Image Processing-based Performance Evaluation of KNN and SVM Classifiers for Lung Cancer Diagnosis

Kavitha B C[1], Naveen K B[2]*
Research Scholar, Department of Electronics and Communication Engineering, BGS Institute of Technology, Adichunchanagiri University, B G Nagara, India, 571448[1]
Professor, Department of Electronics and Communication Engineering, BGS Institute of Technology, Adichunchanagiri University, B G Nagara, India, 571448[2]

***Abstract*—It is important to note that the cure rates in cases of advanced stages of lung cancer are remarkably low, which stresses out the importance for early detection as means to increase survival chances. A strong area of focus when it comes to increased research in the lung cancer diagnosis is the search for ways through which this disease can be identified at its early stages. The methodology described below is proposed as a means to facilitate early detection of lung cancer. There are two phases in this approach. The study deals with effectiveness of three types of classifiers K-Nearest Neighbors (KNN), Random Forest and Support Vector Machine (SVM) to identify cases related to lung cancer via relevant medical data assessment. In this application, the eval axis performs profiling or measures the accuracy of applying these classifiers and discriminating between cancerous instances versus non-cancerous ones within the dataset. To rate the adequacy of classifiers in distinguishing classes, performance metrics like accuracy, precision, recall and F1- score are used. Furthermore, the research compares KNN, Random Forest and SVM, explaining their specific advantages as well as disadvantages logically referring to how they can or cannot be applied while detecting lung cancer. This investigation shows helpful results in suggesting the possibility that machine learning techniques could assist to identify lung cancer as exact and timely as possible, providing more successful diagnostic procedures and patient outcomes. The experimental findings show that SVM gives the best result at 95.06%, KNN comes second with a percentage of 86.89.**

Keywords*—K-Nearest Neighbors; lung cancer detection; machine learning; medical data; performance metrics; support vector machine

I. Introduction

A seamless connectivity to users is provided by wireless cancer stands as a formidable threat to human life, frequently leading to fatalities globally due to delayed diagnoses. The primary role of the lungs involves supplying the body with oxygen and expelling carbon dioxide during essential bodily functions. Lung cancer develops from uncontrolled tissue and cell growth within the lungs, which, left unchecked, can spread and harm neighboring tissues. It claims around 1.3 million lives yearly globally, with 30–40,000 new cases in Turkey every year, and is the top cause of cancer-related deaths among men and the second-leading cause among females [1].

The impact of this disease is profound, evidenced by its higher mortality rate compared to combined rates of correctional, pancreatic, and breast cancers. For example, in 2020, an estimated 30,000 Canadians were expected to be diagnosed, resulting in approximately 21,000 deaths. Globally, the burden of cancer is predicted to double by 2050, with lung cancer at the forefront [2]. Late-stage diagnoses often lead to the fatality of lung cancer.

A comprehensive understanding of its development, along with effective early detection methods and suitable treatments, significantly influences better outcomes. Lung cancer often becomes lethal because of late-stage diagnosis. Improved outcomes are largely dependent on a thorough understanding of the pathophysiology, appropriate medications, and efficient early detection techniques.

As a result, it is still crucial to detect lung cancer as soon as possible, particularly in high-risk groups like smokers or people who work in toxic settings or are exposed to oil fields [3, 4]. Novel biomarkers are desperately needed to help with this population screening. Lung cancer symptoms could not cause serious problems until the illness is fairly advanced [5, 6].

The primary cause of lung cancer's extreme hazard is its ability to progress without showing any signs. About 25% of cancer patients experience no symptoms at all. Most people find out that lung X-rays from another illness cause lung cancer. When it comes to lung cancer, early diagnosis is crucial [7, 8]. Due to the fact that lung cancer frequently spreads quickly to the brain, liver, adrenal glands, and bones.

However, the average life expectancy and quality have grown with the recently developed lung cancer treatment approaches. Thanks to developments in imaging methods like low-dose spiral computed tomography, lung cancer can now be identified early on.

II. Literature Survey

One of the deadliest illnesses a person can have is cancer. The late diagnosis can result in deaths worldwide. The lungs are responsible for fundamental biological activities such as breathing in oxygen and exhaling carbon dioxide. Lung cancer develops when lung tissues and cells multiply uncontrollably.

As they grow unchecked in their natural habitat, these tumours pose a threat to adjacent tissues via metastasis. While lung cancer ranks second for women, it tops the list for males when it comes to cancer-related deaths [9,10]. An estimated 1.3 million individuals worldwide pass away from lung cancer each year. Every year in Turkey, between 30,000 and 40,000 new cases of lung cancer are reported. Lung cancer symptoms could not cause serious problems until the illness is fairly advanced. The primary feature that makes lung cancer so hazardous is its ability to progress without symptoms [11, 12]. About 25% of cancer patients experience no symptoms at all. Most people find out that lung X-rays from another illness cause lung cancer. Lung cancer identification at an early stage is crucial. Due to the fact that lung cancer frequently spreads quickly to the brain, liver, adrenal glands, and bones. On the other hand, the average life expectancy and quality have grown with the recently developed lung cancer treatment approaches [13,14]. Thanks to developments in imaging methods like low-dose spiral computed tomography, lung cancer can now be identified early on. Moreover, accurate diagnosis is essential for developing the best possible treatment plans for each lung cancer patient [15, 16]. Therefore, in order to improve prognosis and treatment results, the identification of sensitive and specific biomarkers for early detection has arisen as a crucial necessity in the field [17, 18]. Building machine learning models to improve lung cancer detection and prediction accuracy is the main goal of this work. This study aims to find the best model for early-stage lung cancer detection by using several classifiers and comparing their performance measures [19, 20]. Regarding lung cancer, this effort has significant potential for enhancing patients' outcomes with early detection. The proposed strategy in this study is quite strong as well, enabling for an impressively high accuracy ratio of predicting lung cancer early on. In this work, the dataset used is gathered from reputable research centers and that of the Machine Learning Repository. To determine and compare the ratios of their accuracy, we have utilized two different classifiers out here SVM and KNN.

III. METHODOLOGY

The lung cancer detection and prediction have been developed in MATLAB using an effective algorithm based on the image processing techniques. One of the detection methods utilised by this algorithm is that of a multi-stage classification process in diagnosing lung cancer, and so forth. First, it verifies if there can be ones or more cells in the input image impacted by cancer; otherwise it proceeds with determining if lung cancer is going to occur. The cancer algorithm has an additional stage after identifying that it is a case of cancer. This involves further malignancies being classified or divided into early, intermediate and advanced staging of the disease. There is a series of classification steps before each of them to develop a variety of segmentation and picture enhancement techniques. Techniques like contrast enhancement, colour space modification, and image scaling are all part of image enhancement. For segmentation purposes, the algorithm utilizes thresholding and marker-controlled watershed-based techniques. The comprehensive workflow of this proposed system is illustrated in Fig. 1, outlining the sequential stages and processes involved in the detection and prediction of lung cancer through image processing methods.

IV. DATA COLLECTION

Gather a diverse dataset containing relevant patient information, including clinical attributes and diagnostic indicators related to lung cancer. The effectiveness of machine learning models relies heavily on datasets sourced from credible origins, particularly those encompassing a substantial volume of images. Numerous datasets are accessible, facilitating lung disease detection, and some researchers opt to create their datasets for enhanced accuracy [21, 22]. Even so, there are several other publicly accessible datasets for research and teaching on the internet. Examples of lung CT scan pictures used in this study's experimental analysis are shown graphically in Fig. 2. These images serve as representative samples from the dataset used, demonstrating the type and quality of data utilized for the research's experimental investigations [23]. The model has been trained and tested using the LIDC-IDRI lung CT scan dataset for the proposed task. In our efforts to verify and confirm the findings of our study, we obtained a specific set of samples from Adichunchanagiri Hospital and Research Centre (AHRC) as well as Adichunchanagiri Institute of Medical Sciences (AIMS), located in BG Nagara, Mandya.

V. DATA PREPROCESSING

Address outliers, inconsistencies, missing numbers, and clean up the dataset. Make sure that the feature scales are consistent by normalising or standardising the data. An essential part of the standard work flow when developing a machine learning model entails image pre-processing. Since datasets commonly include image files of different dimensions, Formats and the possibility of having noise or blurriness, pre-processing these images prior to machine learning training is necessary. CLAHE Contrast Limited Adaptive Histogram Equalisation, Wiener filtering, Adaptive Gaussian Filtering and Gaussian and Gabor filtering are some of the most commonly used pre-processing methods for analysis (see Fig. 3). Through application of these techniques to the data, one can normalize the images, correct for size and format variations and minimize noise or smudging thus benefiting optimally from the data for more productive use in machine learning models [24, 25].

VI. MODEL TRAINING

A. K-Nearest Neighbors (KNN) Classifier

The KNN algorithm classifies datasets based on the similarity of one sample with another in terms of a set of neighbors. The number of neighboring datasets considered is K. It uses the process of Euclidean Distance (ED) measurement to classify by comparing similarity between test sample and other samples in the database. The provided sentence describes the given distance in terms of the Euclidean plane between two sets of coordinates, X and Y.

$$ED\ (x, y) = \sqrt{\sum_{j=1}^{k} (X_i - Y_i)^2} \quad (1)$$

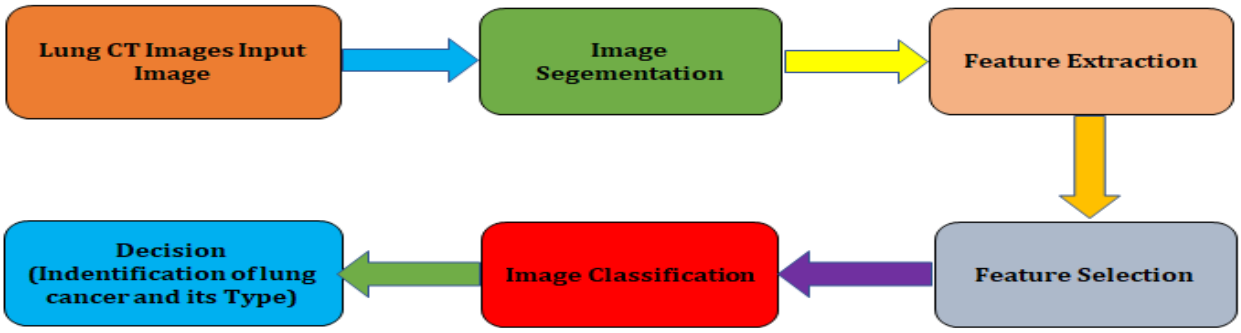

Fig. 1. Lung cancer prediction system.

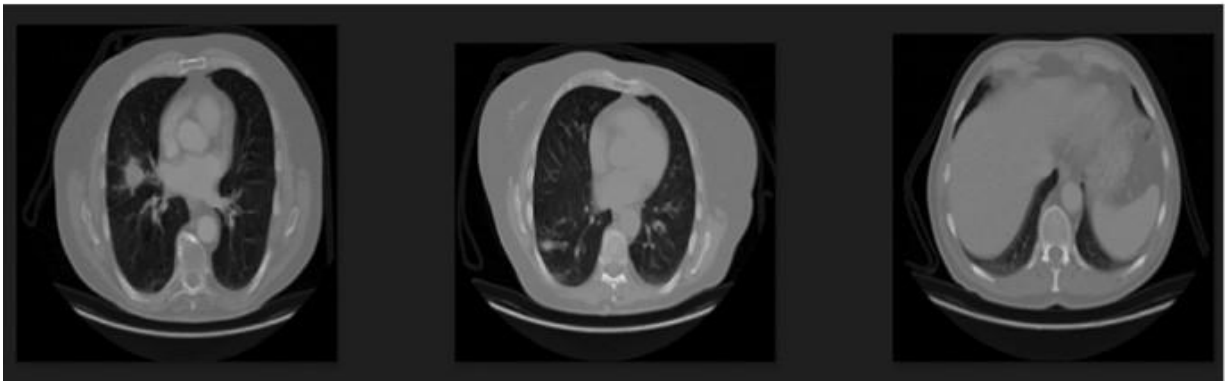

Fig. 2. Sample CT image.

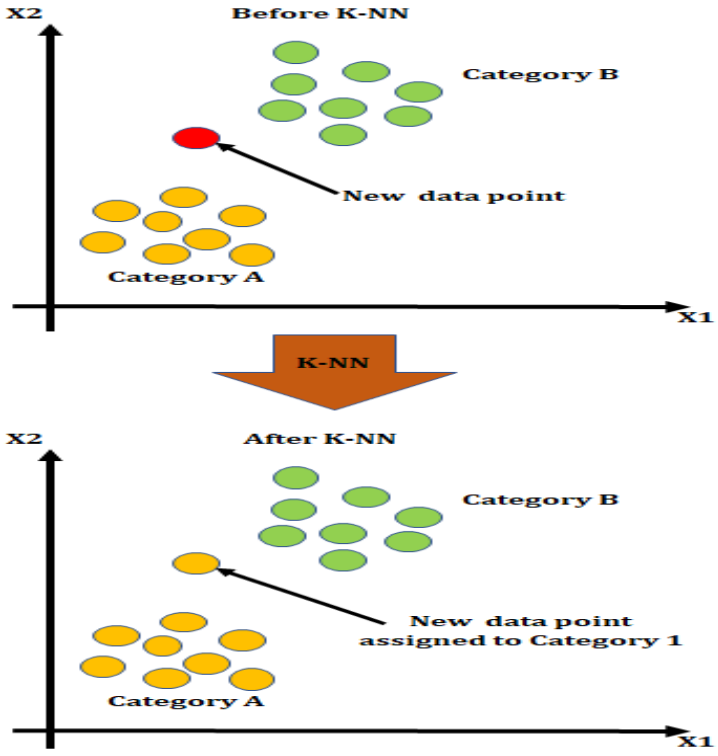

Fig. 3. KNN classification.

The diagnosis of small cell lung cancer (SCLC) images is based on the Entropy Degradation Method (EDM), as proposed by Qing Wu et al. in their study [13]. This includes a set of inputs called EDM associated with every training set which upon being transformed into an EDM score, applied through the logistic function gives it a probability [26]. The testing phase, an input is used which contains no markings and injected into a neural system. The output is subsequently computed by making use of the scores probabilities. So this final output is very crucial because it helps in determining the category to which the testing data belongs i. It can tell that whether patient is suffering with cancer or not and in non-cancerous person, will be found with either normal function.

$$a_{m,n} = \log(\sum_m \sum_m^N \overrightarrow{Point}\ (p = j)x\ \overrightarrow{x_j} + \log\left(\sum \overrightarrow{Point}\ (p = n)\right) / (\sum \sum \overrightarrow{Point})) \quad (2)$$

The total size of input x is N which, for values of n being either 0 or 1, where p=n serves as an indicator function and M ranges from 1 to the size of the input x; the estimated signals of maximum entropy are Y = cd f (y). The calculation of its value involves the function h, expressed as follows:

$$h = \log(\det(det(W)) + \frac{1}{N} \sum(\log(e + 1 - Y^2)) \quad (3)$$

Where, W denotes an identity 5x5 matrix.

$$W = w + eta \times (g)$$

where, eta defines the rate of convergence and the gradient matrix is given by g.

B. SVM Classifier

SVM is one of the popular approaches used in supervised learning, which is frequently adopted to address problems touching on both regression and classification (see Fig. 4). This is a form of machine learning mostly used in classification tasks, SVM. It ensures that the margin of separation between classes is maximised by building a hyperplane. For instance, SVM applies its method to generate a hyperplane that separates different categories of NSCLC in the field classification of lung cancer varieties. This aim is to determine the lung cancer type using NSCLC categories accurately. For an n-dimensional space, SVM algorithm aims at producing that magical line or decision border in space which is best in organizing the classes. This decision boundary is also sometimes called a hyperplane and determines how future additional data points will be classified on the accuracy. Support Vector Machine is the name of some particular method because these situations we are discussing are called support vectors. The Sample cases of Normal, Benign and Malignant CT images as shown in Fig. 5. This way the algorithm's ability to correctly classify data is enhanced in the process of determining the optimum decision boundary with aid of these support vectors.

C. Random Forest

Machine learning's Random Forest method pools the power of several decision trees to provide accurate forecasts (see Fig. 6). This method comes up with a huge number of decision trees while training, each employing different subset of the features and data. Random Forest achieves a diversified prediction by outputting multiple diverse trees produced through the process of bootstrap sampling and feature randomness. During categorization, it sums up the outcomes from individual trees through a majority voting scheme; for regression tasks, it averages the results of constituent trees. Importantly, Random Forest is known as a method that can work with high-dimensional data and helps to minimize overfitting as well as reveals the importance of features. It, therefore, has a very wide-ranging application in many fields where precise predictions are important.

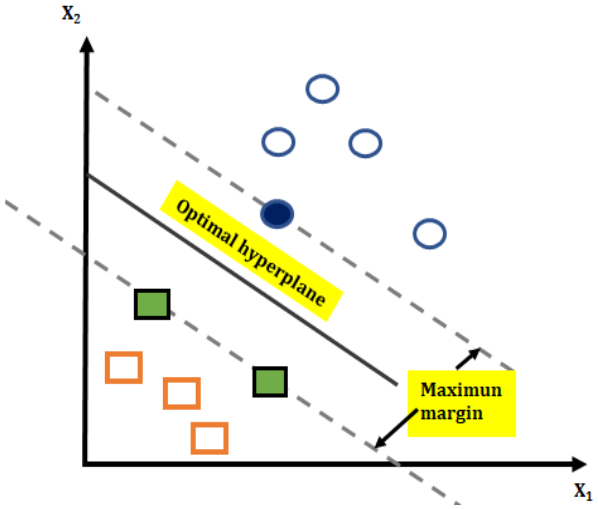

Fig. 4. SVM classification.

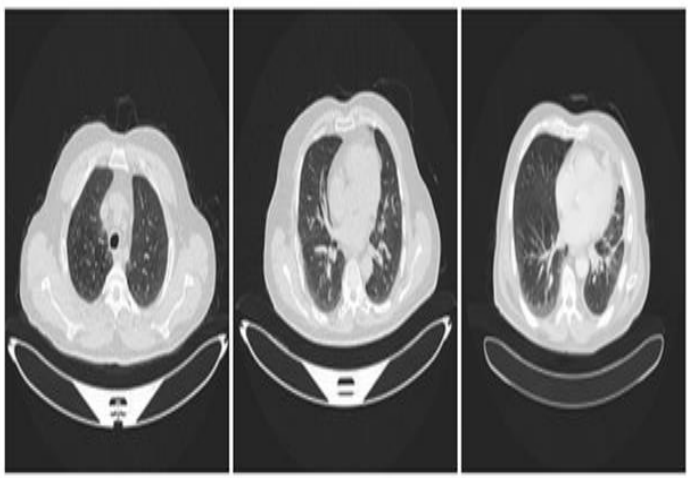

Fig. 5. Normal, Benign and Malignant CT images.

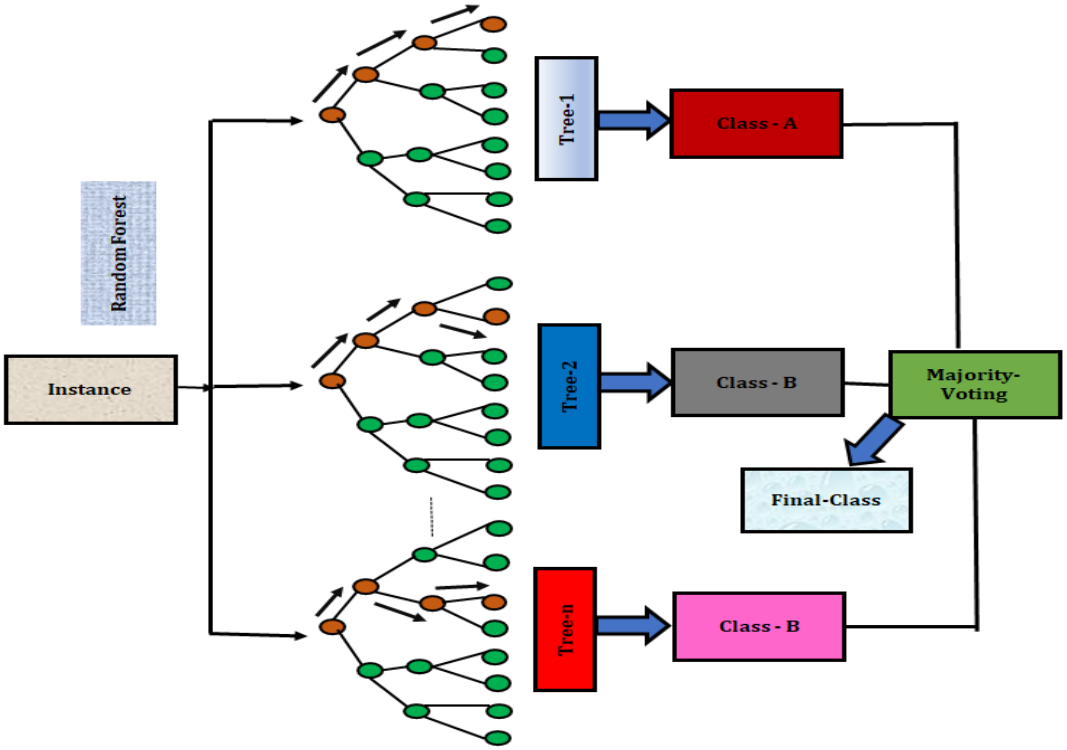

Fig. 6. Random forest.

VII. RESULTS

The CT images used in this study were obtained from Lung Image Database Consortium (LIDC) of the NIH-NCI. For the purpose of lung cancer diagnosis and screening, the Lung Image Database Consortium Image Collection (LIDC-IDRI) contains microscopic slides of thoracic computed tomography images labelled with lesions. 1) CancerTool is a freely available online application designed specifically for the identification and diagnosis of lung cancer using computer-aided diagnostic methods. This dataset provides a comprehensive and annotated collection for researchers in the field, enabling them to improve on computational methods. For the purpose of our computational processes we have utilized LIDC-IDRI database especially in reference to the training dataset. At the feature extraction stage, analogy of origin CT images that were initially 512x512 x3 are downsized to 256x. From this dataset, we took a sample of 500 images to do our experiment. In this investigation, two mutually exclusive labels were taken into consideration: (i) abnormal with Non-Small Cell Lung Cancer (NSCLC) in various stages; (ii) normal, i.e., images displaying no radiological abnormalities. Thirty percent of the available data was set aside for accuracy evaluation and result verification, and the remaining seventy percent was used to train the model. A tabular format was created by computing and organizing metrics such as Accuracy, False Positive Rate, and True Positive Rate (TPR). A small number of samples are taken from the AHRC and AIMS, BG Nagara, Mandya, in order to confirm our findings. In order to validate the data we received, the proposed task was carried out under the supervision of AIMS, a radiologist. The inventors claim that the recommended diagnostic system gives renowned physicians a precise and quick diagnosis.

Performance evaluation metrics like, Recall, Precision, False measure, Sensitivity, Specificity, The machine learning model's evaluation employs accuracy alongside FAR and FRR calculations using the following equations.

$$Re\,c\,all\ R = \frac{TP}{TP+FN} \quad (4)$$

$$Pr\,e\,cision\ \ P = \frac{TP+TN}{TP+FN+FP+FN} \quad (5)$$

$$F-measure\ \ FM = \frac{2\times Re\,call\times Pr\,ecision}{Re\,call + Pr\,ecision} \quad (6)$$

$$Sensitivity\ \ S_e = \frac{TP}{TP+TN} \quad (7)$$

$$Specificity\ \ S_p = \frac{TN}{TP+TN} \quad (8)$$

$$Accuracy\ \ A = \frac{TN+TP}{TP+TN+FP+FN} \quad (9)$$

$$TP\ rate = \frac{TP}{precision}; \quad (10)$$

$$TN\ rate = \frac{TN}{precision} \quad (11)$$

$$FAR = \frac{FP}{FP+TN} \quad (12)$$

$$FRR = \frac{FP}{TP+FN} \quad (13)$$

Table I presents the performance evaluation outcomes of lung cancer prediction utilizing the SVM classifier. The use of a 5-fold cross-validation process was done to guarantee that the model's performance was estimated accurately. The overall

accuracy achieved by the system for predicting a specific type of lung cancer stood at 95.06%. In addition to accuracy, this study scrutinized other pivotal metrics essential for gauging overall performance. These metrics encompass precision, recall, AUC (Area Under the Curve), and F1 score. Each metric was calculated and tabulated for individual classes using the identical 5-fold cross-validation approach, and the findings were consolidated for comprehensive analysis and comparison.

Similarly, Table II presents the performance evaluation outcomes of lung cancer prediction utilizing the KNN classifier. This classifier is also able to classify the subtypes of the lung cancer with an accuracy of 86.89%. Table III. Presents the Performance Analysis of Lung Cancer Using Random Forest Classifier.

The NSCLS with different stages is presented in Fig. 7. In the lung cancer classification, SVM typically overcomes KNN and Random Forest because it can effectively handle complex relationships of data and high-dimensional features as shown in Fig. 8. SVM's ability to detect intricate patterns in datasets, especially those with nuanced characteristics, such as lung cancer data often results in more precision and reliability in classifying the results compared to KNN and Random Forest algorithms across the specific medical field attributed confusion matrix of stark face.

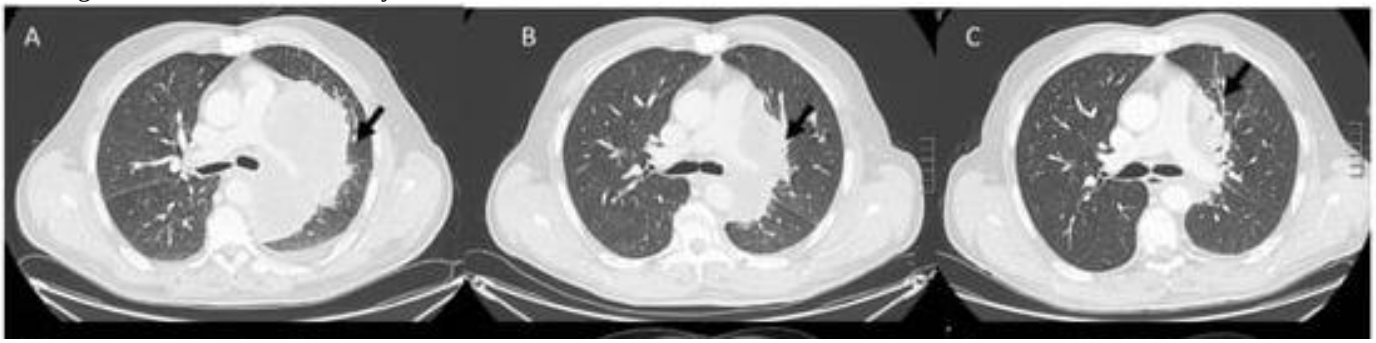

Fig. 7. NSCLC with different stages.

TABLE I. PERFORMANCE ANALYSIS OF LUNG CANCER USING SVM CLASSIFIER

Fold\Class	TPR	FPR	Precision	Recall	F1-Score	Accuracy
1	0.995	0.389	0.994	0.995	0.994	98.90
2	0.991	0.427	0.909	0.991	0.991	91.19
3	0.985	0.048	0.975	0.985	0.980	97.34
4	0.961	0.111	0.947	0.961	0.953	93.71
5	0.999	0.303	0.934	0.999	0.965	94.14
Average	0.99	0.26	0.95	0.99	0.98	95.06

TABLE II. PERFORMANCE ANALYSIS OF LUNG CANCER USING KNN CLASSIFIER

Fold\Class	TPR	FPR	Precision	Recall	F1-Score	Accuracy
1	0.989	0.427	0.893	0.989	0.938	89.85
2	0.988	0.331	0.817	0.988	0.894	86.02
3	0.985	0.299	0.820	0.985	0.895	86.58
4	0.961	0.346	0.807	0.961	0.877	83.84
5	0.999	0.486	0.866	0.999	0.928	88.17
Average	0.98	0.38	0.84	0.98	0.91	86.89

TABLE III. PERFORMANCE ANALYSIS OF LUNG CANCER USING RANDOM FOREST CLASSIFIER

Fold\Class	TPR	FPR	Precision	Recall	F1-Score	Accuracy
1	0.99	0.57	0.76	0.99	0.86	78.84
2	0.99	0.50	0.82	0.99	0.89	83.85
3	0.98	0.37	0.77	0.98	0.86	82.58
4	0.96	0.40	0.77	0.96	0.85	80.78
5	1.00	0.58	0.82	1.00	0.90	83.75
Average	0.98	0.48	0.78	0.98	0.87	81.96

TABLE IV. EVALUATE OUR MODEL AGAINST CURRENT BEST PRACTICES

Sl. No	Authors	Accuracy in %
1	Murphy et al. [17]	84.00
2	Messay et al. [18]	82.66
3	Gomathi et al. [19]	76.90
4	Kumer et al. [20]	86.00
5	Proposed	95.05

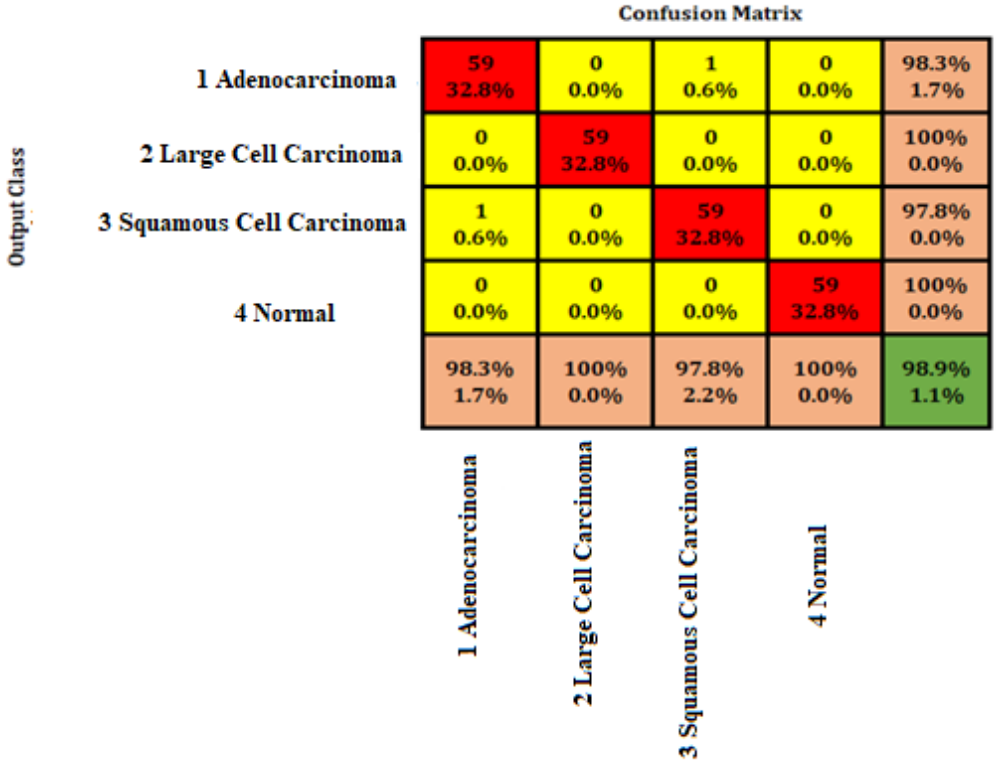

Fig. 8. Confusion Matrix (SVM Classifier).

VIII. DISCUSSIONS

The four main kinds of lung cancer adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal are all correctly identified by the suggested approach. Fig. 9 illustrates the Confusion Matrix associated with the classification for lung cancer. The Confusion Matrix depicted therein highlights correctly categorized samples via green diagonal points, while non-diagonal points signify misclassified samples. This visualization aids in understanding the model's performance in accurately classifying different types of lung cancers, emphasizing correct classifications (green points) and areas needing improvement (non-diagonal points).

Notably, the ROC-AUC values exceed 0.99 for every class, underscoring the remarkable proficiency of the model in accurately distinguishing between these diverse classes. In comparative performance among KNN, Random Forest, and SVM, while both KNN and Random Forest exhibit strengths in certain scenarios due to their simplicity and ensemble nature respectively, SVM stands out for its robustness in handling complex decision boundaries and high-dimensional data as in Fig. 10. SVM often excels when the dataset is characterized by intricate relationships between features, as it effectively finds the optimal hyperplane to separate classes, leading to superior generalization and accuracy in such situations. A comparison between our generated model and the most advanced model in the field is shown in Table IV and Fig. 11. Our model outperforms other models in terms of accuracy, according to the findings. This comparison shows that our model is somewhat more accurate than the current state-of-the-art models in the field.

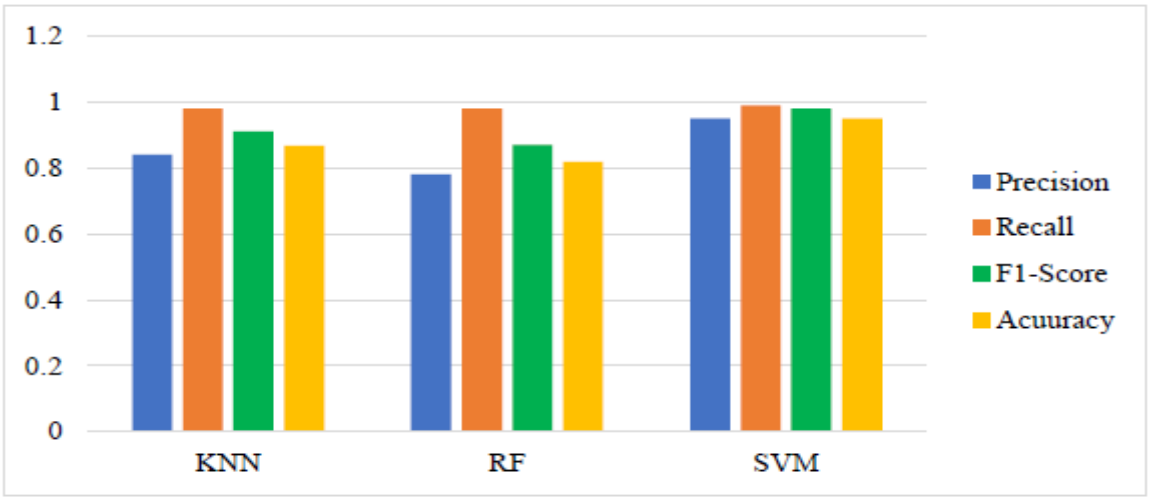

Fig. 9. Performance analysis of classifiers.

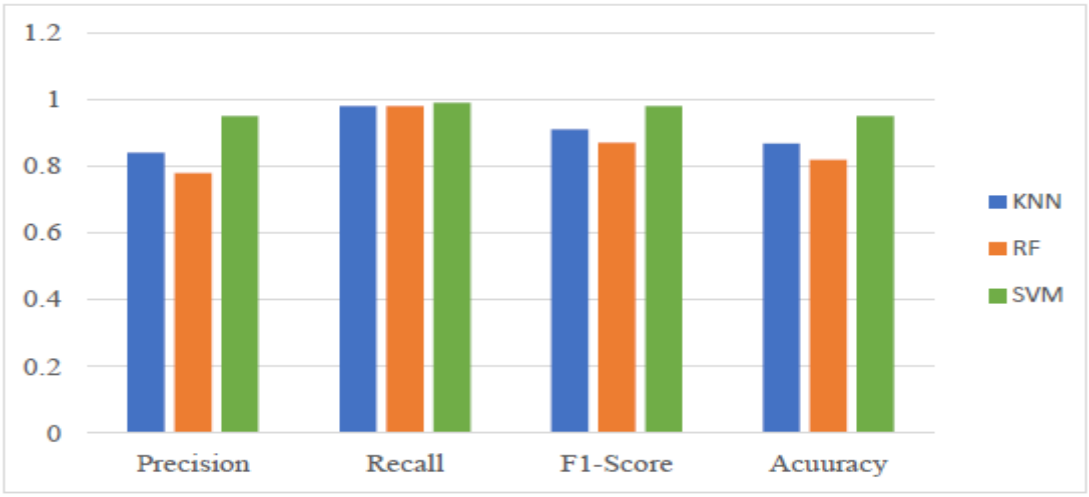

Fig. 10. Comparison of metrics with classifiers.

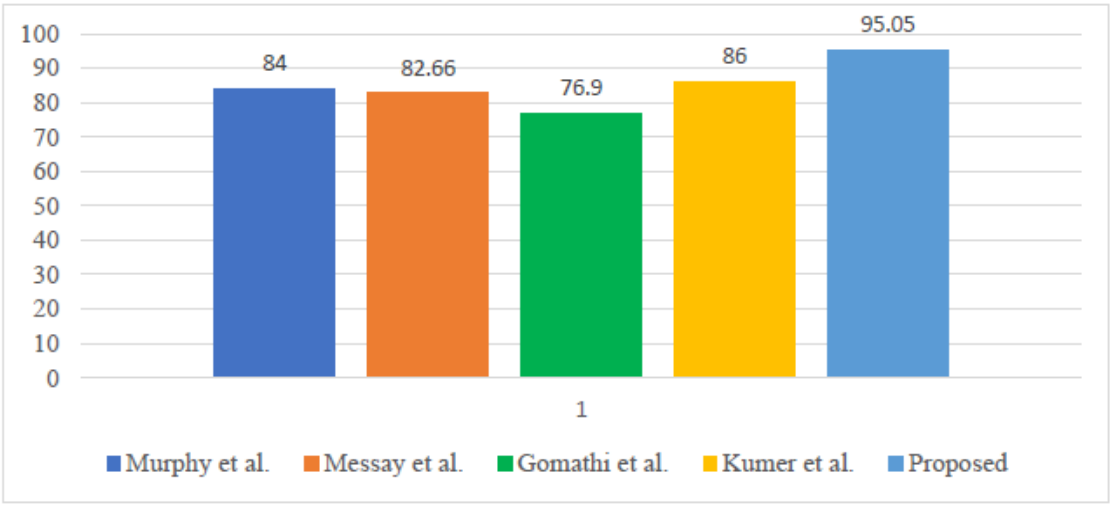

Fig. 11. Compared to current best-practice models, the proposed model.

IX. Conclusion and Future Work

The study utilizes 500 lung images for predictive analysis within the proposed algorithm. The training dataset for lung cancer is obtained from a recognized machine learning database. CT images sourced from the NIH-NCI - LIDC form the basis of this research. Its objective is to assess and contrast the performance of three classifiers in early-stage lung cancer diagnosis. Results reveal that the SVM demonstrated the highest accuracy, reaching 95.05%. This finding highlights SVM's potential in detecting lung cancer in its early stages, potentially contributing to saving numerous lives. Conversely, the random forest and KNN algorithm exhibited lower accuracy, recording 81.96% and 88.40% respectively. Finally, the proposed work underwent validation using samples collected from AHRC and AIMS, BG Nagara, Mandya. Under the supervision of a radiologist at AIMS, these datasets were employed to authenticate our obtained results. As per the developers, this diagnostic system offers top doctors an accurate and swift diagnosis, and we utilized these datasets for validation purposes. Based on the satisfactory outcomes of this research, future work will focus on improving the performance and reliability of the proposed model utilizing a more comprehensive and diverse data collection with images of people from different backgrounds and different stages of lung cancer.

References

[1] Huang, Shigao, Ibrahim Arpaci, Mostafa Al-Emran, Serhat Kılıçarslan, and Mohammed A. Al-Sharafi. “A comparative analysis of classical machine learning and deep learning techniques for predicting lung cancer survivability.” Multimedia Tools and Applications 82, no. 22 (2023): 34183-34198.

[2] Chassagnon, Guillaume, Constance De Margerie-Mellon, Maria Vakalopoulou, Rafael Marini, Trieu-Nghi Hoang-Thi, Marie-Pierre Revel, and Philippe Soyer. “Artificial intelligence in lung cancer: current applications and perspectives.” Japanese Journal of Radiology 41, no. 3 (2023): 235-244.

[3] Dritsas, Elias, and Maria Trigka. “Lung cancer risk prediction with machine learning models.” Big Data and Cognitive Computing 6, no. 4 (2022): 139.

[4] Anil Kumar, C., S. Harish, Prabha Ravi, Murthy Svn, B. P. Kumar, V. Mohanavel, Nouf M. Alyami, S. Shanmuga Priya, and Amare Kebede Asfaw. “Lung cancer prediction from text datasets using machine learning.” BioMed Research International 2022 (2022).

[5] Altuhaifa, Fatimah Abdulazim, Khin Than Win, and Guoxin Su. “Predicting lung cancer survival based on clinical data using machine learning: A review.” Computers in Biology and Medicine (2023): 107338.

[6] Huang, Shigao, Ibrahim Arpaci, Mostafa Al-Emran, Serhat Kılıçarslan, and Mohammed A. Al-Sharafi. “A comparative analysis of classical machine learning and deep learning techniques for predicting lung cancer survivability.” Multimedia Tools and Applications 82, no. 22 (2023): 34183-34198.

[7] Kavitha B C, Dr. Naveen K B, “Detection of Lung Cancer using VGG-16 Model,” Neuro Quantology, September 2022, Volume 20, Issue 9, Page 3747-3751, 2022.

[8] Wang, Fangwei, Qisheng Su, and Chaoqian Li. “Identidication of novel biomarkers in non-small cell lung cancer using machine learning.” Scientific Reports 12, no. 1 (2022): 16693.

[9] Mamun, Muntasir, Afia Farjana, Miraz Al Mamun, and Md Salim Ahammed. “Lung cancer prediction model using ensemble learning techniques and a systematic review analysis.” In 2022 IEEE World AI IoT Congress (AIIoT), pp. 187-193. IEEE, 2022.

[10] He, Ruimin, Xiaohua Yang, Tengxiang Li, Yaolin He, Xiaoxue Xie, Qilei Chen, Zijian Zhang, and Tingting Cheng. “A machine learning-based predictive model of epidermal growth factor mutations in lung adenocarcinomas.” Cancers 14, no. 19 (2022): 4664.

[11] Sarma, Parismita & Rahman, Mirzanur (2024) Mathematical analysis of wavelet-based multi-image compression in medical diagnostics, Journal of Discrete Mathematical Sciences and Cryptography, 27:2-B, 675–687.

[12] Dr. Nagaraju C , SanganaManasa Durga, Rachana N, “Ascertainment of Lung Cancer at an Early Stage”, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, ISSN : 2456-3307,Volume 2, Issue 4, 2017.

[13] Shah, Syed Naseer Ahmad, and Rafat Parveen. “An Extensive Review on Lung Cancer Diagnosis Using Machine Learning Techniques on Radiological Data: State-of-the-art and Perspectives.” Archives of Computational Methods in Engineering 30, no. 8 (2023): 4917-4930.

[14] R. c. Tanguturi and G. Poornima, "An Intelligent System for Remote Monitoring of Patients Health and the Early Detection of Coronary Artery Disease," 2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), Bangalore, India, 2022, pp. 1-6.

[15] B. Ashreetha and N. S. Reddy, "Implementation of a Machine Learning-based Model for Cardiovascular Disease Post Exposure prophylaxis," 2023 International Conference for Advancement in Technology (ICONAT), Goa, India, 2023, pp. 1-5.

[16] Wu, Qing, Yi-Ming Qian, Xiang-Li Zhao, Shou-Mei Wang, Xiao-Jun Feng, Xin-Fang Chen, and Shu-Hui Zhang. “Expression and prognostic significance of centromere protein A in human lung adenocarcinoma.” Lung Cancer 77, no. 2 (2012): 407-414.

[17] Asuntha, A., A. Brindha, S. Indirani, and Andy Srinivasan. “Lung cancer detection using SVM algorithm and optimization techniques.” J. Chem. Pharm. Sci 9, no. 4 (2016): 3198-3203.

[18] Kavitha B C, Dr. Naveen K B, “Image Acquisition and Pre-processing for Detection of Lung Cancer using Neural Network”, Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT- 2022), December 26-27, 2022.

[19] S. R. Kawale, S. P. Diwan, "Intelligent Breast Abnormality Framework for Detection and Evaluation of Breast Abnormal Parameters," 2022 International Conference on Edge Computing and Applications (ICECAA), 2022, pp. 1503-1508.

[20] Bharati, Subrato, Prajoy Podder, and Pinto Kumar Paul. “Lung cancer recognition and prediction according to random forest ensemble and RUSBoost algorithm using LIDC data.” International Journal of Hybrid Intelligent Systems 15, no. 2 (2019): 91-100.

[21] Altuhaifa, Fatimah Abdulazim, Khin Than Win, and Guoxin Su. "Predicting lung cancer survival based on clinical data using machine learning: A review." Computers in Biology and Medicine (2023): 107338.

[22] Ghazaala Yasmin, Parismita Sarma, A. Azhagu Jaisudhan Pazhani, A novel method of data compression using ROI for biomedical 2D images, Measurement: Sensors, Volume 24, 2022, 100439, ISSN 2665-9174.

[23] Murphy, Stephen J., Marie-Christine Aubry, Faye R. Harris, Geoffrey C. Halling, Sarah H. Johnson, Simone Terra, Travis M. Drucker et al. “Identification of independent primary tumors and intrapulmonary metastases using DNA rearrangements in non–small-cell lung cancer.” Journal of Clinical Oncology 32, no. 36 (2014): 4050.

[24] Messay, Temesguen, Russell C. Hardie, and Steven K. Rogers. “A new computationally efficient CAD system for pulmonary nodule detection in CT imagery.” Medical image analysis 14, no. 3 (2010): 390-406.

[25] Gomathi, E. “A Novel Lung Cancer Segmentationand Classificationusing Ann.” Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12, no. 11 (2021): 4298-4304.

[26] Kumar, Vaibhav, Joshua T. Cohen, David van Klaveren, Djøra I. Soeteman, John B. Wong, Peter J. Neumann, and David M. Kent. “Risk-targeted lung cancer screening: a cost-effectiveness analysis.” Annals of internal medicine 168, no. 3 (2018): 161-169.