The Bioscan 19(3): 95-102, 2024 www.thebioscan.com
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES

Lung Cancer Prediction System based on Machine Learning Algorithms

M.Prabu

Department of Artificial Intelligence and Data Science, Aalim Muhammad Salegh College ofEngineering, Chennai – 600 055, India.

DOI: https://doi.org/10.63001/tbs.2024.v19.i03.pp95-102

KEYWORDS
lung cancer, AdaBoost, decision tree (DT), support vector machine (SVM), random forest classifier (RF).

Received on:
08-08-2024

Accepted on:
10-12-2024

ABSTRACT

Lung cancer is the most dangerous malignant tumour in terms of morbidity and mortality, and it poses a significant threat to human health. Recognizing and predicting lung cancer at the earliest possible stage can significantly enhance patient survival. Machine learning techniques can predict lung cancer early and effectively. We used a publicly accessible dataset from the Kaggle web repository and employed machine learning algorithms to predict lung cancer. After the pre- processing and normalization procedures on the dataset, the dataset is divided into training and testing subsets. To determine the optimum model for lung cancer prediction, this study employs four prominent classifiers, such as AdaBoost, Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF). In this study, experimental results show that the proposed machine learning algorithms achieve accuracy of 71.67%, 85.67%, 97%, and 100% in predicting three levels (low, medium, and high) of lung cancer. Random Forest classifier outperforms the other classifiers with the highest accuracy. The performance of classifiers is compared using parameters such as precision, recall, F1 score, and accuracy.

INTRODUCTION

Lung cancer develops within the lungs and is the main cause of cancer-related deaths in men and women. Non-small cell lung cancer (NSCLC) is the predominant type, comprising 80% of all instances of lung cancer. Small-cell lung cancer (SCLC) and mesothelioma are two other types. Smoking is the primary cause of lung cancer. However, other factors such as second-hand smoke, radon, asbestos, and certain chemicals can also increase the risk. Persistent cough, breathlessness, wheezing, chest discomfort, hoarseness, unintentional weight loss, fatigue, bone soreness, recurrent headaches, speech difficulty, and memory lapses are just a few of the symptoms. Anyone experiencing the symptoms listed above should seek medical attention.

In India, lung cancer emerges as the primary contributor to cancer-related mortality, constituting nearly a quarter of all deaths attributed to cancer. Lung cancer is becoming more common in the country, with an increasing occurrence. Notably, it has risen to become the most common cancer among Indian women. Smoking is the primary cause of lung cancer in India, responsible for a large number of cases. In India, approximately 40% of men and 10% of women smoke. Furthermore, over 30% of non-smokers in India are exposed to second-hand smoke, which is one of the major risk factors for lung cancer. Lung cancer is caused by the use of biomass fuels and exposure to certain chemicals increases the risk of lung cancer. In 2020, an estimated 70,275 lung cancer cases were diagnosed in INDIA. In developed countries the 5-year lung cancer survival rate is 20% and in INDIA its about 10%. Indian government has launched many initiatives to reduce the lung cancer issue in the country and the measures include: Raising awareness about the causes and potential risk factors, help the people who are attempting to quit smoking, allocating funds for the development of new treatments, Increasing the availability of high-quality cancer care in rural areas.

Early prediction of lung cancer can increase the survival rate. Surgical procedures, radiation therapy, chemotherapy, and targeted therapy are the lung cancer treatment options. Treatment depends on the stage and type of lung cancer. Surgery is the most common treatment in early stage and the primary goal of surgery is to remove the malignant tumour. Thoracoscopy is used to remove the tiny and localised tumour. If the malignancy grows in size, a thoracotomy, which involves a wider incision in the chest, may be required. Chemotherapy uses pharmaceutical medications to destroy cancer cells. Radiation therapy is used to treat advanced lung cancer. In radiation therapy, cancer cells are killed using high-energy beams. Targeted therapy in advanced lung cancer uses medications that specifically target molecules in cancer cells. Apart from these traditional treatments, a number of novel treatments for lung cancer are being developed. Furthermore, Immunotherapy uses the body's immune system to combat cancer and gene therapy uses genes to either repair damaged DNA or destroy cancer cells.

1. Related Work

Prediction models were proposed using hybridised genetic algorithms based on existing learning models to detect breast cancer tumours [1]. A machine learning framework was developed to check whether the water used for drinking purposes is safe [2]. Furthermore, a heterogeneous deep learning-based ensemble model was developed for effective breast cancer prediction. This model works in three stages: feature extraction, generation of stacked feature sets, and the extracted features fed to RF algorithm for prediction [3]. A method of calculating the film thickness of IC-deposited films employs the Auto SKlearn system to construct a prediction model and uses three main methodologies, such as meta-learning, Bayesian optimisation, and model integration [4]. Principal Component Analysis (PCA), Particle Swarm Optimisation (PSO), and Extreme Learning Machine (ELM) were adapted to improve the accuracy of dissolved oxygen prediction in river water [5]. Sentiment analysis was conducted using standard recurrent neural network (RNN) with gated recurrent units (GRU) and bidirectional gated recurrent units (BiGRU) [6]. A pneumonia detection model was

developed using a convolutional neural network (CNN). Pneumonia detection technique was based on X-ray images from pneumonia patients and healthy people [7]. XGBoost, Decision Tree, and RF algorithms were used to classify water quality indexes, and their performance was evaluated using the metrics such as accuracy, precision, recall, and the F-1 score [8]. ELM was recommended for predicting the risk of heart disease. The stack ensemble learning method employs a set of weak learners, including multi-layer perceptron classifiers, DT classifiers, SVM classifiers, and logistic regression [9]. Customers loss was predicted using XGBoost and RF algorithm [10].

Furthermore, SVM and DT algorithms were used to predict the symptoms of diabetes [11]. CNNs were employed to create efficient models to identify high-risk groups for lung cancer, allows for early identification and avoids long-term effects from intervention [12]. RF logistic regression, SVM, Adaboost classifier, and gradient boosting were employed to forecast mental health issues among employees in the workplace [13]. A RF classifier was employed to differentiate earthquake and non-earthquake vibrations [14]. Artificial neural network (ANN) techniques in medical research can enhance intelligent diagnosis by developing medicine prescription and maintaining the patient [15]. Alibaba Tianchi Lung Cancer Detection Competition dataset was used in Noisy U-Net (NU- Net) to detect the Long Nodule patients [16]. 3D Deep Convolutional Neural Network (3DDCNN) was used to create a unique computer-aided decision support system for lung nodule diagnosis [17]. The Tuna Swarm Algorithm with Deep Learning model was applied for colon and lung cancer detection [18]. Transferable texture Convolutional Neural Networks (CNN) employed to Long Nodule in CT scans [19]. A Deep learning model was proposed to detect the lung cancer type from CT images. DenseNet and adaboost used to classify malignant tumour and aggregate multiple classification [20]. SVM was used to identify the image as benign or malignant [21]. AlexNet was used to classify the tissues of lung and colon for lung cancer diagnosis in Histopathology Images [22]. An improved multidimensional region-based fully convolutional network-based automated decision support system for lung nodule identification and classification was presented [23]. Machine learning algorithms were employed across different fields for prediction. Many of the existing methods apply machine learning algorithms to CT scan images and Histopathology Images to identify the lung cancer. Furthermore, machine learning algorithms employed to predict the breast cancer. In this study, the classifiers such as AdaBoost, DT, SVM and RF used to determine the optimum model for lung cancer prediction. Dataset is fed as a CSV file to the machine Learning algorithms.

2. Methodology

A workflow diagram of the proposed study for lung cancer prediction is shown in Figure 1. Dataset from the Kaggle web repository is used in the proposed work. Furthermore, 21 parameters are considered as the cause of lung cancer in the dataset. Data pre-processing refers to preparatory steps and techniques in data science applied to raw data before it can be used for analysis and modelling. Data pre-processing is important since the quality of the data directly affects the efficiency of machine learning models. Data pre-processing is not a one-size-fits-all approach, and it differs based on the nature of the data, the research objectives, and the type of modelling. Pre- processing takes up the most time for data scientists to make sure the data is clean, consistent, and suitable for effective research and model building. Data cleaning, data transformation, data reduction, data normalisation, handling outliers, and feature engineering are the main objectives of data pre-processing.

Data cleaning is the process of identifying and dealing with missing data, duplicate values, and errors in a dataset. Data transformation converts data into analysis-ready formats. Data transformation involves scaling numerical features, encoding categorical variables, and applying mathematical modification to make the data suitable for modelling. Furthermore, data reduction reduces the data volume without compromising important information. Dimensionality reduction can be used to reduce duplicate and irrelevant features. Data normalisation ensures that all features are on a similar scale and that no feature dominates the other features during analysis. Rigorous statistical procedures can be adopted to remove the outliers. However, feature engineering creates new features from existing data to provide more insights and improve the performance of the model. From the dataset 30% data used for training and 70% data used for testing.

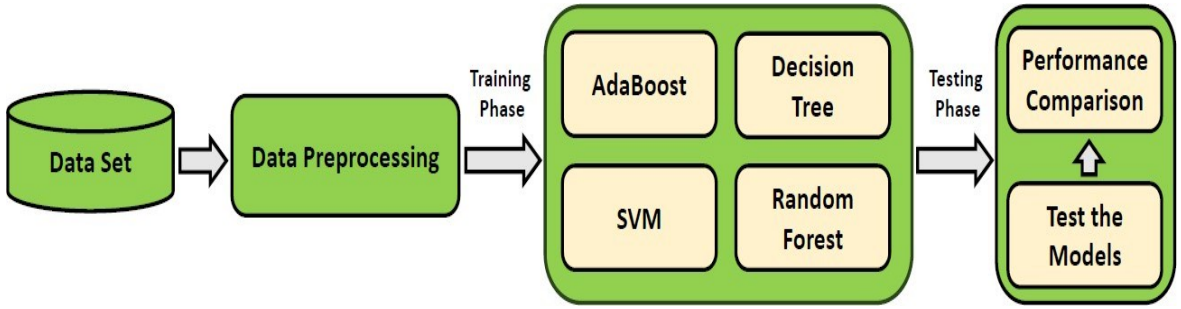


Fig.1. Workflow diagram of the proposed study for lung cancer prediction

The AdaBoost classifier belongs to the class of ensemble learning approaches that combine a number of ineffective learners to produce a strong learner. Weak learners merged to create a highly accurate model, even though they are less accurate individually. AdaBoost trains weak learners sequentially through an iterative technique. The approach applies weights to the training examples at each iteration, emphasising misclassified examples from the prior learner. Then, using this weighted dataset, the weak learner that follows is trained. Until the desired number of weak learners is reached, this technique is repeated. The weighted total of the predictions made by each weak learner yields the AdaBoost classifier's final prediction. The weights obtained during training are inversely proportionate to each learner's error. AdaBoost is a potent algorithm that excels in many classification tasks with high accuracy. Furthermore, AdaBoost is resistant to data noise and outliers. Hence, this study adapts the AdaBoost classifier. However, AdaBoost training on large datasets can be computationally demanding, just like random forest training. The choice and calibre of the

weak learners have a significant impact on AdaBoost's performance. Additionally, AdaBoost faces the danger of overfitting the training data if too many weak learners are used. AdaBoost is a strong and adaptable algorithm that may be used for a variety of classification applications. It performs well in situations with unbalanced data or datasets with few features. To get the best results, consideration should be given to computational expenses and the selection of suitable weak learners.

Both classification and regression applications can use decision tree classifiers. Decision Tree creates a model that resembles a tree to generate predictions by learning straightforward decision rules from the input features. Furthermore, data is divided into a labelled dataset, where each data point has a target label and a collection of input features that match. The algorithm starts creating the tree once the data is ready. The best characteristic from the dataset is first chosen by the algorithm to divide the data. Finding the feature that best divides the data into the various divisions is the objective. This is accomplished by analysing a number of factors such as entropy, information gain, and Gini impurity. The data is separated into subgroups based on the best feature's values. This procedure is continued recursively for each branch of the tree, with each subset becoming a branch, and so on until a halting condition is satisfied.

The halting condition could be when the tree has reached its maximum depth, when there aren't enough samples in a leaf node, or when every sample in a node is from the same class. A tree is used to generate predictions on fresh, unforeseen data after it has been constructed by following the decision rules from root to leaf node. The projected class for input data is subsequently assigned as the majority class in that leaf node. This study employs decision trees because they are straightforward to understand, manage both category and numerical data, and are resistant to outliers. However, DT is vulnerable to overfitting when the tree is too deep. Several methods, like pruning, ensemble methods like random forests or gradient boosting, or placing growth restrictions on the tree, are used to combat overfitting. It is crucial to employ strategies to avoid overfitting because it can occur.

Both classification and regression applications can use SVM classifiers. The main purpose of an SVM classifier is to find the best hyperplane to divide data points into distinct classes in a feature space, which is accomplished by increasing the margin between the two classes. The separation of each class's closest data points, or support vectors, from the hyperplane is known as margin. A data point is represented as a vector of features by SVM, which operates in a high-dimensional feature space. The performance of the SVM depends on the feature selection. The SVM discovers the hyperplane that can properly separate the data points of various classes. Multiple hyperplanes may accomplish this, but the SVM seeks to identify the one with the greatest margin. The separability of data is frequently imperfect in real-world situations. SVM employs a soft margin strategy in these situations, enabling some misclassifications to provide a greater overall separation, and the "C" parameter regulates this trade-off. A large C value will attempt to minimise misclassifications at the expense of a lower margin, while a small C value will allow for a larger margin with some misclassifications. SVM can effectively handle non-linearly separable data using the kernel method. The kernel function maps the input data to a higher-dimensional space with data points separated by hyperplanes.

Depending on the soft margin setting and the kernel being used, SVM optimises a convex quadratic objective function during the training phase to identify the best hyperplane that maximises the margin or reduces the classification error. Once trained, the SVM classifier can assign categories to novel data points based on which side of the hyperplane they reside on. In this study, SVM is used because it operates well with both linearly and non-linearly separable data and is effective in high-dimensional environments. It can also handle enormous feature sets. Furthermore, SVM needs to use a portion of the training data, the use of support vectors makes it memory-efficient. SVM training can be computationally demanding, particularly for large datasets. In SVM, choosing the proper kernel and tweaking the hyperparameters are difficult tasks. The model's interpretability may be constrained, especially when using non-linear kernels. The selection of features, kernel, and hyperparameters has a significant impact on the system's performance and generalizability.

A RF classifier can be applied to both classification and regression problems. RF classifier is one of the powerful and most preferred machine learning algorithms due to its versatility, precision and longevity. In RF classifiers training phase, number of decision trees are constructed and their predictions are integrated to produce the decision. Construction of RF algorithm involves the following steps: bootstrapped sampling, building decision trees, and voting or averaging. RF classifier begin by bootstrapping various subsets of the original data. Bootstrapped sampling entails picking data points at random from the training set and replacing them. Results in each subset having a different combination of samples. Furthermore, a decision tree is trained on the corresponding data for each of the bootstrapped subsets. However, during the construction of each tree, only a random subset of features is considered for splitting at each node. This process introduces randomness and diversity among the individual trees. After all of the decision trees have been built, they make predictions based on new data points. The final prediction for classification tasks is determined by a majority vote among the individual tree predictions. The final prediction for regression problems corresponds to the average of the individual predictions generated by the trees. This study uses the random forest method because of its characteristics such as low overfitting, high accuracy, adaptability, feature importance, parallelization, robustness, and suitability for different data types.

Multiple trees were combined to offer a reliable model. When compared to a single decision tree, the random forest technique is more accurate on complicated datasets. The random forest method can also be used with a wide range of data types, including categorical and numerical features. Understanding the value of each feature in the prediction process is made easier with the help of the Random Forest algorithm's feature importance score. Because of ensemble averaging, random forests are less susceptible to outliers and noisy data points. While random forest is powerful, its individual decision trees can be highly complex and less interpretable compared to simpler models like linear regression. It can be computationally expensive to train multiple decision trees, especially when there are many trees and a significant number of datasets. After Data pre- processing, the dataset is splitted into training and testing datasets. Advanced machine learning techniques such as AdaBoost classifiers, decision tree classifiers, SVM classifiers, and RF classifiers were used in this study to identify the best assessment model for lung cancer prediction. Metrics like accuracy, precision, recall, and F-measure are used to evaluate the performance of adapted classifiers.

3. Results and Discussion

Precision refers to the degree of accuracy exhibited by a classifier's positive predictions. Precision is defined as the ratio between true positive instances and the sum of

true and false positive instances. Precision quantifies the proportion of cases that the classifier designates as positive that are genuinely positive. An adeptly precise classifier will yield minimal instances of false positives, which correspond to cases incorrectly classified as positive. Mathematically, precision is calculated using the below formula:

$$Precision = \frac{True\ Positives}{(True\ Positives + False\ Positives)} \quad (1)$$

predicted a positive outcome even though the reality was negative. Recall is an indicator that evaluates the accuracy of a classifier's positive predictions. Recall is defined as the ratio between positive instances and the sum of true positives and false negatives. Recall captures the proportion of actual positive instances that the classifier successfully predicts. A classifier with increased recall used to have a minimum number of false negatives, which represent instances mistakenly labelled as negative. Mathematically, recall is calculated using the below formula:

$$Recall = \frac{True\ Positives}{(True\ Positives + False\ Negatives)} \quad (2)$$

When it's crucial to prevent false negatives, recall is a valuable parameter for classifier evaluation. A high-recall cancer classifier won't overlook any cancer cases. This is crucial since false-negative cancer instances can cause individuals to experience major health issues. However, recall is not the sole significant parameter for classifier evaluation. F-measure measures the accuracy of a model on a dataset. F-measure combines precision and recall, and it's the harmonic mean of precision and recall. Furthermore, F-measure is a more balanced metric than precision and recall. Mathematically, the F-measure is calculated using the below formula:

$$F - measure = \frac{(2 * Precision * Recall)}{(Recall + Precision)} \quad (3)$$

As we know, precision is the percentage of positive incidents that are actually expected to be positive. The proportion of positive incidents that actually occur and are as expected is known as recall. The F-measure has a range of 0 to 1, and higher values of the F-measure indicate better performance. Precision, recall, and F-measure are calculated to select the best model to predict lung cancer.

Table 1
Performance values obtained for AdaBoost classifier

AdaBoost Classifier			
Parameter	Class 1	Class 2	Class 3
Precision	1	0	0.57
Recall	1	0	1
F-measure	1	0	0.72

Performance values obtained for the AdaBoost classifier are listed in Table 1. Precision for classes 1, 2, and 3 is obtained as 1, 0, 0.57. Recall for classes 1, 2, and 3 is obtained as 1, 0, 1, and the F-measure is obtained as 1, 0, 0.72. AdaBoost classifiers perform better with classes 1 and 3. Furthermore, AdaBoost underperforms in class 2. Figure 2 shows the performance of the AdaBoost classifier based on parameters such as precision, recall, and F-Measure.

True positives are the number of cases that are both truly positive and predicted by the classifier to be positive. False positives are the number of instances in which the classifier incorrectly

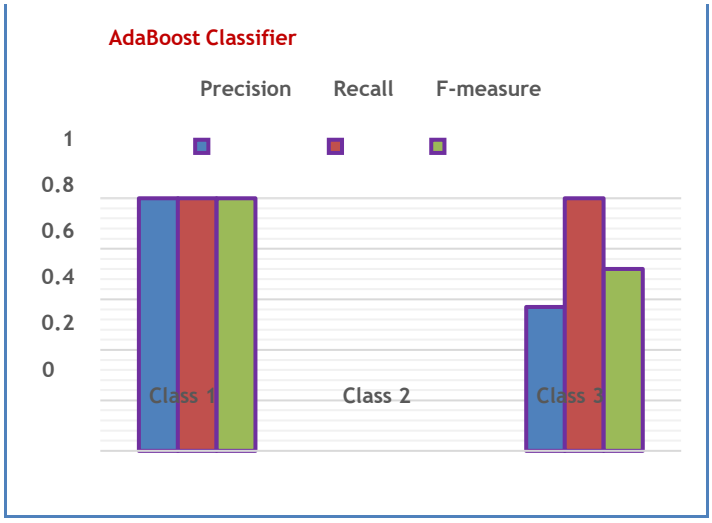


Fig. 2. Performance of AdaBoost classifier

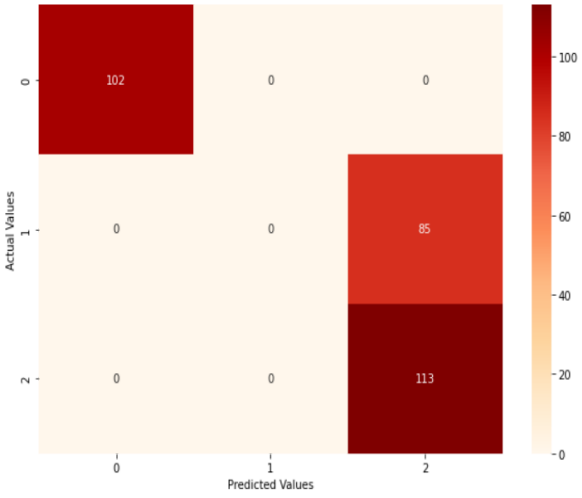


Fig. 3. Confusion Matrix of AdaBoost classifier

Confusion:

matrix is an organised table that shows the effectiveness of a classification model. It serves as a valuable tool for comprehending the model's proficiency in differentiating a class among multiple classes. The confusion matrix is a square table with two dimensions: the predicted class and the actual class. The table consists of nine quadrants, and each quadrant represents a different

prediction. The nine quadrants are: true positives for class 1 (TP_1), true negatives for class 1 (TN_1), false positives for class 1 (FP_1), false positives for class 2 (FP_2), true positives for class 2 (TP_2), false negatives for class 2 (FN_2), false positives for class 3 (FP_3), false negatives for class 3 (FN_3), and true positives for class 3 (TP_3). Array of metrics such as precision, recall, and accuracy can be calculated using confusion matrix that measure the model's efficiency. The confusion matrix for the AdaBoost classifier is shown in Figure 3. The performance values obtained for the decision tree classifier are listed in Table 2. Precisions for classes 1, 2, and 3 are obtained as 0.97, 1, and 0.95. Recall for classes 1, 2, and 3 is obtained as 1, 0.93, and 0.97, and the F-measure is obtained as 0.99, 0.96, and 0.96. Unlike the AdaBoost classifier, the DT classifier performs well with all the classes. Figure 4 shows the performance of the decision tree classifier based on parameters such as precision, recall, and F-measure. The confusion matrix of the Decision Tree classifier is shown in Figure 5.

Table 2
Performance values obtained for Decision Tree classifier

	Decision Tree Classifier		
Parameter	Class 1	Class 2	Class 3
Precision	0.97	1	0.95
Recall	1	0.93	0.97
F-measure	0.99	0.96	0.96

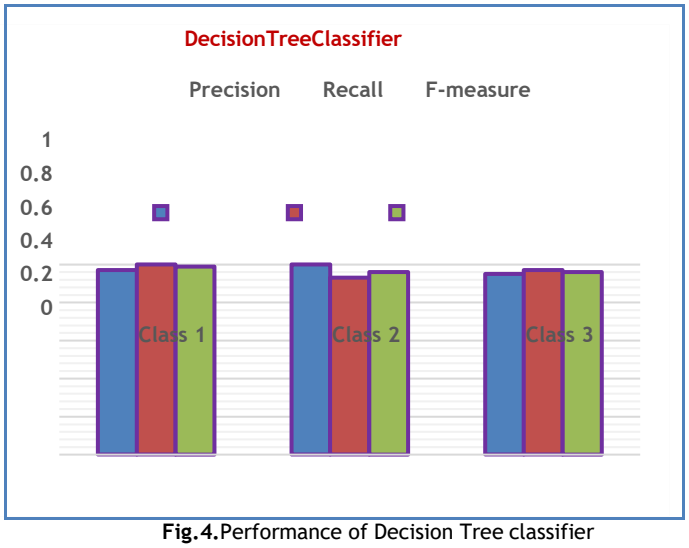


Fig.4. Performance of Decision Tree classifier

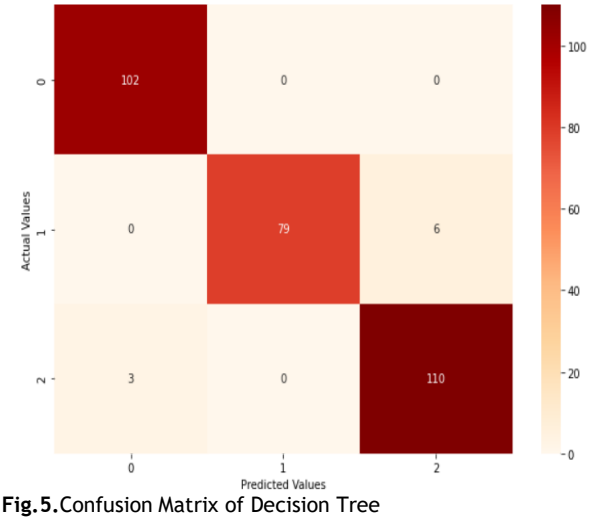


Fig.5. Confusion Matrix of Decision Tree classifier

Performance values obtained for the SVM classifier are listed in Table 3. Precisions for classes 1, 2, and 3 are obtained as 0.97, 1, and 0.95. Recall for classes 1, 2, and 3 is obtained as 1, 0.93, and 0.97, and the F-measure is obtained as 0.99, 0.96, and 0.96. The performance of the decision tree classifier and the SVM classifier is exactly similar. Figure 6 represents the performance of the SVM classifier based on parameters such as precision, recall, and F-measure. The confusion matrix of the SVM classifier is shown in Figure 7.

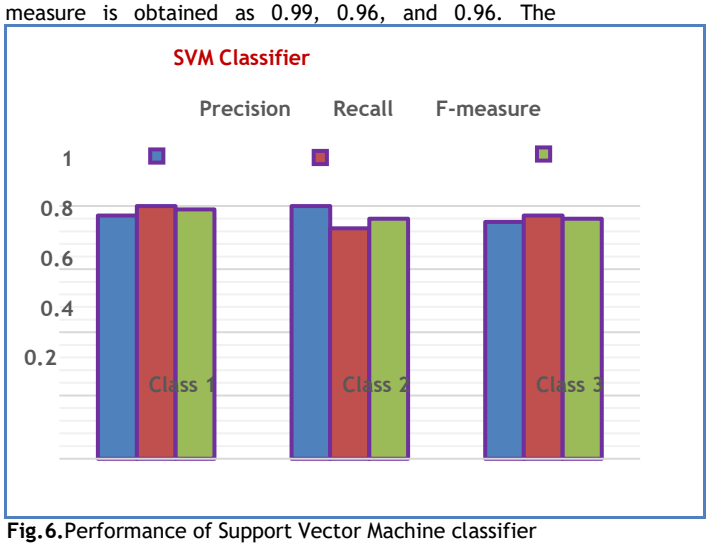


Fig.6. Performance of Support Vector Machine classifier

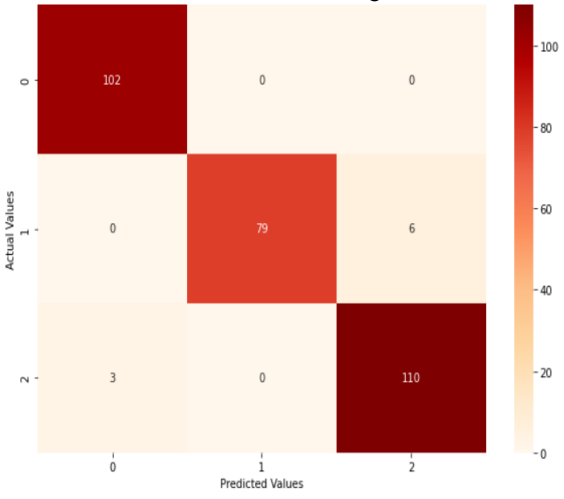


Fig.7. Confusion Matrix of Support Vector Machine classifier

Table 3

Performance values obtained for Support Vector Machine classifier

SVM Classifier			
Parameter	Class 1	Class 2	Class 3
Precision	0.97	1	0.95
Recall	1	0.93	0.97
F-measure	0.99	0.96	0.96

Table 4
Performance values obtained for Random Forest classifier

Random Forest Classifier			
Parameter	Class 1	Class 2	Class 3
Precision	1	1	1
Recall	1	1	1
F-measure	1	1	1

The performance values obtained for the Random Forest classifier are listed in Table 4. Precision for classes 1, 2, and 3 is obtained as 1. Recall and the F-Measure for classes 1, 2, and 3 are obtained as 1. A random forest classifier performs better than AdaBoost, Decision Tree, and SVM classifiers. Figure 8 represents the performance of the Random Forest classifier based on parameters such as precision, recall, and F-measure. The confusion matrix of the RF classifier is shown in Figure 9.

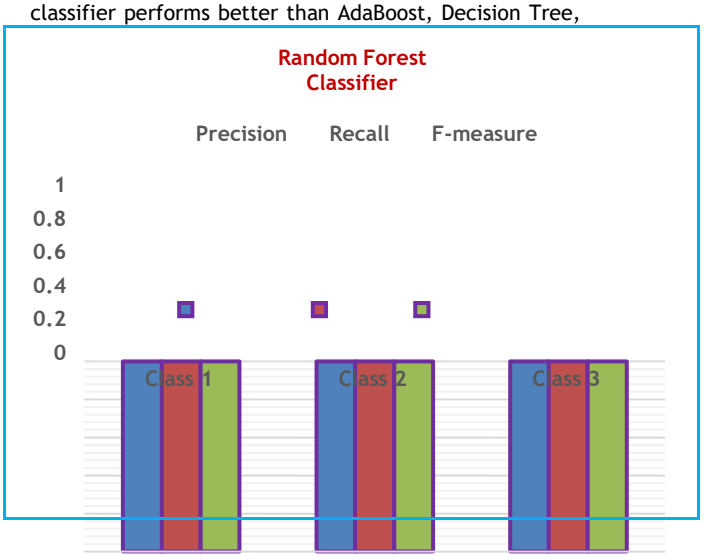


Fig.8. Performance of Random Forest classifier

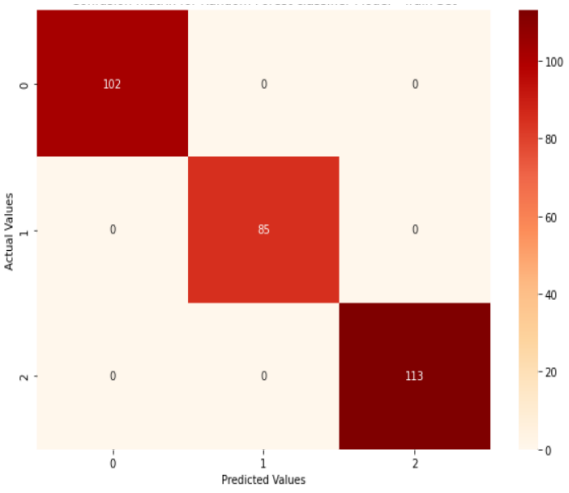
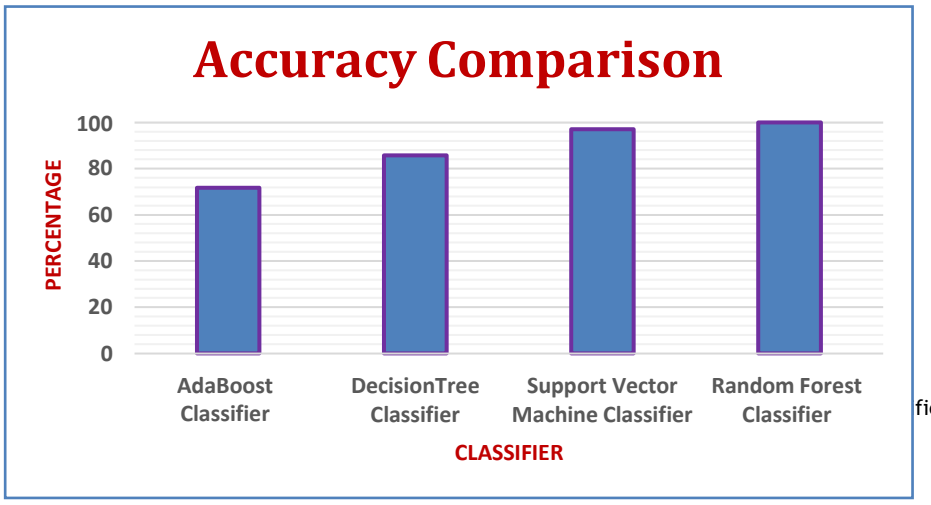


Fig.9. Confusion Matrix of Random Forest classifier

The accuracy of different classifiers is listed for evaluation in Table 5. The AdaBoost classifier performs less compared to other classifiers, with an accuracy of 71.67%. The decision tree classifier performs better than the AdaBoost classifier but less than the SVM and RF classifiers, with an accuracy of 85.67. SVM classifier outperforms AdaBoost and DT classifiers with an accuracy of 97%, and it performs less than compared to the RF classifier. The performance of the RF classifier surpasses all other classifiers with an accuracy of 100%. Figure 10 represents the performance of all the classifiers based on accuracy.

Table 5
Performance of different classifiers

Algorithm	Efficiency
AdaBoost Classifier	71.67
Decision Tree Classifier	85.67
SVM Classifier	97.00
RF Classifier	100



fier

CONCLUSION

Machine learning is used to examine large datasets in order to recognise patterns and trends within them. Machine learning has the ability to build predictive models for determining a person's susceptibility to lung cancer. These models consider personal characteristics such as age, smoking history, and family medical history. Early detection of cancer minimises the risk of complications and increases the probability of a cure. Various types of classifiers, such as AdaBoost, Decision Tree, SVM, and RF Classifier were adapted in this study to determine the best model for lung cancer prediction. The RF classifier outperformed all of the other classifiers. Algorithm's efficiency and performance can be increased by properly choosing the hyperparameters. In this study, RF classifier outperformed the other three machine learning algorithms with highest accuracy. Furthermore, the ensemble algorithm can be adapted to increase the ability of the system to predict lung cancer perfectly. Early detection can help to enhance the quality of cancer patient's life and that can be achieved using prediction models.

REFERENCES

- Ayoola, Joyce A., and Tokunbo Ogunfunmi. "A Comparative Analysis of Hybridized Genetic Algorithm in Predictive Models of Breast Cancer Tumors." IEEE Access (2023). https://doi.org/10.1109/GHTC55712.2022.9911033.
- Aleem, Isha, Attique Ur Rehman, Sabeen Javaid, and Tahir Muhammad Ali. "An Integrated Machine Learning Framework for Effective Classification of Water." In 2023 International Conference on Energy, Power, Environment, Control, and Computing (ICEPECC), pp. 1-6. IEEE, 2023. https://doi.org/10.1109/ICEPECC57281.2023.10209495.
- Jadoon, Ehtisham Khan, Fiaz Gul Khan, Sajid Shah, Ahmad Khan, and Muhammed Elaffendi. "Deep Learning-Based Multi-Modal Ensemble Classification Approach for Human Breast Cancer Prognosis." IEEE Access (2023). https://doi.org/10.1109/ACCESS.2023.3304242.
- Shi, Yumeng, Yu Cai, and Yining Chen. "A Method of Predicting the Deposited Film Thickness in IC Fabrication Based on Automatic Machine Learning." In 2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD), pp. 299-304. IEEE, 2023. https://doi.org/10.1109/ICAIBD57115.2023.10206117.
- Guo, Lijin, and Dawei Fu. "River Water Quality Prediction Model Based on PCA-APSO-ELM Neural Network." In 2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD), pp. 512-517. IEEE, 2023. https://doi.org/10.1109/ICAIBD57115.2023.10206249.
- Rudra, Rizwanul Islam, and Anilkumar Kothalil Gopalakrishnan. "Sentiment Analysis of Consumer Reviews Using Machine Learning Approach." In 2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 49-54. IEEE, 2023. https://doi.org/10.1109/JCSSE58229.2023.10202106.
- Zhang, Guotao, Xin Li, Dingfu Zhang, and Teoh Teik Toe. "Pneumonia Detection Method based on Convolutional Neural Network." In 2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD), pp. 716-721. IEEE, 2023. https://doi.org/10.1109/ICAIBD57115.2023.10206204.
- Ceesay, Rattana, Thapana Boonchoo, and Prapaporn Rattanatamrong. "Machine Learning Approaches for Quality Control in Pulp Packaging Manufacturers." In 2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 385-390. IEEE, 2023. https://doi.org/10.1109/JCSSE58229.2023.10202113.
- Shaikh, Yasmeen, V. K. Parvati, and S. R. Biradar. "Heart Disease Prediction using Ensemble Learning." In 2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC) https://doi.org/10.1109/ICAISC58445.2023.10200375.
- Adiputra, I. Nyoman Mahayasa, and Paweena Wanchai. "Customer Churn Prediction Using Weight Average Ensemble Machine Learning Model." In 2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 90-94. IEEE, 2023. https://doi.org/10.1109/JCSSE58229.2023.10202105.

- Sonawane, Chaitanya, Kalpesh Somwanshi, Raj Patil, and Roshani Raut. "Diabetic Prediction Using Machine Algorithm SVM and Decision Tree." In 2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC), pp. 1-7. IEEE, 2023. https://doi.org/10.1109/ICAISC58445.2023.10199539.
- Prabu, S., G. Prathibha, S. Meivizhi, K. Janani, and N. Dharani. "Artificial Intelligence Based Lung Cancer Prediction Using Convolutional Neural Network Algorithm." In 2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC), pp. 1- 7. IEEE, 2023. https://doi.org/10.1109/ICAISC58445.2023.10199641.
- Almaleh, Abdulaziz. "Machine Learning-Based Forecasting of Mental Health Issues Among Employees in the Workplace." In 2023 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), pp. 118-124. IEEE, 2023. https://doi.org/10.1109/IAICT59002.2023.10205620.
- Junior, Rio, Ary Murti, and Dien Rahmawati. "Implementation of Random Forest Classifier for Real-time Earthquake Detection System." In 2023 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), pp. 227-231. IEEE, 2023. https://doi.org/10.1109/IAICT59002.2023.10205761.
- Parveen, R., M. Nabi, F. A. Memon, S. Zaman, and M. Ali. "A review and survey of artificial neural network in medical science." Journal of Advanced Research in Computing and Applications 3, no. 1 (2016): 7-16.
- Huang, Wenkai, and Lingkai Hu. "Using a noisy U-net for detecting lung nodule candidates." IEE https://doi.org/10.1109/ACCESS.2019.2918224
- Masood, Anum, Po Yang, Bin Sheng, Huating Li, Ping Li, Jing Qin, Vitaveska Lanfranchi, Jinman Kim, and David Dagan Feng. "Cloud-based automated clinical decision support system for detection and diagnosis of lung cancer in chest CT." IEEE journal of translational engineering https://doi.org/10.1109/JTEHM.2019.2955458.
- Obayya, Marwa, Munya A. Arasi, Nuha Alruwais, Raed Alsini, Abdullah Mohamed, and Ishfaq Yaseen. "Biomedical image analysis for colon and lung cancer detection using tuna swarm algorithm with deep learning model." IEEE Access (2023). https://doi.org/10.1109/ACCESS.2023.3309711.
- Ali, Imdad, Muhammad Muzammil, Ihsan Ul Haq, Amir A. Khaliq, and Suheel Abdullah. "Efficient lung nodule classification using transferable texture convolutional neural network." Ieee https://doi.org/10.1109/ACCESS.2020.3026080.
- Pang, Shanchen, Yaqin Zhang, Mao Ding, Xun Wang, and Xianjin Xie. "A deep model for lung cancer type identification by densely connected convolutional networks and adaptive boosting." IEEE Access 8 (2019): 4799-4805. https://doi.org/10.1109/ACCESS.2019.2962862.
- Abbas, Waseem, Khan Bahadar Khan, Muhammad Aqeel, Muhammad Adeel Azam, Muhammad Hamza Ghouri, and Fawwad Hassan Jaskani. "Lungs nodule cancer detection using statistical techniques." In 2020 IEEE 23rd International Multitopic Conference (INMIC), pp. 1-6. IEEE, 2020.https://doi.org/10.1109/INMIC50486.2020.9318181.
- Peng, Chun-Cheng, and Jia-Wei Wu. "Deep Learning-Assisted Lung Cancer Diagnosis from Histopathology Images." In 2023 IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS), pp. 17-20. IEEE, 2023. https://doi.org/10.1109/ECBIOS57802.2023.10218594.
- Masood, Anum, Bin Sheng, Po Yang, Ping Li, Huating Li, Jinman Kim, and David Dagan Feng. "Automated decision support system for lung cancer detection and classification via enhanced RFCN with multilayer fusion RPN." IEEE Transactions on Industrial Informatics 16, no. 12 (2020): 7791-7801. https://doi.org/10.1109/TII.2020.2972918.

Access 7 (2019): 67905-67915.

in health and medicine 8 (20

Access 8 (2020): 175859-175870.