World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

Generalizable Ensemble Learning Models for Early Lung Cancer Detection

Shakil Khan [1], Mehedi Hasan [2], Mohammed Imam Hossain Tarek [2], Mostafizur Rahman Shakil [3], Md Fakrul islam Polash [4] and Istiak Kabir [5,*]

[1] *Department of Business Analytics, International American University, Los Angeles, CA 90010, USA.*
[2] *Department of Business Administration, International American University, Los Angeles, CA 90010, USA.*
[3] *Department of Engineering Management, Westcliff University, Irvine, CA 92614, USA.*
[4] *Department of Engineering Project Management, Westcliff University, Irvine, CA 92614, USA.*
[5] *Department of Computer Science, Westcliff University, Irvine, CA 92614, USA.*

World Journal of Advanced Engineering Technology and Sciences, 2025, 17(01), 200-217

Publication history: Received on 30 August 2025; revised on 04 October 2025; accepted on 07 October 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.17.1.1390

Abstract

Lung cancer means the birth of malignant cell inside the body which is out of control. A rising number of death rates in both genders has prompted researchers in the medical field to figure out ways on how it can be detected at an early stage for the purpose of mitigation, which also increases the patient's survival. Although there have been number of researches done using machine learning specially ensemble models, there still remains a gap on the research which is to have a comparative analysis done between the ensemble models such as Hybrid Majority Voting and Ensemble Stacking on tabular data. The objective of this study is to apply machine learning specially Ensemble models and compare their results on different datasets to identify the general pattern of the algorithms in this field and figure out if any particular method of ensemble performs better than the other in predicting lung cancer. Two Datasets were collected from public online sources and analysed to make sure it follows the distribution properly and there are no outliers. A pool of 9 Machine Learning algorithms with 50 hyper-parameter settings were studied to pick the best 3 Machine Learning models. After that a number of ensemble techniques were applied such as Majority Hard Voting, Weighted Hard Voting, Soft Voting, and Ensemble Stacking and their performance were analysed. Different Evaluation metrics such as Accuracy, F1-Score, ROC-AUC Score, Average Precision and Confusion Matrices were applied which highlighted the superior performance of the Ensemble Models. Particularly, Weighted Ensemble Learning Model for Dataset 1 achieved 89.04% Accuracy and F1-Score and Ensemble Stacking for Dataset 2 achieved 87.95% Accuracy and F1-Score, which indicates the superior effectiveness and generalizability of the ensemble models.

Keywords: Lung Cancer; Tabular Data; Machine Learning; Classification; Confusion Matrix; Heat Map

1. Introduction

Among the cancers, Lung Cancer is considered one the deadliest types of cancer in the world surpassing even Colon, Uterus and Skin Cancer as well [1], contributing significantly to the annual number of cancer-related deaths. This respiratory disease has increased enormously over the last few decades due to the polluted atmosphere. United Nations in 2015 has declared 17 Sustainable Development goals (SDGs) and among them Target 3.9 is focuses on reducing the number of death due to hazardous chemical pollution in air, water and soil [2]. The prognosis of cancer varies widely depending on the type, stage and overall health of the patient [3]. Some initial symptoms of Lung Cancer include shortness of breath and back pain. The back pain is caused when pressure is exerted on the tumour itself. The tumour can also cause issues by spreading to different parts of the body [4]. Additional symptoms including persistent coughs, spit out blood when coughing, chest pain when breathing or laughing, weakness, fatigue, reduced appetite and weight and also recurring respiratory infections like pneumonia and bronchitis [5].

* Corresponding author: Istiak Kabir

Traditional diagnostic methods such as imaging and biopsies, often detect this disease at a pretty later stage which limits treatment options and reduces survival rates. Recent studies [6], [7], [8] in deep learning have enabled the early detection and automated analysis of clinical data. Particularly using tabular data such as Demographics, Clinical Parameters, Lifestyle & Environment, Symptoms & Comorbidities, and Treatment Data [9]. Usage of these kind of data has enabled the deployment of variety of machine learning for predictive modelling of lung cancer diagnosis. Although machine learning has shown promising performance in cancer prediction, balancing high predictive performance with interpretability remains a significant challenge. Many advanced machine learning models offer limited transparency, often being criticized as "black boxes," which hampers clinical acceptance where understanding the rationale behind predictions is vital for trust and informed decision-making [10], [11].

To improve predictive accuracy in lung cancer detection, ensemble learning strategies have emerged as effective solutions. These methods combine multiple machine learning models to leverage their collective strengths, employing approaches such as Majority Hard Voting, Majority Soft Voting, and Ensemble Stacking [12]. By integrating the outputs of diverse models, ensemble learning aims to improve overall prediction robustness and accuracy [13], [14]. Among the ensemble techniques Majority Hard Voting, Majority Soft Voting and Ensemble Stacking are particularly noteworthy [15]. Majority Hard Voting takes predictions from all of the classifiers and selects the class with the majority vote [16]. Majority Soft Voting sums all the classifiers prediction probabilities to determine the final prediction. Ensemble Stacking involves training a meta-classifier to combine to predictions of the base classifiers leveraging the strength of each to improve performance.

A thorough assessment of these two ensemble approaches for lung cancer prediction from tabular data is absent. This study aims to address the following critical research questions: (i) Which individual machine learning model provides better performance for the early prediction of lung cancer from tabular data of affected individuals? (ii) Which majority voting classifier provides better performance for the early prediction of lung cancer from tabular data of affected individuals? (iii) What are the techniques that can be applied to improve the performance of the model? This study investigates the implementation and comparative evaluation of Hybrid Majority Voting and Ensemble Stacking. To achieve this, two lung cancer datasets were collected from Kaggle, a detailed data preprocessing such as handling missing values, data scaling, data encoding, and handling outliers were done. To the best of our knowledge, these two datasets are completely new and no prior study has systematically examined the performance of Hybrid Majority Voting—which includes Majority Hard Voting, Majority Soft Voting, and Weighted Hard Voting—alongside Ensemble Stacking within a unified experimental framework for predicting lung cancer in early stage. The study employed nine distinct machine learning algorithms with 50 different hyperparameter configurations spanning nine fundamentally different algorithms to identify the most effective base classifiers. This extensive exploration serves two primary purposes: Firstly, it systematically identifies strong standalone classifiers by providing a wide range of complexity performance trade-offs, secondly by casting a broad across decision trees, gradient based methods, instance-based learners and neural architecture, we minimize the risk of overlooking a potentially superior model which is crucial of ensemble models' success as it maximizes the heterogeneity among base classifiers. Existing literature has not explored such an extensive hyperparameter tuning process, nor have these optimized models been consistently integrated into ensemble architectures for lung cancer prediction. This paper makes the following three key contributions:

- A detailed comparative analysis of advanced ensemble techniques tailored to lung cancer prediction from structured tabular data.
- A rigorous evaluation of model selection and optimization procedures prior to ensemble construction.
- This research aims to highlight the comparative strengths and limitations of individual and ensemble models, while assessing their practical effectiveness in enhancing predictive accuracy and reliability in lung cancer diagnostics.

2. Related Works

Recently biomedical research has improved therapeutic outcomes. Novel approaches in precision wound healing leverage regenerative therapies integrated with smart technologies to enhance patient-specific recovery strategies [17]. Similarly, hybrid temozolomide nanoconjugates demonstrate how polymer–drug designs can improve drug stability and efficacy in glioblastoma therapy [18], offering a promising route for overcoming conventional treatment limitations. Complementing these developments, the concept of molecular erasers introduces protein degradation as a means to reprogram cancer immunity, opening new avenues for immuno-oncology [19]. In precision farming, AI enhances efficiency and productivity through automatic navigation and self-driving technologies in agricultural machinery [20]. In computer vision, models with attention mechanisms have improved road segmentation for autonomous driving [21]. In healthcare, hybrid CNN-SVM models on enhanced MRI data show promise in accurately classifying Alzheimer's disease [22]. Additionally, multimodal deep learning frameworks such as MultiSenseNet help predict machine failure

risks in industrial settings [23]. In plant pathology, deep stacking models combining CNNs with gradient boosting have improved detection of leaf diseases [24]. Overall, these advancements illustrate the growing effectiveness of AI in addressing complex real-world challenges across key sectors.

Dama et al. [25] proposed an explainable machine learning framework aimed at predicting lung cancer in relation to the impact of pesticide exposure. Using three datasets – the Thai Dataset (680 instances), Lung cancer Prediction Dataset (1000 Samples), and a Survey Lung cancer Dataset (309 Patients) – the study focused on numerical attributes such as pesticide exposure, duration, age, occupation, symptoms like coughing blood, smoking history. Various machine learning models such as XGBoost, Logistic Regression, Decision Tree, Gaussian Naïve Bayes, Multi-Layer Perceptron, Random Fores, Support Vector Machine (SVM) was used that delivered superior performance. Two Stage Feature Selection methods such as Extra Tree Classifier and PCA for Dimensionality Reduction were applied. XGBoost with SMOTE combined with ENN (Edited Nearest Neighbour) with PCA Preprocessing performed the best results, achieving 99% accuracy in all of the datasets. Despite its high performance the study acknowledged its limitations such as overfitting risks due to high-dimensional data. Future directions suggested validating models on larger multi-omics data and developing real-time risk monitoring tools, investigating deep learning approaches for further accuracy enhancement.

Asuntha et al. [26] conducted an experiment to study the comparative performance of various machine learning algorithms for lung cancer prediction. The study used public datasets from Kaggle Comprising of 310 samples capturing data such as patient's habits (e.g. smoking, alcohol consumption) and symptoms (e.g. yellow fingers, chest pain, anxiety). The features also included demographic details alongside clinical symptoms. Multiple machine learning models such as Logistic Regression, Gaussian Naïve Bayes, Bernoulli Naïve Bayes, SVM, Random Forest, K-Nearest Neighbours (KNN), Extreme Gradient Boosting (XGB), Extra Trees, AdaBoost and ensemble methods combining XGB and AdaBoost, as well as a Multi-Layer Perceptron (MLP). The data preprocessing steps involved removing duplicate entries, splitting the data into 80:20 ratio for training and testing and performing 10-fold cross validation. KNN Achieved the highest accuracy achieving an accuracy of 92.86% followed closely by Bernoulli Naïve Bayes and Gaussian Naïve Bayes. Persons Correlation heat maps were also used for feature correlation analysis. Some limitations of the study were the usage of a very small dataset, and using only one ensemble technique and not exploring the other techniques such as bagging, stacking etc.

Al-Huseiny et al. [27] developed a machine learning framework using XGBoost for the classification of Benign Malignant pulmonary nodules based on low-dose spiral computerized tomography (LDCT) and clinical data from Screening. The study utilized two datasets, the primary dataset was collected from physical examination data from the Institute of Health management of the General Hospital of the Chinese People's Liberation Army (PLA) which container 1335,503 participants and an external validation dataset was collected from the Henan Provincial People's Medical Health Examination Centre and Sichuan Provincial People's Hospital which contained 5.146 participants. Both demographic data and Clinical Data were collected for the experiment. The top 15 clinical features were selected via XGBoost feature importance ranking. Processing involved encoding the categorical Data's, Splitting the data in 80:20 ration for training and testing, hyperparameter tuning through grid search and automatic handling of missing data using XGBoost. The model achieved an AUC of 0.76 and an accuracy of 0.75 in internal validation, and an improved AUC of 0.87 an accuracy of 0.80 in external validation. Although achieving good results, some limitations of the study included using limited number of data samples which limited generalizability and also, usage of only one type of Ensemble technique (The other ensemble technique could potentially improve performance of the models).

Reddy et al. [28] conducted a study based on nomogram-based and machine learning-based methods for survival prediction of non-small cell lung cancer (NSCLC) patients. The authors used a dataset that contained 6,586 patients from the Cancer Hospital Affiliated to Chongqing University (CUCH), China. The dataset included clinical, pathological, demographic, and treatment-related features such as age, sex, weight, smoking history, Tumour Staging, and treatment modalities. Several Machine Learning Algorithm was implemented such as Logistic Regression, Random Forest, XGBoost, Decision Tree, and Light Gradient Boosting Machine. Feature selection was done using pairwise Spearman's rank correlation and the Boruta Method. Among the models Random Forest performed the best across multiple points suggesting it's reliability for long-term survival prediction. Limitations of the study included lack of external validation and potential improvement to model's using different ensemble techniques. Rule-based knowledge could also be integrated to enhance the predictions of the machine learning models.

Manapov et al. [29] developed machine learning models using publicly available Kaggle dataset containing 309 patients for Lung Cancer Prediction. The dataset contained variety of demographic and clinical feature such as coughing, shortness of breath, chest pain and other health indicators such as chronic diseases and allergy history. A wide range of machine learning models were explored such as Naïve Bayes (NB), Bayesian Network, Stochastic Gradient Descent,

Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Random Tree (RT), Reduced Error Pruning Tree (RepTree), Rotation Forest (RotF), and AdaBoostM1. Feature selections were done using Gain Ratio and Random Forest rankings. SMOTE was also applied for class balancing in the data. Rotation Forest achieved the best accuracy of 97.1% with an AUC of 99.3%. Some limitations of the study noted by the authors were the non-generalizability of the models, and non-clinical features of the dataset. Future work also included using deep learning techniques such as LSTM and CNN.

Debnath et al. [30] proposed a novel multivariate boosting classification method name Multivariate Ruzicka Regressed eXtreme Gradient Boosting Data Classification (MRRXGBDC). The authors used four datasets: The Lung cancer dataset (1000 instances), the Thoracic Surgery dataset (470 instances), the Air Quality Lung cancer dataset (2602 countries), and the Lung 16 lung cancer dataset (551, 065 instances). All of the datasets were publicly available. The dataset contained features such as ID, age, gender, air pollution exposure, alcohol use, dust allergy, and occupational hazards. One of the unique features of the proposed model was its ability to handle missing values and handle imbalanced data. Limitations identified for this study includes only exploring only one form of Ensemble technique, reducing computational complexity and improving generalizability of the disease prediction system.

Ogundana et al. [31] conducted the machine learning based study on the investigation of gender-specific prognosis in lung cancer patients for the Surveillance, Epidemiology and SEER database comprising of 28,458 cases. A range of machine learning models such as Naïve Bayes, Decision Trees, Random Forest, Extreme Gradient Boosting, K-Nearest Neighbour, Logistic Regression, and Support Vector Machine with Random Forest achieving the best overall performance. For one-year survival XGB model achieved 90.75% accuracy, for three-year survival prediction LR achieved the best accuracy of 75.65% accuracy, and for Five-year Survival prediction LR again achieved the best accuracy getting about 71.79% accuracy. Major limitations of the study was the restricted scope of clinical features in the SEER database, major features were absent such as smoking and drinking history. Also, the comparison the other ensemble methods such as hard bagging, stacking and soft voting were not explored in this study.

Authors [32] explored machine learning approaches for predicting early death among lung cancer patients with bone metastases. Using the SEER database, the study utilized demographic and clinical characteristics of the patients to be used as predictive features. Logistic Regression, Extreme Gradient Boosting, Random Forest, Neural Network, Gradient Boosting Machine were applied for comparative studies. Gradient Boosting Machine performed the best among the other machine learning models. To interpret the results. Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive exPlanations (SHAP) were employed. Some of the limitations of this study were lack of certain clinical predictors such as the quantity and specific location of the bone metastases, and the absence of external validation sets.

Shah et al. [33] investigated deep learning-based approaches for predicting lung cancer survival periods using the SEER database. The dataset included both categorical and quantitative variables such age, number of tumours, chemotherapy status, and AJCC TNM Staging factors. Several machine learning and deep learning models were applied, such as Artificial Neural, Convolutional Neural Network, Recurrent Neural Network, Random Forest, Support Vector Machine, Naïve Byes, Gradient Boosting Machines and Linear Regression for comparative analysis. A custom ensemble was also developed, which was based on stacking. Feature selection was done using LASSO regression. ANN performed the best for classification task with an accuracy of 71.18% and CNN was the most effective for regressions tasks achieving 13.50% RMSE and 50.66% R^2. Some limitations of the study were the data included imbalanced dataset and higher error rates for predicting longer survival periods.

3. Methodology

3.1. Experimental Setup

The Experimental setup for this study was conducted in a Local Machine which used Python 3.11.4 as a runtime environment. The Local Machine used Ryzen 1600 with 6 Core and 12 Threads CPU, 16GB of Ram and an RTX 3060 12GB of GPU which was used for efficient training of the models. Pandas, Seaborn and Scikit Learn were used for Data Preprocessing and Splitting, scaling and model creation. Imbalanced Learn was used for Over Sampling of the data. Matplotlib was used to visualize the training curves, evaluate the models and create confusion matrices. This setup allowed for efficient and effective experimentation with no compute unit constraints.

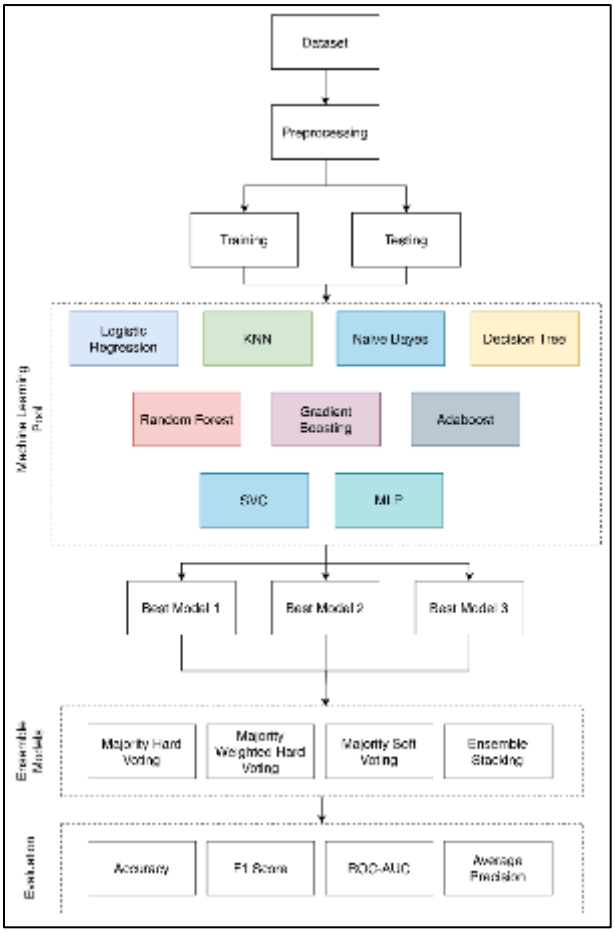

Figure 1 Workflow Diagram

The proposed study methodology for lung cancer prediction pipeline is illustrated in

Figure 1.The following section outlines the methodology of the study which includes the steps taken for data collection, processing and evaluation, creation of machine learning pool for selecting the best machine learning models, constructing ensemble models from the best machine learning model and evaluation. Detailed explanation of the steps of the approaches are as follows:

3.2. Dataset Preparation and Analysis

The datasets were collected from Kaggle [17], [18]. A description of the dataset is given in Figure 1. Dataset 1 consisted of 5000 Patient records with 18 features whereas Dataset 2 consists of 20000 records with 16 attributes. Both the dataset includes demographic information such as medical history, lifestyles and symptoms associated with pulmonary disease. The dataset before processing was analysed statistically for null values and if it contained any outliers. Pearson Correlation heatmap was also plotted which is provided in Figure 2 to assess the importance of different attributes among each other. The left side of the figure refers to the dataset 1 and right side to the dataset 2. For dataset 1 strong correlations are observed between stress immune and immune weakness, as well as, mental stress and stress immune. For dataset 2, the dataset focuses more on diagnosis-specific features such as yellow fingers, chronic disease etc. The correlations can be seen as weak which allows the machine learning model to learn from the data independently and evaluate feature contributions automatically.

Table 1 Dataset Description

Dataset Name	Attributes	Values	Statistics						
			Count	Mean	Std	Min	50%	75%	Max

Dataset 1 (lung_cancer_dataset)	AGE	Numeric Value	5000	57.22	0.50	30	57	71	84
	GENDER	1 [Female], 0 [Male]		0.50	0.50	0	1	1	1
	SMOKING	1 [Yes], 0 [No]		0.66	0.47	0	1	1	1
	FINGER_DISCOLORATION	1 [Yes], 0 [No]		0.60	0.48	0	1	1	1
	MENTAL_STRESS	1 [Yes], 0 [No]		0.53	0.49	0	1	1	1
	EXPOSURE_TO_POLLUTION	1 [Yes], 0 [No]		0.51	0.499	0	1	1	1
	LONG_TERM_ILLNESS	1 [Yes], 0 [No]		0.43	0.49	0	0	1	1
	ENERGY_LEVEL	Numeric Value		55.03	7.91	23.25	55.05	60.32	83.04
	IMMUNE_WEAKNESS	1 [Yes], 0 [No]		0.39	0.48	0	0	1	1
	BREATHING_ISSUE	1 [Yes], 0 [No]		0.80	0.39	0	1	1	1
	ALCOHOL_CONSUMPTION	1 [Yes], 0 [No]		0.35	0.47	0	0	1	1
	THROAT_DISCOMFORT	1 [Yes], 0 [No]		0.69	0.45	0	1	1	1
	OXYGEN_SATURATION	Numeric Value		94.99	1.48	89.92	94.97	95.98	99.79
	CHEST_TIGHTNESS	1 [Yes], 0 [No]		0.60	0.48	0	1	1	1
	FAMILY_HISTORY	1 [Yes], 0 [No]		0.30	0.45	0	0	1	1
	SMOKING_FAMILY_HISTORY	1 [Yes], 0 [No]		0.20	0.40	0	0	0	1
	STRESS_IMMUNE	1 [Yes], 0 [No]		0.20	0.40	0	0	0	1
	PULMONARY_DISEASE	1 [Yes], 0 [No]		Null	Null	Null	Null	Null	Null
Dataset 2 (lcs_synthetic_20000)	GENDER	M [Male], F [Female]	20000	Null	Null	Null	Null	Null	Null
	AGE	Numeric Value		62.20	8.2	30.00	62.00	68.00	87.00
	SMOKING	2 [Yes], 1 [No]		1.56	0.49	1.00	2	2	2
	YELLOW_FINGERS	2 [Yes], 1 [No]		1.57	0.49	1	2	2	2

	ANXIETY	2 [Yes], 1 [No]		1.53	0.49	1	2	2	2
	PEER_PRESSURE	2 [Yes], 1 [No]		1.50	0.5	1	2	2	2
	CHRONIC DISEASE	2 [Yes], 1 [No]		1.50	0.49	1	2	2	2
	FATIGUE	2 [Yes], 1 [No]		1.67	0.47	1	2	2	2
	ALLERGY	2 [Yes], 1 [No]		1.56	0.50	1	2	2	2
	WHEEZING	2 [Yes], 1 [No]		1.55	0.49	1	2	2	2
	ALCOHOL CONSUMING	2 [Yes], 1 [No]		1.55	0.49	1	2	2	2
	COUGHING	2 [Yes], 1 [No]		1.57	0.49	1	2	2	2
	SHORTNESS OF BREATH	2 [Yes], 1 [No]		1.64	0.47	1	2	2	2
	SWALLOWING DIFFICULTY	2 [Yes], 1 [No]		1.47	0.49	1	1	2	2
	CHEST PAIN	2 [Yes], 1 [No]		1.55	0.49	1	2	2	2
	LUNG_CANCER	YES, NO		Null	Null	Null	Null	Null	Null

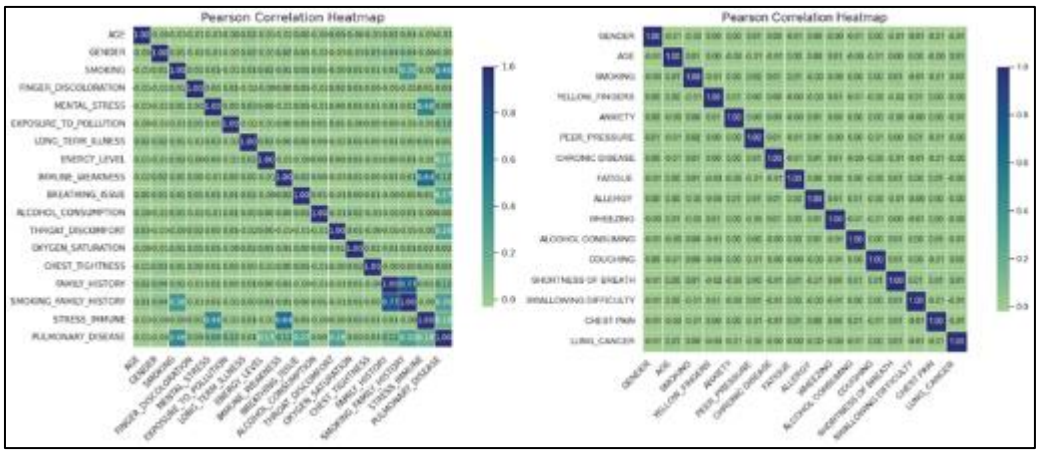

Figure 2 Pearson Correlation Heatmap

Preprocessing is crucial step in any machine learning experiment. It prepares the data to be fed inside the machine learning model [34]. A class-based count was measured to check class imbalance which is shown in Figure 3. As there was imbalance between the classes which could potentially make the machine learning model biased. Synthetic Minority Oversampling (SMOTE) was applied to balance the classes. The data was split into two sets for training (80%) and testing (20%). A standard scaling was applied to normalize the data [35], [36]. It transforms the data in such a way that the data has a mean of zero and a standard deviation of 1.

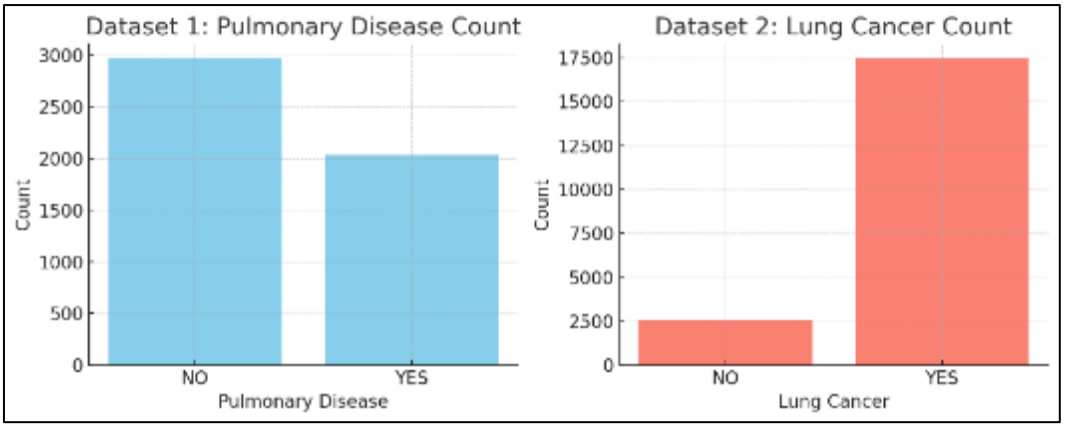

Figure 3 Class Distribution of Datasets

3.3. Machine Learning Pool

After preprocessing a comprehensive set of machine learning classifiers were selected for the detection of Lung Cancer from the tabular dataset. A pool of 9 machine learning models were created with 50 different hyperparameters settings by fine tuning. The hyperparameters of the models are given in Table . The hyperparameters were selected based on a discrete search space of the nine candidate algorithms covering regularization strengths and for logistic regression, estimator counts and feature sampling strategies for Random Forest and boosting methods. These choices were informed by standard practice in the literature and preliminary experiments. A grid-styled sweep was performed over the combination of these parameters on the training data to guard against overfitting and ensure balanced performance. This exhaustive tuning of the models helped identify the strongest base learners and created the performance diversity necessary for a robust ensemble creation [37].

Logistic Regression: Logistic Regression supervised machine learning model which is used widely binary classification problems. It uses the logistic (Sigmoid) function to map the linear combination of input features into probability values ranging from 0 to 1 [38]. The probability indicates the likelihood of that given input corresponds to one of the predefined classes. In this study Logistic Regression is used due to its simplicity, interpretability and effectiveness.

Table 2 Machine Learning Pool Description

Model ID	Model Type	Parameters
LR1	Logistic Regression	penalty=l2, C=0.75, solver=liblinear, multi_class=ovr
LR2	Logistic Regression	penalty=None
LR3	Logistic Regression	penalty=l2, C=0.75, solver=lbfgs
LR4	Logistic Regression	penalty=l2, C=0.5
LR5	Logistic Regression	penalty=l2, C=1.0
LR6	Logistic Regression	penalty=l2, C=0.8
KNN	KNN	n_neighbors=2, algorithm=kd_tree, metric=manhattan
3NN	KNN	n_neighbors=3
5NN	KNN	n_neighbors=5
6NN	KNN	n_neighbors=6, algorithm=brute, p=1, metric=cosine
9NN	KNN	n_neighbors=9, algorithm=ball_tree, metric=euclidean
10NN	KNN	n_neighbors=10, algorithm=brute, p=2, metric=minkowski
NB1	Naive Bayes	GaussianNB

NB2	Naive Bayes	BernoulliNB
DT1	Decision Tree	criterion=gini, min_samples_leaf=4
DT2	Decision Tree	criterion=gini, max_features=sqrt, min_samples_leaf=2, min_impurity_decrease=0.01
DT3	Decision Tree	criterion=gini, max_features=log2, min_samples_leaf=2, min_impurity_decrease=0.01
DT4	Decision Tree	criterion=entropy, splitter=random, max_depth=6, min_samples_leaf=2
DT5	Decision Tree	criterion=log_loss, max_depth=8, min_samples_leaf=3, min_impurity_decrease=0.001
RF1	Random Forest	criterion=entropy, n_estimators=100
RF2	Random Forest	n_estimators=500
RF3	Random Forest	n_estimators=1000
RF4	Random Forest	criterion=log_loss, n_estimators=500
RF5	Random Forest	criterion=gini,max_features=log2, n_estimators=500
RF6	Random Forest	criterion=gini, max_features=sqrt, n_estimators=1000, bootstrap=False
RF7	Random Forest	criterion=gini, max_features=log2, n_estimators=500, bootstrap=False
RF8	Random Forest	criterion=entropy, n_estimators=500, bootstrap=False
GB1	Gradient Boosting	n_estimators=500
GB2	Gradient Boosting	loss=log_loss, n_estimators=1000, criterion=squared_error
GB3	Gradient Boosting	loss=log_loss, n_estimators=400, criterion=friedman_mse
GB4	Gradient Boosting	n_estimators=1100, criterion=squared_error, min_impurity_decrease=0.01
GB5	Gradient Boosting	loss=log_loss, n_estimators=500, criterion=friedman_mse, min_impurity_decrease=0.001
GB6	Gradient Boosting	loss=log_loss, n_estimators=800, min_samples_leaf=3, criterion=squared_error
GB7	Gradient Boosting	loss=log_loss, n_estimators=1200, min_impurity_decrease=0.01
GB8	Gradient Boosting	n_estimators=500, min_samples_split=5, min_impurity_decrease=0.0001
AdaBoost1	AdaBoost	default
AdaBoost2	AdaBoost	n_estimators=200, algorithm=SAMME
AdaBoost3	AdaBoost	n_estimators=500, algorithm=SAMME, learning_rate=2
AdaBoost4	AdaBoost	n_estimators=1000, algorithm=SAMME
AdaBoost5	AdaBoost	n_estimators=1200
SVC1	SVC	probability=True
SVC2	SVC	probability=True, C=0.5, kernel=linear
SVC3	SVC	probability=True, C=0.75, kernel=linear
SVC4	SVC	probability=True, C=1.0, kernel=linear
SVC5	SVC	probability=True, C=1.25, kernel=linear
MLP1	MLP	max_iter=500
MLP2	MLP	max_iter=200
MLP3	MLP	learning_rate=adaptive, max_iter=500

MLP4	MLP	learning_rate=invscaling, max_iter=1000
MLP5	MLP	hidden_layer_sizes=(100,100), max_iter=1000

K-Nearest Neighbour: It is a non-parametric supervised machine learning model. The main methodology behind this classifier is that it classifies an object based on the plurality vote by its neighbour. The neighbours to be considered or k, and the distant metric to be used in the voting is considered as a hyperparameter [39]. KNN was introduced in this study for its robustness and its classification performance which it does without making any assumptions about the data distribution.

Naïve Bayes: This is a simple classifier that assigns probabilities to different labels from the vector of feature values. It is based on the Bayes Theorem. It calculates the posterior probability of each class based on the input features and selects the class with the highest probability [40]. Despite, its complexity it often performs well in high dimensional dataset. This model was added to the study to analyse the benchmark of probabilistic modelling.

Support Vector Machine: Otherwise known as SVM is a supervised max-margin model that can be used for both classification and regression analysis. It tries the find the optimal hyperplane that separates the different classes by maximizing the margin of difference between them. It can handle both linear and non-linear input features. For non-linear input features it uses different kernel functions to convert the high dimensional data to lower dimensional data [41]. In this study SVM was used to capture complex decision boundaries and enhance classification accuracy.

Decision Tree: It is a recursive model that splits the data intro a structured tree based on the input features. In DT each branch represents the outcome of the test and the leaf nodes represent the class labels. The whole decision tree represents the classification rules. The models were used in this study to explore feature importance.

Random Forest: This is an ensemble of weak learners grouped together. Usually, decision trees are used as the weak learners. By introducing randomness in its feature selection, it creates variations in the individual trees. The model was chosen for its robustness in classification tasks.

Gradient Boosting Machine (GBM): It builds an ensemble of weak learners in a sequential manner where each model's output is fed to another model. The task of each model is to minimize the error of the previous model. It optimizes its loss using gradient descent to lead to a highly accurate model [42]. GBM was utilized in this study for its high capability of capturing complex patterns in the data.

AdaBoost: Another form of Ensemble Learning that combines multiple machine weak classifiers to form a strong classifier. It assigns higher weights to misclassified instances in successive iterations, which allows the model to focus on harder case. In this study AdaBoost was employed to examine the performance in boosting weak learners for classification.

Multi-Layer Perceptron: This is a feedforward neural network which consists of fully connected layers using nonlinear activation functions and backpropagation for learning. MLP has the ability to model complex relationships and it is widely used in deep learning approaches.

3.4. Ensemble Model Construction

After evaluating the models from the machine learning pool, three best machine learning models were selected. Using these top models, several ensemble techniques were constructed.

- **Majority Hard Voting:** In majority hard voting, the final prediction is based on the most frequent prediction among the models.
- **Majority Soft Voting:** It uses probability estimates from all the models and averages them to get the final probability.
- **Weighted Hard Voting:** Models are assigned weights proportional to their individual performance.
- **Stacking Ensemble:** Prediction of the models are used as input for another machine learning model.

4. Result Analysis

The Accuracy and F1 score of the 50 Machine Learning models are given in Table for both Dataset 1 and Dataset 2. For Dataset 1 RF7, MLP2 and GB8 were the best 3 models having accuracy of 0.89, 0.86, and 0.86 respectively and for Dataset

2 RF6, MLP5, and KNN produced the best results having an accuracy of 0.84, 0.78 and 0.77. These models demonstrate that Random Forest and Multi-Layer Perceptron consistently performs well in both the datasets, highlighting the model's robustness in lung cancer classification tasks.

Table 3 Evaluation of Machine Learning Pool

Model ID	Dataset 1		Dataset 2	
	Accuracy	F1	Accuracy	F1
LR1	0.845148	0.849126	0.61285	0.616319
LR2	0.84557	0.849425	0.612814	0.61627
LR3	0.845148	0.849126	0.612814	0.61627
LR4	0.845148	0.849126	0.612814	0.61627
LR5	0.845148	0.849054	0.612814	0.61627
LR6	0.844937	0.848887	0.612814	0.61627
KNN	0.784177	0.764544	0.77093	0.765034
3NN	0.816667	0.822323	0.751425	0.775123
5NN	0.821941	0.828337	0.74239	0.769882
6NN	0.83903	0.835358	0.739271	0.748684
9NN	0.829114	0.835404	0.726399	0.755462
10NN	0.845992	0.844402	0.725145	0.740095
NB1	0.822785	0.825935	0.613818	0.618514
NB2	0.822152	0.828467	0.612276	0.616453
DT1	0.81962	0.816134	0.736259	0.74012
DT2	0.718565	0.765519	0.499839	0.33327
DT3	0.718565	0.765519	0.499839	0.33327
DT4	0.823207	0.81399	0.499839	0.33327
DT5	0.855907	0.851557	0.581765	0.589635
RF1	0.8827	0.880105	0.832813	0.837466
RF2	0.879536	0.876935	0.834283	0.838806
RF3	0.879958	0.877382	0.833566	0.838062
RF4	0.882068	0.879572	0.833244	0.838032
RF5	0.879536	0.876935	0.834283	0.838806
RF6	0.88481	0.88294	0.836506	0.838439
RF7	0.886076	0.884351	0.836435	0.838386
RF8	0.822785	0.823382	0.806747	0.819663
GB1	0.858017	0.857035	0.649564	0.658597
GB2	0.852321	0.851531	0.659711	0.669583
GB3	0.859705	0.858773	0.647987	0.656373
GB4	0.778059	0.793244	0.500018	0.666683
GB5	0.855696	0.854825	0.650389	0.659524

GB6	0.844093	0.842322	0.597218	0.609395
GB7	0.778059	0.793244	0.500018	0.666683
GB8	0.863502	0.861716	0.630168	0.637603
AdaBoost1	0.827426	0.836456	0.61493	0.616302
AdaBoost2	0.827426	0.836237	0.616077	0.619523
AdaBoost3	0.697046	0.689755	0.542397	0.537384
AdaBoost4	0.834388	0.84217	0.615969	0.618537
AdaBoost5	0.832278	0.840148	0.613854	0.616122
SVC1	0.861814	0.860198	0.68108	0.694091
SVC2	0.844304	0.845031	0.613424	0.622187
SVC3	0.844304	0.84497	0.613352	0.622116
SVC4	0.844726	0.845385	0.613316	0.622015
SVC5	0.844937	0.845565	0.613245	0.621969
MLP1	0.85	0.84861	0.676849	0.679861
MLP2	0.864557	0.862641	0.676455	0.683033
MLP3	0.85	0.84861	0.676849	0.679861
MLP4	0.839241	0.837976	0.676849	0.679861
MLP5	0.83038	0.832261	0.780252	0.791777

Table 4 Model Comparison on Dataset 1

Model	**Accuracy**	**F1 Score**	**ROC AUC**	**Average Precision**
Gradient Boosting	0.860	0.860	0.860	0.815
MLP	0.861	0.861	0.861	0.813
Random Forest	0.890	0.890	0.890	0.847
Ensemble Hard	0.869	0.869	0.869	0.826
Ensemble Weighted	0.890	0.890	0.890	0.847
Ensemble Soft	0.866	0.866	0.866	0.819
Stacking (MLP)	0.889	0.889	0.889	0.846

Table presents the Accuracy, F1 scores, ROC-AUC Score and Average Precision of the top performing machine learning models compared with their ensemble counterparts that is built using the top models from the machine learning pool on Dataset 1. The Weighted Ensemble Model gives better performance than the two individual models like MLP and Gradient Boosting in all of the given metrics. It shows the performance having the ROC_AUC, Accuracy and F1 Score of 89.04%, which is followed closely by Random Forest which itself is an Ensemble method, and it also achieved the same ROC_AUC, Accuracy and F1 Score. The ensemble stacking also exhibits the potential of achieving similar performance like the weighted ensemble models for lung cancer prediction on this dataset 1.

The above plot at Figure 4 illustrates the precision-recall curves for different machine learning models evaluated on Dataset 1. The curve highlights how each model performs on different thresholds. Random Forest and Ensemble Weighted Voting exhibit nearly identical curves, which is characterized by a high precision that is well maintained across a broad range of recall values with only a slight declined toward the end-indicating these models maintain strong predictive confidence even as they capture more true positives. Among all the classifiers, Random Forest and Ensemble Weighted Voting achieved the highest average precision score of 0.8471, which indicates strong performance in identifying positive instances. Ensemble Stacking and Ensemble Hard Voting followed closely with an average precision

of 0.825 and 0.818 respectively which highlights the strong generalization and the robustness in improving classification accuracies across various data distributions.

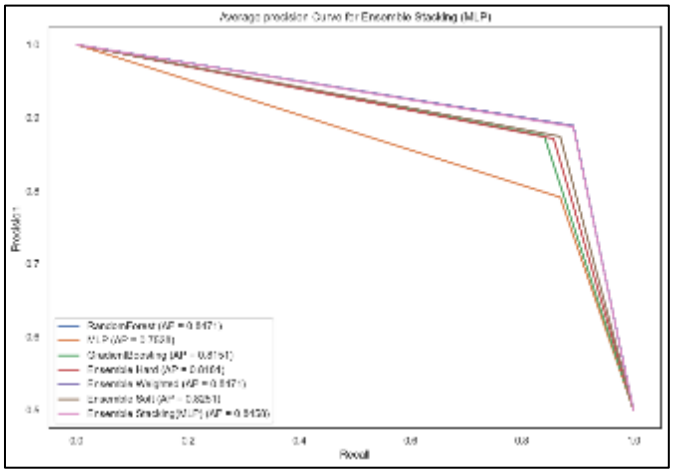

Figure 4 Precision Recall Curve for Dataset 1

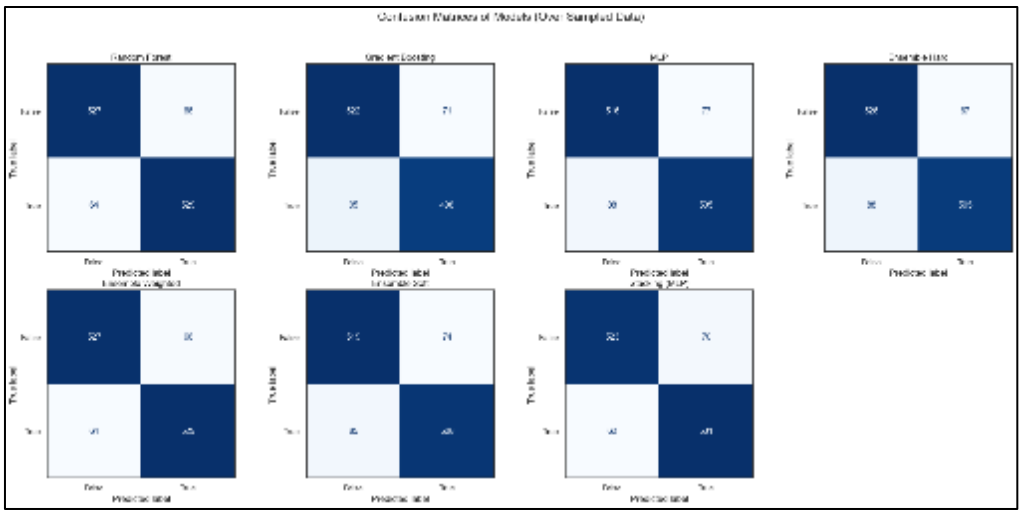

Figure 5 Confusion Matrix for models trained on Dataset 1

Figure 5 illustrates the confusion matrices for the best performing models compared with the Ensemble Techniques on Dataset 1. Random Forest and Weighted Ensemble Model shows strong diagonal dominance, indicating a high rate of correct classifications, with very few false positives and false negatives.

Table 5 Model Comparison on Dataset 2

Model	Accuracy	F1 Score	ROC AUC	Average Precision
Random Forest	0.855	0.854	0.855	0.801
MLP	0.786	0.784	0.786	0.712
KNN	0.783	0.783	0.783	0.725
Ensemble Hard	0.841	0.841	0.841	0.784
Ensemble Weighted	0.855	0.854	0.855	0.801
Ensemble Soft	0.827	0.827	0.827	0.758
Ensemble Stacking (MLP)	0.879	0.878	0.879	0.865

Table shows the Accuracy, F1-Score, ROC-AUC Score and Average Precision for the different machine learning models evaluated on Dataset 2. Similar to Dataset 1, the ensemble models, particularly Weighted Ensemble and Stacking demonstrated improved performance that the other machine learning models achieving 85.5% and 87.9% accuracy, F1-Score and 80.1% and 86.5% average precision respectively. This consistent pattern across different datasets suggests strong robustness capabilities of ensemble learning when combining strong, diverse models.

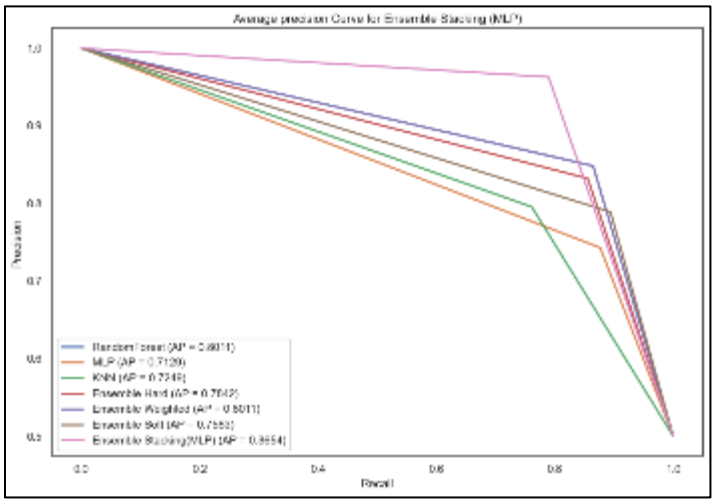

Figure 6 Precision Recall Curve for Dataset 2

The precision-recall curve at Figure 6 reveals the comparative analysis between several classifiers in terms of average precision (AP) and ROC-AUC. Among the models, Ensemble Stacking (MLP) model stands out with the most dominant curve, maintaining very high precision across the entire recall range before dropping sharply near full recall. This shape suggest that the models is highly confident in its predictions until it begins to retrieve nearly all positive cases, at which point precision starts to decline.

Ensemble Stacking (MLP) achieved the highest average precision of 0.865, showing the models superior capability in capturing complex patterns in the data. Both random forest and weighted ensemble method followed closely with an average precision of 0.801 and 0.801 respectively showing strong consistent results. Among the machine learning models K-Nearest Neighbour (KNN) and MLP achieved an average precision of 0.713 and 0.725, which reflects moderate effectiveness. These findings demonstrate that ensemble-based strategies particularly Ensemble Stacking offered notable improvements in average precision for Dataset 2.

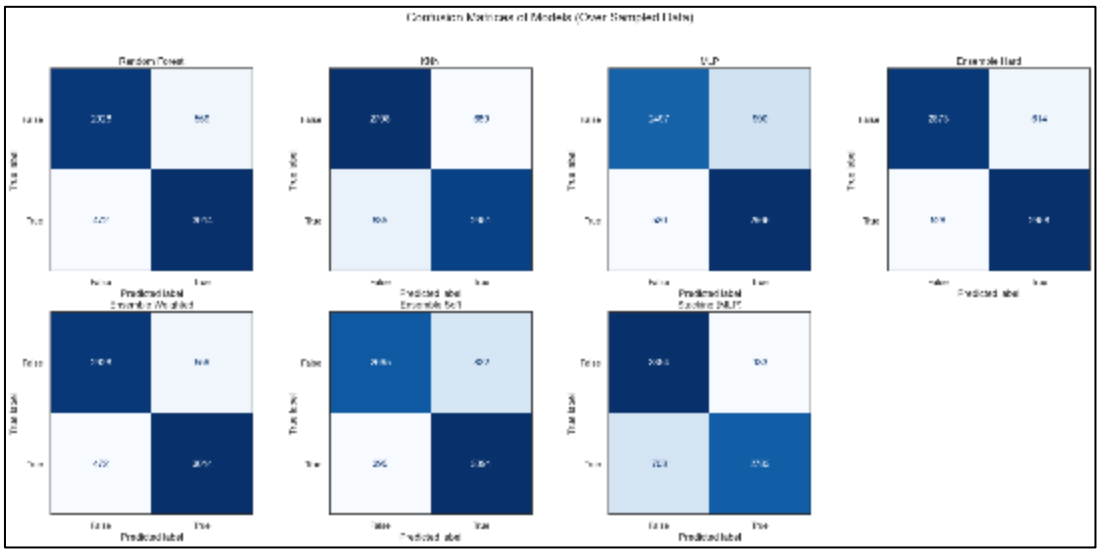

Figure 7 Confusion Matrix for models trained on Dataset 2

Figure 7 displays the confusion matrices for the top models on Dataset 2. Random Forest and Ensemble methods maintained high true positive and true negative rates which is evident in their strong diagonal entries. Compared to Dataset 1, Dataset 2's confusion matrices showed slightly better specificity and recall, suggesting the models trained on Dataset 2 are more reliable at correctly identifying both the positive and negative cases. Across both datasets, ensemble methods built using the top performing models consistently performed better results compared to the individual models. On Dataset 1, RF7, MLP2, and GB8 were used to create the ensemble model. Meanwhile, on Dataset 2 RF6, MLP5 and KNN were used to construct the Ensemble model. Ensemble Techniques such as Weighted Hard Voting and Stacking Ensemble approaches showed meaningful improvements in both accuracy and F1-Scores. The confusion matrices also showed the models True Positives and True Negatives improving. These findings confirm that carefully selecting the best performing models to be combined into an ensemble model substantially improves the predictive performance of Ensemble models.

5. Discussion

In this study we have demonstrated that ensemble learning methods—particularly weighted hard voting and ensemble stacking consistently outperformed individual base learners in the task of lung cancer prediction from tabular clinical data. On Dataset 1, the weighted voting ensemble achieved an accuracy, F1-Score, and ROC-AUC of 89.04%, matching the performance of the best single Random Forest model. On Dataset 2, ensemble stacking with multilayer perceptron achieved 87.9% accuracy and F1-Score with an average precision of 86.5%, which surpassed not only the individual models such as Random Forest (85.5%) and MLP (78.0%) but also soft and hard voting classifiers. These findings directly address our primary research questions by confirming that (1) while Random Forest and MLP are strong predictors, (2) ensemble strategies such as weighted ensemble and meta-learning models provide statistically meaningful gains in both discriminative power and threshold-independent metrics such as ROC-AUC and Average Precision.

Our precision-recall analysis from Figure 4 and Figure 6 revealed that Weighted Ensemble Voting and Ensemble Stacking maintained high precision across a broad range of recall values, indicating robust confidence as models capture more true positives in highly imbalanced contexts. Confusion matrices at Figure 5 and Figure 7 show that these ensemble models reduce both false-positives and false-negatives relative to the individual classifiers, which is more critical in clinical settings as the cost of misdiagnosis is very high. Together, the results show that combining diverse algorithms via hybrid weighted majority voting and ensemble stacking harness complementary decision boundaries which mitigates overfitting and variance inherent to single models.

Importantly, For RQ3, it is directly answered by our methodological choices where for Data Preprocessing methods such as handling missing values, SMOTE oversampling after splitting the data which removes data leakages, and standard scaling ensured clean, well distributed inputs for the machine learning models. For all the classifiers extensive hyperparameter optimization across nine algorithms and fifty parameter configurations allowed us to select the best parameter settings for each of the machine learning classifiers. Feature level analysis such as Pearsons Correlation Heatmap allowed us to remove weak predictive attributes, reducing noise and overfitting. The exploration of the Ensemble Paradigms demonstrated Weighted Ensemble Approach and Ensemble Stacking yielded the greatest gains which was further confirmed by metrics such as Accuracy, F1-Score and threshold independent metrics such as ROC-AUC and average precision.

6. Conclusion

This study presents a comprehensive analysis on machine learning and ensemble learning techniques for lung cancer prediction using structured tabular data. With different techniques applied for preprocessing such as Label Encoding, Over Sampling to balance the dataset. A pool of 50 machine learning models with different hyper parameter settings were created and among them top performing models such as Random Forest, Multi-Layer Perceptron, KNN were selected to construct different Ensemble Models such as Majority Hard Voting, Weighted Hard Voting, Soft Voting, and Ensemble Stacking. Experimental results show that the ensemble models specially Ensemble Stacking and Ensemble Weighted Hard Voting performed consistently across different datasets and outperforming the machine learning models in terms of accuracy, F1-Score, ROC-AUC curve and Average Precision score, and. Notably, Weighted Ensemble and Stacking approaches showed superior generalization capability and an analysis confirms lower false positives and false negatives. The findings support the integration of Ensemble Learning in clinical decision support systems for lung cancer prediction as it leverages the strengths of diverse base learners. Despite the promising results there are some potentials threats to validity. Firstly, both the datasets were collected from Kaggle which doesn't reflect real world scenarios that include noise and diversity. Secondly, the class imbalance was addressed using SMOTE which may

introduce bias and artifacts. Lastly, no external dataset was used for validation which may affect generalizability of the finding to other populations and settings. For future work, incorporating additional real-world datasets with diverse temporal features, and imaging data may help enhance the predictive accuracy. Moreover, exploring deep learning ensemble techniques and explainable AI methods can also help improve both performance and interpretability, making models more practical for clinical deployment.

Compliance with ethical standards

Disclosure of conflict of interest

There is no conflict of interest.

References

[1] N. Banerjee and S. Das, "Prediction Lung Cancer- in Machine Learning Perspective," 2020 International Conference on Computer Science, Engineering and Applications, ICCSEA 2020, Mar. 2020, doi: 10.1109/ICCSEA49143.2020.9132913.

[2] R. Javed, T. Abbas, A. H. Khan, A. Daud, A. Bukhari, and R. Alharbey, "Deep learning for lungs cancer detection: a review," Artificial Intelligence Review 2024 57:8, vol. 57, no. 8, pp. 1–39, Jul. 2024, doi: 10.1007/S10462-024-10807-1.

[3] Md. R. Ahmed et al., "Hierarchical Swin Transformer Ensemble with Explainable AI for Robust and Decentralized Breast Cancer Diagnosis," Bioengineering 2025, Vol. 12, Page 651, vol. 12, no. 6, p. 651, Jun. 2025, doi: 10.3390/BIOENGINEERING12060651.

[4] R. Haque et al., "Explainable deep stacking ensemble model for accurate and transparent brain tumor diagnosis," Comput Biol Med, vol. 191, p. 110166, Jun. 2025, doi: 10.1016/J.COMPBIOMED.2025.110166.

[5] M. Liu et al., "The value of artificial intelligence in the diagnosis of lung cancer: A systematic review and meta-analysis," PLoS One, vol. 18, no. 3, p. e0273445, Mar. 2023, doi: 10.1371/JOURNAL.PONE.0273445.

[6] M. I. H. Siddiqui et al., "Accelerated and accurate cervical cancer diagnosis using a novel stacking ensemble method with explainable AI," Inform Med Unlocked, vol. 56, p. 101657, Jan. 2025, doi: 10.1016/J.IMU.2025.101657.

[7] R. Haque et al., "A transfer learning-based computer-aided lung cancer detection system in smart healthcare," IET Conference Proceedings, vol. 2024, no. 37, pp. 594–601, Mar. 2025, doi: 10.1049/ICP.2025.0858.

[8] R. Haque et al., "Advancing Early Leukemia Diagnostics: A Comprehensive Study Incorporating Image Processing and Transfer Learning," BioMedInformatics 2024, Vol. 4, Pages 966-991, vol. 4, no. 2, pp. 966–991, Apr. 2024, doi: 10.3390/BIOMEDINFORMATICS4020054.

[9] F. H. Tang et al., "Recent advancements in lung cancer research: a narrative review," Transl Lung Cancer Res, vol. 14, no. 3, pp. 975–990, Mar. 2025, doi: 10.21037/TLCR-24-979/COIF.

[10] A. S. U. K. Pranta et al., "A Novel MaxViT Model for Accelerated and Precise Soybean Leaf and Seed Disease Identification," Computers 2025, Vol. 14, Page 197, vol. 14, no. 5, p. 197, May 2025, doi: 10.3390/COMPUTERS14050197.

[11] A. Al Noman et al., "ViX-MangoEFormer: An Enhanced Vision Transformer–EfficientFormer and Stacking Ensemble Approach for Mango Leaf Disease Recognition with Explainable Artificial Intelligence," Computers 2025, Vol. 14, Page 171, vol. 14, no. 5, p. 171, May 2025, doi: 10.3390/COMPUTERS14050171.

[12] R. Haque, S. H. Laskar, K. G. Khushbu, M. J. Hasan, and J. Uddin, "Data-Driven Solution to Identify Sentiments from Online Drug Reviews," Computers 2023, Vol. 12, Page 87, vol. 12, no. 4, p. 87, Apr. 2023, doi: 10.3390/COMPUTERS12040087.

[13] H. Guo, K. Xu, G. Duan, L. Wen, and Y. He, "Progress and future prospective of FDG-PET/CT imaging combined with optimized procedures in lung cancer: toward precision medicine," Ann Nucl Med, vol. 36, no. 1, pp. 1–14, Jan. 2022, doi: 10.1007/S12149-021-01683-8/METRICS.

[14] R. Haque et al., "Data-Centric Approach for Leather Quality Control Using Advanced Vision Transformer Models," Proceedings - International Conference on Next Generation Communication and Information Processing, INCIP 2025, pp. 200–205, 2025, doi: 10.1109/INCIP64058.2025.11019741.

[15] M. R. Ahmed, R. Haque, S. M. A. Rahman, A. W. Reza, N. Siddique, and H. Wang, "Vision-audio multimodal object recognition using hybrid and tensor fusion techniques," Information Fusion, vol. 126, p. 103667, Feb. 2026, doi: 10.1016/J.INFFUS.2025.103667.

[16] J. Hasan et al., "Transforming Leukemia Classification: A Comprehensive Study on Deep Learning Models for Enhanced Diagnostic Accuracy," PEEIACON 2024 - International Conference on Power, Electrical, Electronics and Industrial Applications, pp. 266–271, 2024, doi: 10.1109/PEEIACON63629.2024.10800693.

[17] A. H. Malik and S. Rahman, "Toward precision wound healing: Integrating regenerative therapies and smart technologies," International Journal of Science and Research Archive, vol. 16, no. 3, pp. 244–257, Sep. 2025, doi: 10.30574/IJSRA.2025.16.3.2492.

[18] A. H. Malik and S. Rahman, "Hybrid Temozolomide Nanoconjugates: A polymer–drug strategy for enhanced stability and glioblastoma therapy," International Journal of Science and Research Archive, vol. 16, no. 3, pp. 258–268, Sep. 2025, doi: 10.30574/IJSRA.2025.16.3.2493.

[19] A. H. Malik and S. Rahman, "Molecular erasers: Reprogramming cancer immunity through protein degradation," World Journal of Advanced Engineering Technology and Sciences, vol. 16, no. 3, pp. 277–291, Sep. 2025, doi: 10.30574/WJAETS.2025.16.3.1335.

[20] M. Sabbir Hossain et al., "Automatic Navigation and Self-Driving Technology in Agricultural Machinery: A State-of-the-Art Systematic Review," IEEE Access, vol. 13, pp. 94370–94401, 2025, doi: 10.1109/ACCESS.2025.3573324.

[21] M. S. Hossain, M. Rahman, M. Ahmed, M. M. Kabir, M. F. Mridha, and J. Shin, "DeepLabv3Att: Integrating Attention Mechanisms in DeepLabv3 for Enhanced Road Segmentation," 2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2024, pp. 711–718, 2024, doi: 10.1109/3ICT64318.2024.10824285.

[22] M. Rahman, M. S. Hossain, A. A. Eva, M. M. Kabir, M. F. Mridha, and J. Shin, "Alzheimers Disease Classification with a Hybrid CNN-SVM Approach on Enhanced MRI Data," 2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2024, pp. 50–57, 2024, doi: 10.1109/3ICT64318.2024.10824633.

[23] M. Rahman, M. S. Hossain, U. Rozario, S. Roy, M. F. Mridha, and N. Dey, "MultiSenseNet: Multi-Modal Deep Learning for Machine Failure Risk Prediction," IEEE Access, vol. 13, pp. 120404–120416, 2025, doi: 10.1109/ACCESS.2025.3586978.

[24] M. S. Hossain, M. Rahman, M. G. R. Abir, J. Maua, and A. Rahman, "Plant Leaf Disease Detection Using Deep Stacking: Integrating CNNs and Gradient Boosting for Enhanced Classification Accuracy," Studies in Computational Intelligence, vol. 1202, pp. 15–26, 2025, doi: 10.1007/978-981-96-4520-6_2.

[25] E. Dama et al., "Biomarkers and lung cancer early detection: State of the art," Cancers (Basel), vol. 13, no. 15, p. 3919, Aug. 2021, doi: 10.3390/CANCERS13153919/S1.

[26] A. Asuntha and A. Srinivasan, "Deep learning for lung Cancer detection and classification," Multimed Tools Appl, vol. 79, no. 11–12, pp. 7731–7762, Mar. 2020, doi: 10.1007/S11042-019-08394-3/METRICS.

[27] M. Al-Huseiny, A. S.-I. J. of Electrical, and undefined 2021, "Transfer learning with GoogLeNet for detection of lung cancer," academia.eduMS Al-Huseiny, AS SajitIndonesian Journal of Electrical Engineering and computer science, 2021•academia.edu, vol. 22, no. 2, pp. 1078–1086, 2021, doi: 10.11591/ijeecs.v22.i2.pp1078-1086.

[28] N. S. Reddy and V. Khanaa, "Intelligent deep learning algorithm for lung cancer detection and classification," Bulletin of Electrical Engineering and Informatics, vol. 12, no. 3, pp. 1747–1754, Jun. 2023, doi: 10.11591/EEI.V12I3.4579.

[29] F. Manapov et al., "PET/CT for Target Delineation of Lung Cancer Before Radiation Therapy," Semin Nucl Med, vol. 52, no. 6, pp. 673–680, Nov. 2022, doi: 10.1053/J.SEMNUCLMED.2022.05.003.

[30] J. Debnath et al., "LMVT: A hybrid vision transformer with attention mechanisms for efficient and explainable lung cancer diagnosis," Inform Med Unlocked, vol. 57, p. 101669, Jan. 2025, doi: 10.1016/J.IMU.2025.101669.

[31] I. J. Ogundana, S. Y. Foo, Z. Yu, P. Chaturvedi, A. Jhamb, and V. Nemade, "Prediction and Classification of Lung Cancer Using Machine Learning Techniques," IOP Conf Ser Mater Sci Eng, vol. 1099, no. 1, p. 012059, Mar. 2021, doi: 10.1088/1757-899X/1099/1/012059.

[32] S. R. R. B R, S. Sen, R. Bhatt, M. L. Dhanetwal, M. Sharma, and R. Naaz, "Stacked neural nets for increased accuracy on classification on lung cancer," Measurement: Sensors, vol. 32, p. 101052, Apr. 2024, doi: 10.1016/J.MEASEN.2024.101052.

[33] A. A. Shah, H. A. M. Malik, A. H. Muhammad, A. Alourani, and Z. A. Butt, "Deep learning ensemble 2D CNN approach towards the detection of lung cancer," Scientific Reports 2023 13:1, vol. 13, no. 1, pp. 1–15, Feb. 2023, doi: 10.1038/s41598-023-29656-z.

[34] A. Hossain et al., "Transformer-Based Ensemble Model for Binary and Multiclass Oral Cancer Segmentation," 2025 International Conference on Electrical, Computer and Communication Engineering, ECCE 2025, 2025, doi: 10.1109/ECCE64574.2025.11012921.

[35] A. Al Noman et al., "Monkeypox Lesion Classification: A Transfer Learning Approach for Early Diagnosis and Intervention," 2024 7th International Conference on Contemporary Computing and Informatics (IC3I), pp. 247–254, Sep. 2024, doi: 10.1109/IC3I61595.2024.10828678.

[36] M. A. Imran et al., "IoT-Enabled Smart Manhole Management System for Real-time Status, Water Level, and Gas Detection," 2024 International Conference on Intelligent Systems for Cybersecurity, ISCS 2024, 2024, doi: 10.1109/ISCS61804.2024.10581265.

[37] J. Debnath et al., "Hybrid Vision Transformer Model for Accurate Prostate Cancer Classification in MRI Images," 2025 International Conference on Electrical, Computer and Communication Engineering, ECCE 2025, 2025, doi: 10.1109/ECCE64574.2025.11013952.

[38] S. Khan et al., "Ensemble-Based Explainable Approach for Rare Medicinal Plant Recognition and Conservation," 2025 10th International Conference on Information and Network Technologies, ICINT 2025, pp. 88–93, 2025, doi: 10.1109/ICINT65528.2025.11030872.

[39] H. Rahman et al., "Automated Weed Species Classification in Rice Cultivation Using Deep Learning," 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1–6, Feb. 2025, doi: 10.1109/ECCE64574.2025.11014047.

[40] M. S. Rahman et al., "Effective Disease Recognition in Cucumbers: A Web-Based Application Using Transfer Learning Models," 2024 IEEE 3rd International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things, RAAICON 2024 - Proceedings, pp. 59–64, 2024, doi: 10.1109/RAAICON64172.2024.10928353.

[41] M. I. Hossain Siddiqui et al., "Eggplant Disease Diagnosis Using a Robust Ensemble of Transfer Learning Architectures," 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1–6, Feb. 2025, doi: 10.1109/ECCE64574.2025.11013918.

[42] R. Haque et al., "Advancements in Jute Leaf Disease Detection: A Comprehensive Study Utilizing Machine Learning and Deep Learning Techniques," 2024 IEEE International Conference on Power, Electrical, Electronics and Industrial Applications (PEEIACON), pp. 248–253, Sep. 2024, doi: 10.1109/PEEIACON63629.2024.10800378.