):e17819, 2011.

[7] P. Basak and A. Nath, "Detection of different stages of lungs cancer in ct-scan image using image processing techniques", International Journal of Innovative Research in Computer and Communication Engineering, vol. 5, pp. 9708-9719, 2017.

[8] Ms. Sweta Minj. (2012). Design and Analysis of Class-E Power Amplifier for Wired & Wireless Systems. International Journal of New Practices in Management and Engineering, 1(04), 07 - 13. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/9

[9] Va Dominic, Dr. Deepa Gupta, Sangita Khare, and Aggarwal, Ab, "Investigation of chronic disease correlation using data mining tech- niques", in 2015 2nd International Conference on Recent Advances in Engineering and Computational Sciences, RAECS 2015, 2015.

[10] James D. Brierley BSc, MB, FRCP, FRCR, FRCPC, Mary K. Gospo- darowicz MD, FRCPC, FRCR (Hon) Christian Wittekind MD, "TNM Classification of Malignant Tumours Eighth Edition", Union for In- ternational Cancer Control (UICC), Pages:1-241, 2017.

[11] Castro, J., González, J., Rodríguez, C., Garcia, J., & Jóhann, Þorvaldsson. Predicting Student Dropout Rates in Engineering Programs with Machine Learning. Kuwait Journal of Machine Learning, 1(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/105

[12] Suren Makaju, P.W.C. Prasad, Abeer Alsadoon, A.K. Singh, A. Elchouemi, Lung Cancer Detection using CT Scan Images, Procedia Computer Science, Volume 125, 2018, Pages 107-114.

[13] Lingyun Gao, Mingquan Ye, Xiaojie Lu, Daobin Huang, Hybrid Method Based on Information Gain and Support Vector Machine for Gene Selection in Cancer Classification, Genomics, Proteomics Bioinformatics, Volume 15, Issue 6, 2017, Pages 389-395.

[14] Rashmi Mothkur, B N Veerappa, Classification Of Lung Cancer Us- ing Lightweight Deep Neural Networks,Procedia Computer Science, Volume 218, 2023, Pages 1869-1877.

[15] Manoharan, D. S. Sathesh, A. (2020). Early diagnosis of Lung Cancer with Probability of Malignancy Calculation and Automatic Segmentation of Lung CT scan Images. Journal of Innovative Image Processing, 2(4), 175-186. doi:10.36548/jiip.2020.4.002.