International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Machine Learning Approach for Lung Cancer Detection and Classification–A Comparative Analysis

Prakasha Raje Urs M[1], Dr. G N K Suresh Babu[2]

Submitted: 25/01/2024 **Revised**: 03/03/2024 **Accepted**: 11/03/2024

Abstract: The low percentages of cure for advanced stages of lung cancer highlight how crucial early discovery is to improving prognoses. Therefore, identifying the disease in its early stages is a potential direction in lung cancer diagnosis research. Using Principal Component Analysis (PCA) for feature extraction and Gray-Level Co-occurrence Matrix (GLCM) features for detection and classification of lung cancer, the proposed study compares several machine learning techniques. The suggested techniques are assessed using three classifiers: Naive Bayes (NB), Decision Tree (DT), and Support Vector Machine (SVM). The goal of the study is to determine which of these classifiers is the best at correctly recognising and classifying cases of lung cancer. This study advances the understanding of machine learning strategies for improving lung cancer diagnosis and classification, perhaps leading to better patient outcomes, by carefully examining the effectiveness of each approach. Using Local Binary Patterns (LBP) to extract features led to notable improvements in all algorithms. The results show that LBP features are useful in increasing classification performance: Naive Bayes (NB) attained an accuracy of 0.851, Decision Trees (DT) to 0.912, and Support Vector Machine (SVM) to 0.961.

Keywords: *Decision Tree, Gray-Level Co-occurrence Matrix, Lung Cancer, Machine Learning, Naive Bayes, Principal Component Analysis.*

1. Introduction

Nearly Lung cancer is a serious and potentially fatal illness that frequently goes undiagnosed, increasing the death rate. To remove carbon dioxide from the body and give oxygen to the body, the lungs are essential. Uncontrollably growing lung tissue and cells that have the potential to spread to neighbouring tissues are the cause of lung cancer [1]. It is the primary cause of cancer-related deaths in men and the second in women worldwide, accounting for over 1.3 million deaths yearly, with 30–40,000 new cases reported in Turkey. Compared to other cancers like colorectal, pancreatic, and breast cancers, lung cancer has a greater death rate. By 2050, it is anticipated that the number of cancer cases would have doubled, with lung cancer playing a significant role. One factor contributing to the high death rate from lung cancer is late-stage diagnosis [2]. Understanding its genesis, early detection methods, and suitable therapy are necessary for improving results. Along with the population's rapid rise, diseases like cholera, chikungunya, cancer, and others are becoming more common. One of these diseases that is currently the leading cause of death worldwide is cancer [3]. Cancer can begin in any one of the billions of cells that make up the human body. Human cells normally divide and grow in a regulated way to produce new cells as needed by the body. As they degenerate or age, cells undergo programmed cell death to make way for new ones. However, the onset of cancer upsets this well-ordered process [4]. Efficient evaluation is essential for identifying irregularities in medical images, especially in conditions like breast and lung cancer. At different stages of treatment, different image processing techniques are used to enhance image quality for early diagnosis. Modalities like computed tomography (CT), positron emission tomography (PET), and X-ray are often used to assess lung nodules, even though they can be expensive and time-consuming. Therefore, there is an urgent need for a novel Computer-Aided Design (CAD) system that can accurately identify lung cancer in its early stages [5]. Many of the existing solutions do not meet the needs of radiologists, even though CAD systems are indispensable to medical radiology. Studies have indicated that early detection of lung tumours using CT scans might significantly improve patient survival rates, perhaps reaching 90% [6]. Consequently, CT scans are often the suggested modality for CAD systems. These CT images are typically sourced from institutions like the Cancer Image Archive, which specialise in cancer analysis and maintain enormous datasets of medical imaging. A CT scan of Figure 1 reveals lung cancer. The leading cause of cancer-related death worldwide is still non-small cell lung cancer (NSCLC) [7]. The five-year survival rate for NSCLC patients is less than 25%. Due to the rarity of early symptoms, most NSCLC lesions are already locally advanced at the time of initial diagnosis. NSCLC is divided into two main histological subtypes: adenocarcinoma (ADC) and squamous cell

[1]*Research Scholar, Srishti College of Commerce and Management, Banashankari, Bengaluru, Affiliated to University of Mysore, India, 560085.*
[2]*Professor, Department of Computer Science, Srishti College of Commerce and Management, Banashankari, Bengaluru, Affiliated to University of Mysore, India, 560085.*

carcinoma (SCC). The best course of action for each patient's treatment is the best way to lower the NSCLC death rates. One of the most important prognostic factors is still accurate lesion characterization of the difference between ADC and SCC [8]. The lesion's histological subtype determines the best course of treatment. individuals with ADC may benefit from treatment with tyrosine kinase inhibitors, a drug that has shown therapeutic efficacy in ADC cases. However, more testing is required to identify whether EGFR (epidermal growth factor receptor) mutations are present in these individuals. Treatment for SCC does not benefit from these inhibitors [9]. The efficiency of treatment and the likelihood of side effects varies for SCC and ADC, and the same is true for immunotherapy and chemotherapy. NSCLC sub-type identification is primarily based on cytological or histological examination [10]. Regretfully, there is a significant chance of problems with these testing. The sensitivity of these approaches ranges from 79% to 98%, depending on the location and procedures of the tumour [11]. There is a continuous search for novel techniques that enable the non-invasive detection of NSCLC histological subtypes. Adenocarcinoma, squamous cell carcinoma, and small cell carcinoma are the three most common kinds of lung cancer, with an accuracy rate of almost 71%, according to an analysis of microscopic pictures [12]. Deep convolution has been used in this instance to classify carcinoma and adenocarcinoma. A result of almost 80% was obtained from other studies that used histological images of two forms of non-small cell lung cancer, squamous cell carcinoma and adenocarcinoma [13]. A three-step automatic classification technique based on deep learning was utilised to achieve this figure, which identified both statistical and morphological features. and a random forests-based regression model [14].

Researchers can use machine learning algorithms to evaluate how well different methods identify and categorise cases of lung cancer. By accurately assessing medical imaging, such as CT scans, machine learning (ML) contributes significantly to the early diagnosis and treatment planning of lung cancer [15]. Machine learning algorithms can identify minute patterns and irregularities in lung pictures that might not be readily discernible to human viewers. This might result in prompt intervention and enhanced patient results [16]. As machine learning (ML) continues to progress, radiologists may rely on ML to help them identify lung cancer early on, improving patient care overall and screening program effectiveness. The introduction, Methodology, Results and Discussion, and Conclusion parts make up the paper's organizational structure. The Introduction sets the scene for lung cancer diagnosis while highlighting the importance of cutting-edge machine learning techniques. The methodology descry bes the methods used, such as different classifiers and GLCM feature extraction. The findings and their interpretation are presented in the Results and Discussion section, and future study directions and critical insights are summarized in the Conclusion.

2. Materials and Methods

In the field of medical diagnostics, accurate lung cancer diagnosis and categorization are crucial since they greatly impact treatment options for patients and their overall prognosis. The application of more sophisticated deep learning and machine learning techniques has resulted in a discernible shift in the direction of automating and enhancing this diagnostic procedure. By utilizing machine learning algorithms and certain feature extraction techniques, scientists can distinguish between patterns of malignant and non-cancerous tissue in lung pictures. Moreover, these sophisticated models can classify different subtypes of lung cancer according to distinctive features that may be identified in the imaging data. Through enabling early detection, facilitating rapid intervention, and ultimately improving treatment efficacy, such advancements hold significant potential for improving patient outcomes in the management of lung cancer. The chance of surviving at an advanced stage of the disease is reduced when compared to the treatment and lifestyle to survive cancer therapy when diagnosed at an early stage. This paper outlines ML-based approach for lung cancer detection from CT images. It begins by extracting lung regions from the images and then segments each slice within these regions.

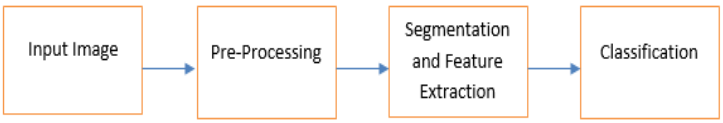

Figure 1 Lung cancer Prediction System

3. Proposed Model

The methodology for identifying and categorizing lung cancer entails multiple crucial processes. First, a dataset containing pictures from lung CT scans—both healthy and malignant cases—is gathered. To guarantee consistency across the dataset and boost quality, these photos go through preprocessing procedures like scaling, normalization, and contrast enhancement. Then, each image's important features are extracted, capturing structural patterns and texture using methods like Principal Component Analysis (PCA) and Gray-Level Co-occurrence Matrix (GLCM). GLCM provides information about the spatial distribution of pixel values in a picture, whereas PCA reduces the dimensionality of the feature space by translating the original features into a lower-dimensional space.

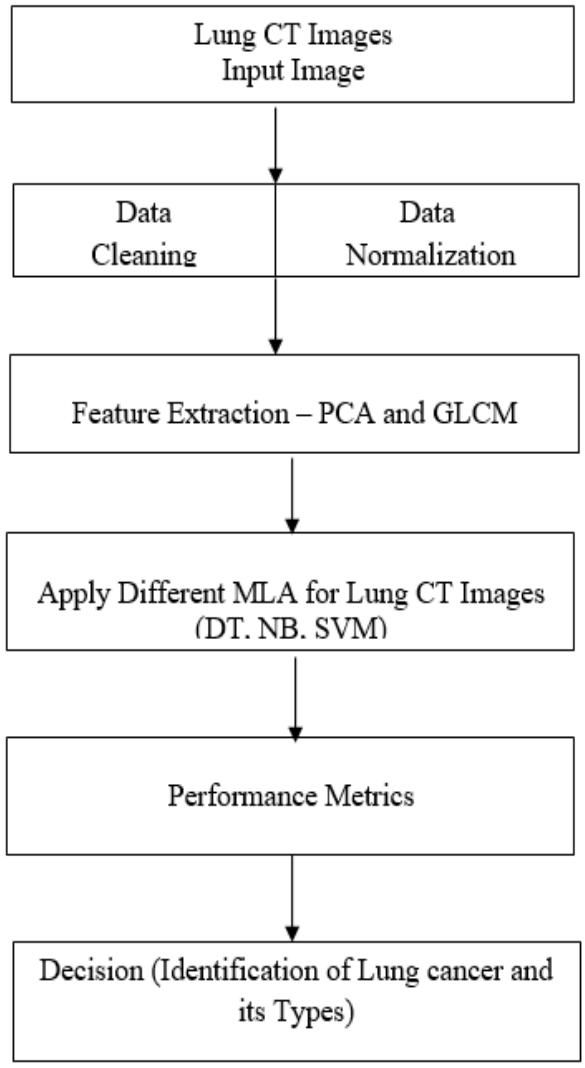

Figure 2: Proposed model for Lung cancer analysis

Subsequently, the dataset is divided into testing and training sets, often with a 70–30 split. The training set is used to train machine learning classifiers like Naive Bayes (NB), Decision Tree (DT), and Support Vector Machine (SVM). These classifiers identify the relationships between the retrieved features and the presence or absence of lung cancer. Finally, the performance of the trained classifiers is evaluated using the testing set. Metrics like accuracy, precision, recall, and F1-score are used to assess how well each classifier determines and groups cases of lung cancer. All things considered, this methodology builds a robust system for lung cancer identification and classification by combining machine learning classifiers with feature extraction techniques like GLCM and PCA.

3.1 Dataset:

Lung cancer research frequently makes use of several publicly accessible datasets, especially when it comes to machine learning and medical imaging analysis. For the research the dataset has been gathering from LIDC-IDRI dataset. This data set includes. Benign and malignant. Along with these various classes like Adenocarcinoma, Large cell carcinoma, Squamous cell carcinoma and Normal.

3.2 Pre-processing

Pre-processing for lung cancer typically involves extracting relevant features from CT scan images to enhance classification accuracy. In the proposed research it includes resizing, normalization, and contrast enhancement to ensure consistency and improve dataset quality. CLAHE and Gaussian filtering are utilized to normalize and denoise images, preparing them for machine learning analysis.

Principal Component Analysis (PCA)

One of the most important methods for dimensionality reduction in the field of lung cancer detection and classification is Principal Component Analysis (PCA). In terms of mathematics, PCA seeks to maintain the maximum variance of the original data while converting a high-dimensional dataset into a lower-dimensional subspace. By locating the main components that represent the most variance in the data, PCA is a technique used in the analysis of CT scan pictures for lung cancer that reduces the dimensionality of the images [17]. This decrease enables more effective feature extraction, assisting in the minimal information loss in the classification of lung cancer.

3.3 Gray-Level Co-occurrence Matrix (GLCM)

The Gray-Level Co-occurrence Matrix (GLCM) is a powerful tool in lung cancer classification, capturing texture patterns in medical images by quantifying spatial relationships between pixel intensities [18]. Texture features extracted from GLCM, including energy, entropy, contrast, correlation, and homogeneity, provide discriminative markers for differentiating lung cancer types and stages. Integrating GLCM-based texture analysis into classification pipelines enhances diagnosis precision, enabling personalized treatments and better patient outcomes. GLCM feature extraction involves using equations to quantify the occurrence frequency of two pixels in lung images [19].

$$contrast = \sum_{k=0}^{n} p_{i,j} i - j^2$$

$$homogeneity = \sum_{i,j=0}^{n-1} \frac{p_{i,j}}{1+ij^2}$$

$$angular\ \ second\ \ moment = \sum_{i,j=0}^{n-1} p_{i,j}^2$$

$$entropy = -\sum_{i,j=0}^{n-1} p_{i,j} \ln\ \ p_{i,j}$$

3.4 Model Training:

3.4.1 Naive Bayes:

The Naive Bayes algorithm is utilized for cancer prediction, leveraging feature independence to compute probabilities for different groups based on available features, such as characteristics extracted from CT images for lung cancer prediction [20]. After training with labeled data, it generates predictions for new images, and classification accuracy is assessed using a confusion matrix [21]. In Figure 4, labels C1, C2, and C3 represent distinct classes, with the highest probability indicating that an incoming data point X likely belongs to class C1.

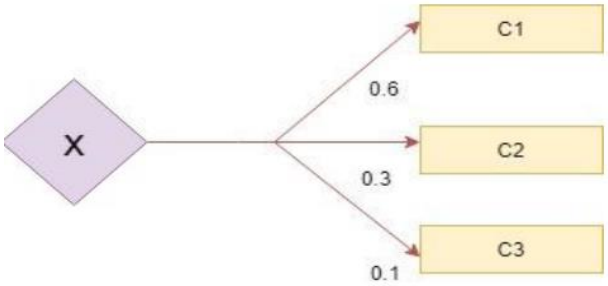

Figure 3: Example of Data and Probability Relations

The following equation were used to calculate the probabilities.

$$\frac{P(C1/X)}{P(C2/X)} > 1$$

$$\frac{P(C1/X)}{P(C2/X)} = \frac{\frac{P(X/C1).P(C1)}{P(X)}}{\frac{P(X/C2).P(C2)}{P(X)}} \qquad (2)$$

3.4.2 Decision tree algorithm

The decision tree algorithm, commonly used for classification tasks, is also applicable for cancer prediction. It constructs a tree based on available data, analyzing features to determine the best splits for accurate classification. The objective is to create a predictive model that assigns classes to instances based on their features [22]. Within a decision tree framework, individual nodes represent specific symptoms drawn from the set S = {s1, s2, s3, ..., sj}, where 'S' denotes conditional attributes. Each branch, denoted by vi, 'k', corresponds to the values of each symptom, representing the hth range for the ith symptom. At the leaves, decisions D = {d1, d2, ..., dk} are presented along with their binary values, wdk={0, 1}. By documenting each path from the root to the leaves, a collection of association rules was generated by converting the decision tree structure [23].

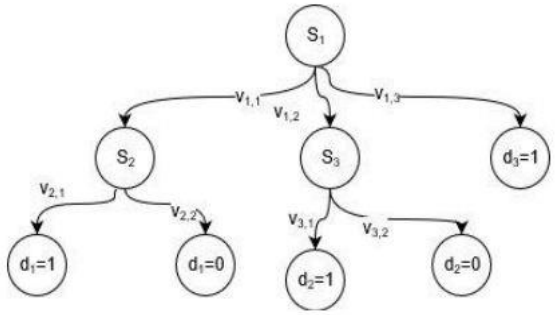

Figure 4: Sample Structure of Decision Tree

The set of association rules for the above given tree are:

$$(S_1, v_{1,1}) \wedge (S_2, v_{2,1}) \Rightarrow (d_1 = 1)$$
$$(S_1, v_{1,3}) \wedge (S_2, v_{2,2}) \Rightarrow (d_1 = 0)$$
$$(S_1, v_{1,2}) \wedge (S_3, v_{3,3}) \Rightarrow (d_2 = 1)$$
$$(S_1, v_{1,2}) \wedge (S_3, v_{3,2}) \Rightarrow (d_2 = 0)$$
$$(S_1, v_{1,3}) \Rightarrow (d_3 = 1) \qquad (1)$$

3.5 Support Vector Machine (SVM) :

The Support Vector Machine (SVM) algorithm is a powerful tool for cancer prediction, especially with high-dimensional image data like CT scans. It operates by separating the dataset into two classes using kernel functions and a hyperplane in a three-dimensional space [24]. For instance, in lung cancer detection from CT scans, SVM with a radial basis function (RBF) kernel is employed after pre-processing the images to extract relevant features. The model is trained with labeled images, distinguishing between normal and abnormal tumors, and then tested with new data. Evaluation using a confusion matrix reveals the accuracy of the SVM classifier's predictions [25].

4. Results and Discussions

The lung cancer dataset used in the research consists of 600 images, with 150 images per class (Adenocarcinoma, Large cell carcinoma, Squamous cell carcinoma, and Normal). MATLAB was used on an i5 system for model development. Machine learning methods such as Naive Bayes (NB), Decision Trees (DT), and Support Vector Machines (SVM) have been developed for lung cancer detection, employing feature selection techniques like Gray-Level Co-occurrence Matrix (GLCM) and Local Binary Patterns (LBP) to enhance accuracy and robustness in classification. Random selection of training and testing data ensured unbiased evaluation. Statistical parameters gauged model performance, aiding in accurate lung cancer classification. The classification model underwent 5-fold cross-validation, dividing the dataset into five subsets for testing and training iteratively. This method ensures robust

evaluation and reliable performance assessment across different data partitions. Figure 5 showcases sample images employed for lung cancer classification.

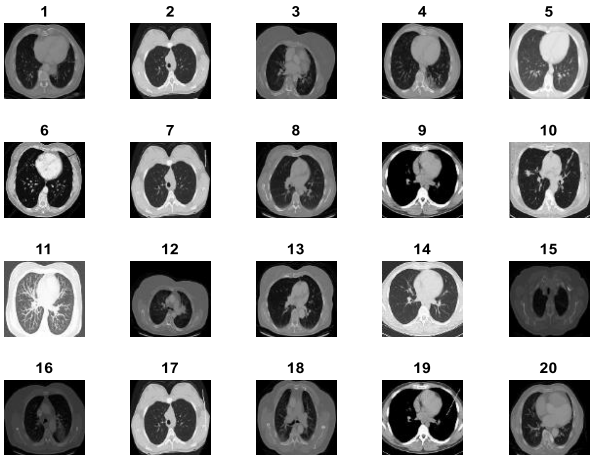

Figure 5: Feature Extraction using LBP

The plot of the minimum objective function value versus the number of function evaluations in MATLAB typically indicates the progress of an optimization algorithm. The plot of minimum objective versus function evaluations serves as a valuable tool for understanding and evaluating the performance of optimization algorithms, guiding their application, and fine-tuning parameters for better results. Figure 6 demonstrates a decreasing trend, indicating convergence towards an optimal solution.

The classification accuracies of NB, DT, and SVM were assessed for a task utilizing GLCM features extracted and LBP as in table 1 and table 2 respectively. Initially, without GLCM feature extraction, NB achieved an accuracy of 0.792, DT reached 0.872, and SVM obtained 0.921. Upon employing LBP feature extraction, noticeable improvements were observed across all algorithms. Specifically, NB's accuracy increased to 0.851, DTs to 0.912, and SVM's to 0.961, showcasing the efficacy of LBP features in enhancing classification performance. Notably, SVM demonstrated the highest accuracy, indicating its effectiveness in leveraging these features for accurate classification. These results underscore the importance of feature selection and extraction techniques in improving machine learning model performance for classification tasks.

Table1: Statistical Parameter for GLCM

Class	Precision	Recall	F1-Score	AUC	Accuracy
NB	0.816	0.910	0.829	0.918	0.792
DT	0.832	0.846	0.892	0.965	0.872
SVM	0.911	0.975	0.987	0.911	0.921

Table 2: Statistical Parameter for LBP

Class	Precision	Recall	F1-Score	AUC	Accuracy
NB	0.89	0.928	0.989	0.980	0.851
DT	0.732	0.935	0.835	0.919	0.912
SVM	0.946	0.918	0.925	0.958	0.961

Table 3 illustrates the statistical parameters obtained through 5-fold cross-validation for the implementation of LBP in conjunction with SVM for lung cancer detection and classification. The table provides a comprehensive overview of the performance metrics achieved by this method, highlighting its efficacy in accurately identifying and categorizing instances of lung cancer.

Table 3: Statistical Parameter for LBP and SVM

Fold	Class	TNR	FNR	FPR	TPR	EER	Precision	Recall	F1-Score	Accuracy	AUC
1	1	0.94	0.15	0.06	0.85	0.11	0.82	0.85	0.83	0.92	0.97
	2	0.95	0.19	0.05	0.81	0.12	0.85	0.81	0.83	0.92	0.99
	3	0.99	0.07	0.01	0.93	0.04	0.94	0.93	0.93	0.97	0.98
	4	0.98	0.02	0.02	0.98	0.02	0.95	0.98	0.96	0.98	0.99
2	1	0.94	0.05	0.06	0.95	0.05	0.90	0.95	0.93	0.94	0.96
	2	0.97	0.11	0.03	0.89	0.07	0.85	0.89	0.87	0.96	0.98
	3	0.98	0.10	0.02	0.90	0.06	0.90	0.90	0.90	0.96	0.99
	4	0.97	0.02	0.03	0.98	0.03	0.93	0.98	0.95	0.97	0.98
3	1	0.96	0.03	0.04	0.97	0.04	0.93	0.97	0.95	0.96	0.96
	2	0.96	0.08	0.04	0.92	0.06	0.91	0.92	0.92	0.96	0.95
	3	0.96	0.07	0.04	0.93	0.05	0.92	0.93	0.93	0.95	0.96
	4	0.82	0.35	0.18	0.65	0.26	0.56	0.65	0.60	0.78	0.82
4	1	0.94	0.06	0.06	0.94	0.06	0.87	0.94	0.90	0.94	0.99
	2	0.96	0.22	0.04	0.78	0.13	0.83	0.78	0.80	0.93	0.99
	3	0.99	0.13	0.01	0.87	0.07	0.95	0.87	0.91	0.97	0.99
	4	0.98	0.03	0.02	0.97	0.02	0.97	0.97	0.97	0.98	0.99
5	1	0.95	0.07	0.05	0.93	0.06	0.84	0.93	0.88	0.95	0.95
	2	0.95	0.21	0.05	0.79	0.13	0.87	0.79	0.83	0.91	0.95
	3	0.96	0.17	0.04	0.83	0.10	0.86	0.83	0.84	0.93	0.96
	4	0.97	0.05	0.03	0.95	0.04	0.93	0.95	0.94	0.96	0.97

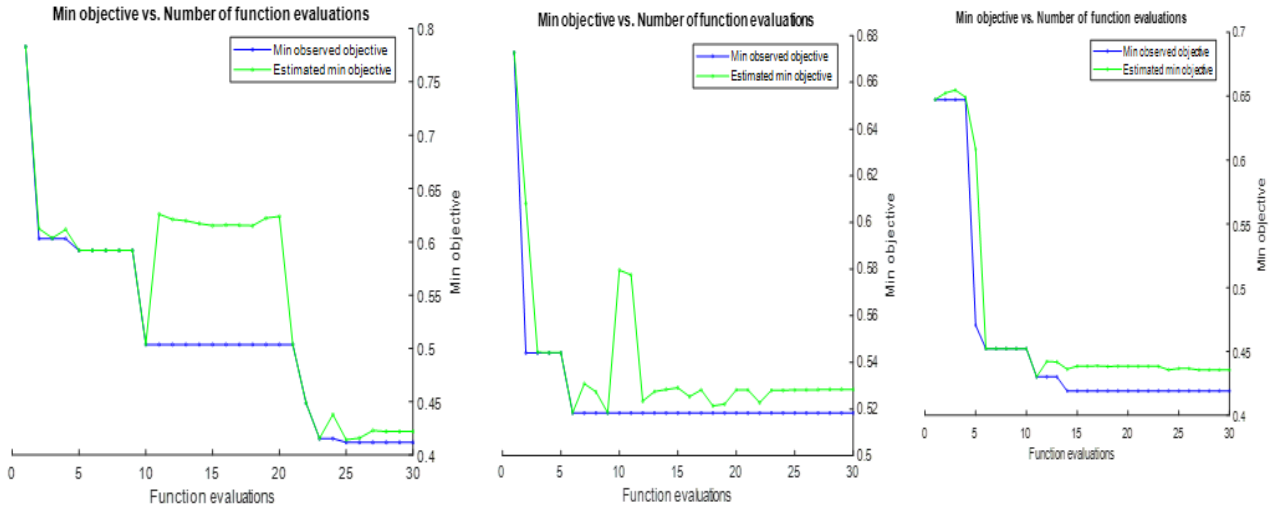

Figure 6: Hyperparameter curve for SVM, NB and DT using LBP

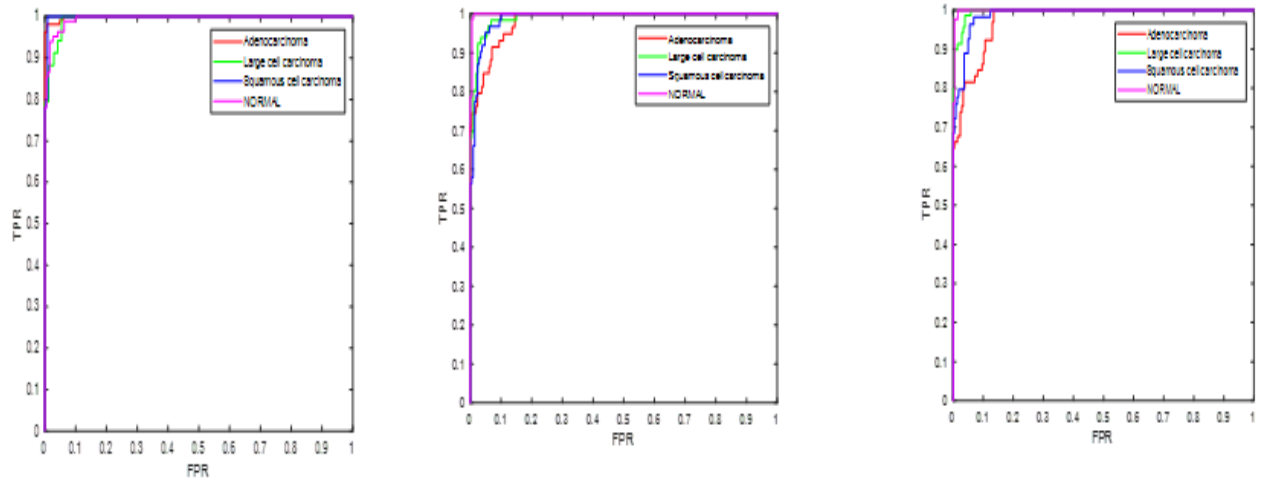

Figure 7: ROC Curve for SVM, NB and DT using LBP

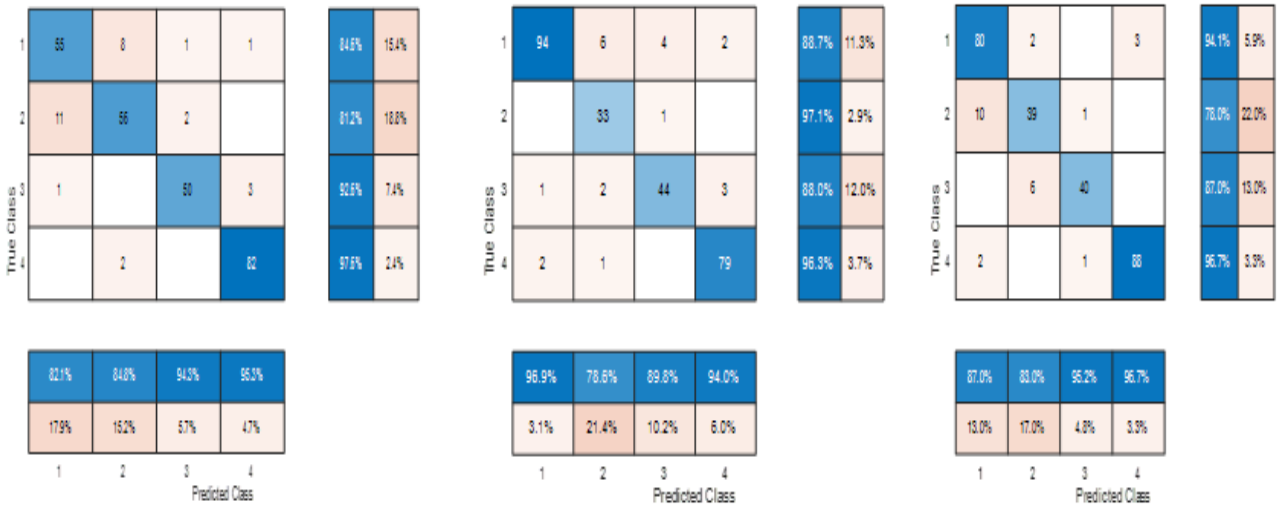

Figure 8: Confusion Matrix for SVM, NB and DT using LBP

Figure 6 shows the hyperparameter curve, Figure 7 the Receiver Operating Characteristic (ROC) curve, and Figure 8 the confusion matrix for the LBP -based NB, DT, and SVM classifiers used to detect and classify lung cancer. When combined, these numbers show how well SVM and LBP work, highlighting their efficacy as the most promising combination for a reliable and accurate classification of lung cancer.

5. Conclusions

By using PCA for feature extraction and GLCM features for detection and classification, the proposed study compares several machine learning techniques for lung cancer detection and classification. Three classifiers—NB, DT, and SVM—are assessed for efficacy in this study. The purpose of this comparison analysis is to evaluate how well the suggested approaches work in correctly identifying and classifying cases of lung cancer. In summary, compared to their performance without feature extraction, the experimentation with GLCM features extracted using LBP greatly improved the classification accuracies of NB, DT, and SVM. The utilization of GLCM-LBP features led to notable accuracy improvements across all algorithms, with SVM achieving the highest accuracy of 0.961. These findings underscore the effectiveness of GLCM-LBP feature extraction in improving the discriminatory power of machine learning models, highlighting its potential for enhancing classification performance in various applications.

REFERENCES

[1] Petty, W. Jeffrey, and Luis Paz-Ares. "Emerging strategies for the treatment of small cell lung cancer: a review." JAMA oncology 9, no. 3 (2023): 419-429.

[2] Chhikara, Bhupender S., and Keykavous Parang. "Global Cancer Statistics 2022: the trends projection analysis." Chemical Biology Letters 10, no. 1 (2023): 451-451.

[3] Liang, Wenhua, Kaican Cai, Qingdong Cao, Chun Chen, Haiquan Chen, Jun Chen, Ke-Neng Chen et al. "International expert consensus on immunotherapy for early-stage non-small cell lung cancer." Translational Lung Cancer Research 11, no. 9 (2022): 1742.

[4] 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.

[5] Hata, Akinori, Takuya Hino, Masahiro Yanagawa, Mizuki Nishino, Tomoyuki Hida, Gary M. Hunninghake, Noriyuki Tomiyama, David C. Christiani, and Hiroto Hatabu. "Interstitial lung abnormalities at CT: subtypes, clinical significance, and associations with lung cancer." Radiographics 42, no. 7 (2022): 1925-1939.

[6] Khan, A., M. A. Alsahli, M. A. Aljasir, H. Maswadeh, M. A. Mobark, F. Azam, K. S. Allemailem, F. Alrumaihi, F. A. Alhumaydhi, and A. S. S. Alwashmi. "Safety, stability, and therapeutic efficacy of long-circulating TQ-incorporated liposomes: Implication in the treatment of lung cancer., 2022, 14, 153.

[7] Bidzińska, Joanna, and Edyta Szurowska. "See lung cancer with an AI." Cancers 15, no. 4 (2023): 1321.

[8] Lingling, Z. H. U., W. A. N. G. Ting, W. U. Juan, Z. H. A. I. Xiaoqian, W. U. Qiang, D. E. N. G. Hanyu, Q. I. N. Changlong, T. I. A. N. Long, and Z. H. O. U. Qinghua. "Updated Interpretation of the NCCN Clinical Practice Guidelines (Version 3. 2023) for Non-small Cell Lung Cancer." Chinese Journal of Lung Cancer 26, no. 6 (2023).

[9] Howlader, Nadia, Gonçalo Forjaz, Meghan J. Mooradian, Rafael Meza, Chung Yin Kong, Kathleen A. Cronin, Angela B. Mariotto, Douglas R. Lowy, and Eric J. Feuer. "The effect of advances in lung-cancer treatment on population mortality." New England Journal of Medicine 383, no. 7 (2020): 640-649.

[10] de Koning, Harry J., Carlijn M. van Der Aalst, Pim A. de Jong, Ernst T. Scholten, Kristiaan Nackaerts, Marjolein A. Heuvelmans, Jan-Willem J. Lammers et al. "Reduced lung-cancer mortality with volume CT screening in a randomized trial." New England journal of medicine 382, no. 6 (2020): 503-513.

[11] Becker, Nikolaus, Erna Motsch, Anke Trotter, Claus P. Heussel, Hendrik Dienemann, Philipp A. Schnabel, Hans-Ulrich Kauczor et al. "Lung cancer mortality reduction by LDCT screening—results from the randomized German LUSI trial." International journal of cancer 146, no. 6 (2020): 1503-1513.

[12] Thallam, Chinmayi, Aarsha Peruboyina, Sagi Sai Tejasvi Raju, and Nalini Sampath. "Early stage lung cancer prediction using various machine learning techniques." In 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1285-1292. IEEE, 2020.

[13] Xu, Yiwen, Ahmed Hosny, Roman Zeleznik, Chintan Parmar, Thibaud Coroller, Idalid Franco, Raymond H. Mak, and Hugo JWL Aerts. "Deep learning predicts lung cancer treatment response from serial medical imaging." Clinical Cancer Research 25, no. 11 (2019): 3266-3275.

[14] Tuncal, Kubra, Boran Sekeroglu, and Cagri Ozkan. "Lung cancer incidence prediction using machine learning algorithms." Journal of advances in information technology 11, no. 2 (2020).

[15] Banerjee, Nikita, and Subhalaxmi Das. "Prediction lung cancer–in machine learning perspective." In 2020 International conference on computer science, engineering and applications (ICCSEA), pp. 1-5.

IEEE, 2020.

[16] Warkentin, Matthew T., Stephen Lam, and Rayjean J. Hung. "Determinants of impaired lung function and lung cancer prediction among never-smokers in the UK Biobank cohort." EBioMedicine 47 (2019): 58-64.

[17] Delzell, Darcie AP, Sara Magnuson, Tabitha Peter, Michelle Smith, and Brian J. Smith. "Machine learning and feature selection methods for disease classification with application to lung cancer screening image data." Frontiers in oncology 9 (2019): 1393.

[18] Benzekry, Sébastien, Mathieu Grangeon, Mélanie Karlsen, Maria Alexa, Isabella Bicalho-Frazeto, Solène Chaleat, Pascale Tomasini, Dominique Barbolosi, Fabrice Barlesi, and Laurent Greillier. "Machine learning for prediction of immunotherapy efficacy in non-small cell lung cancer from simple clinical and biological data." Cancers 13, no. 24 (2021): 6210.

[19] Jayaraj, D., and S. Sathiamoorthy. "Random forest-based classification model for lung cancer prediction on computer tomography images." In 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 100-104. IEEE, 2019.

[20] Wang, Shidan, Donghan M. Yang, Ruichen Rong, Xiaowei Zhan, Junya Fujimoto, Hongyu Liu, John Minna, Ignacio Ivan Wistuba, Yang Xie, and Guanghua Xiao. "Artificial intelligence in lung cancer pathology image analysis." Cancers 11, no. 11 (2019): 1673.

[21] Singh, Gur Amrit Pal, and P. K. Gupta. "Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans." Neural Computing and Applications 31, no. 10 (2019): 6863-6877.

[22] Lakshmanaprabu, S. K., Sachi Nandan Mohanty, K. Shankar, N. Arunkumar, and Gustavo Ramirez. "Optimal deep learning model for classification of lung cancer on CT images." Future Generation Computer Systems 92 (2019): 374-382.

[23] Gu, Qianbiao, Zhichao Feng, Qi Liang, Meijiao Li, Jiao Deng, Mengtian Ma, Wei Wang, Jianbin Liu, Peng Liu, and Pengfei Rong. "Machine learning-based radiomics strategy for prediction of cell proliferation in non-small cell lung cancer." European journal of radiology 118 (2019): 32-37.

[24] Chandra, E. Yatish Venkata, K. Ravi Teja, M. H. C. S. Prasad, and B. Mohammed Ismail. "Lung cancer prediction using data mining techniques." International Journal of Recent Technology and Engineering (IJRTE) 8, no. 4 (2019).

[25] Wiesweg, M., F. Mairinger, H. Reis, M. Goetz, R. F. H. Walter, T. Hager, M. Metzenmacher et al. "Machine learning-based predictors for immune checkpoint inhibitor therapy of non-small-cell lung cancer." Annals of Oncology 30, no. 4 (2019): 655-657.