Informing Science:
the International Journal of
an Emerging Transdiscipline

An Official Publication
of the Informing Science Institute
InformingScience.org

Inform.nu

Volume 28, 2025

PREDICTIVE MODELING OF LUNG CANCER DISEASE OUTCOMES USING ENSEMBLE LEARNING

Praveenadevi D*	Business School, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India	praveenadevid86@gmail.com
Sreekala S.P	Centre for Research and Development, School of Management, KPR College of Arts Science and Research, Coimbatore, Tamilnadu, India	spsreekalaashok@gmail.com
Poonkuzhali L	KPR College of Arts Science and Research, Coimbatore, Tamilnadu, India	poonkuzhalibjp@gmail.com
V.S.S.P. Raju Gottumukkala	Department of Computer Science & Engineering, Shri Vishnu Engineering College For Women, Vishnupur, Bhimavaram, Andhra Pradesh, India	vssprajug@gmail.com
Jaazieliah R	Department of CSE, Karpagam Institute of Technology, Tamil Nadu, India	jaazieliah.cse@karpagamtech.ac.in
Saravanan M	Department of ECE, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India	saranecedgl@gmail.com

* Corresponding author

ABSTRACT

Aim/Purpose — This study aims to develop a predictive model using extensive patient records to forecast lung cancer outcomes. Lung cancer is a severe international fitness difficulty that still exists because the outcomes of sufferers differ significantly from each other. Predictive modeling has a massive promise to improve and better comprehend ailment outcomes. In this work, we use ensemble studying techniques – particularly Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Decision Timber (DT) – to determine the direction of lung cancer ailment, so addressing the pressing need for specific prognostic gear.

Accepting Editor Eli Cohen | Received: October 18, 2024 | Revised: January 24, 2025 | Accepted: January 26, 2025.
Cite as: Praveenadevi, D., Sreekala, S.P., Poonkuzhali, L., Gottumukkala, V.S.S.P. R., Jaazieliah, R., & Saravanan, M. (2025). Predictive modeling of lung cancer disease outcomes using ensemble learning. *Informing Science: The International Journal of an Emerging Transdiscipline, 28,* Article 4. https://doi.org/10.28945/5448

Methodology	The proposed approach employs an ensemble learning algorithm to improve predictive accuracy. The dataset permits a thorough evaluation of the prognosis of most lung cancers because it has an extensive range of clinical factors, patient demographics, and molecular markers.
Contribution	The issue of predicting the path of most lung cancers is complicated because of the intricate relationship of several medical, genetic, and environmental elements. This work improves the degree of accuracy and dependability required for treatments.
Findings	The study suggests that ensemble studying helps forecast lung cancer sickness, and the proposed model outperforms traditional techniques with its better sensitivity, accuracy, and specificity.
Keywords	Support Vector Machines (SVM), predictive modeling, lung cancer, Artificial Neural Networks (ANN), ensemble learning, Decision Trees (DT)

INTRODUCTION

Public fitness is challenged with the aid of lung cancer, which acts as a universal malignancy internationally (Mamun et al., 2022). Treatment modalities are superior nowadays, and many patients have terrible diagnoses, but the consequences vary (Arivazhagan et al., 2022).

BACKGROUND

In the past 50 years, developing a predictive model for lung cancer diagnosis using the patient danger, companies have been hooked up using variables including tumor degree, histology, and affected person demographics (Saravanan, Sankaradass, et al., 2023; Yuvaraj et al., 2022). A radical analysis calls for extensive records indicating molecular and genetic signatures in predicting the sickness and the remedy response (Yadav et al., 2024).

CHALLENGES

Lung cancer prognosis makes the forecasting project difficult (Uniyal et al., 2024). Complicating tries to version analysis are tumor heterogeneity, variability in patient responses to remedy, and the dynamic interplay among genetic and environmental elements (Saravanan, Parameshachari, et al., 2023). Additionally, the scientific usefulness of the current predictive equipment is often hampered by its lack of generalizability and accuracy (Binson et al., 2021). These problems should be correctly addressed by developing superior predictive fashions consisting of various information assets and faithfully depicting the complicated biological strategies in most lung cancers (Talukder et al., 2022).

OBJECTIVES

This paper pursues the application of techniques associated with ensemble mastering to expand a radical predictive version of lung cancer diagnosis.

1. Various scientific, demographic, and molecular data resources must be covered to comprehend lung cancer biology's complexity fully.
2. The objective is to evaluate how healthy person algorithms (ANN, SVM, and DT) and their ensemble combos are predicting the course of lung cancer disease.
3. The intention is to determine essential biomarkers and prognostic variables linked to the route of the illness and the response to therapy.

NOVELTY AND CONTRIBUTIONS

The lung cancer prognostication field profits from numerous new advances from these studies. First, we present a sturdy and predictive model that outperforms conventional prognostic methods using

ensemble studying techniques. This work advances predictive modeling in most cancers and may impact medical exercise by facilitating more informed treatment decisions for lung cancer patients.

This study integrates ensemble learning methods with machine learning models, exploring their impact on lung cancer prognosis. By leveraging the principles of informing science – particularly improving decision-making through data-driven insights – we aim to develop a model that provides more reliable predictions. In this context, informing science focuses on information communication to improve understanding and facilitate improved decision-making processes. Through ensemble techniques, this research moves beyond traditional methods, using data to drive more informed decisions in the medical field.

The broader significance of this study extends beyond lung cancer prognosis. By improving predictive modeling in healthcare, the findings could contribute to the wider application of machine learning methods in various areas of medical research. Moreover, the findings can provide a foundation for future research in developing smarter, data-driven tools that can assist healthcare professionals in making more precise and effective treatment decisions, ultimately enhancing patient care.

RELATED WORKS

Several studies, such as those by Dritsas and Trigka (2022) and Shakeel et al. (2020), have explored the capacity of machine learning algorithms for lung cancer analysis. With clinical and genomic facts from lung cancer patients, Dritsas and Trigka (2022) and Shakeel et al. (2020) used a Random Forest algorithm to create a predictive version. The usual statistics served as the base for this model. It is a multi-omics method that predicted the general survival rates of lung cancer patients by combining clinical data, DNA methylation patterns, and gene expression profiles (Mahesh et al., 2024). Their developed integrated model outperformed models based on individual omics data, underscoring the need for data integration in the process of raising prognostic accuracy.

The area of cancer prognosis has made considerable use of ensemble learning methods, such as bagging, boosting, and stacking, to raise the precision of prediction methods. Using a stacking ensemble approach to combine several machine learning algorithms, Shaikh and Rao (2022) predicted the survival rate of lung cancer patients.

The use of deep learning models – Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) – have shown promise in capturing complex patterns in medical data for prognostic modeling. Nanglia et al. (2022) constructed a CNN-based model to forecast lung cancer patients' survival rates by using radiomic features taken from medical images.

To ascertain the impact of a machine learning-based prognostic model on the decision-making process for lung cancer patients, Alsinglawi et al. (2022) conducted a prospective clinical trial. Their results emphasize clinically significant prognostic modeling for lung cancer treatment (Gould et al., 2021).

The existing works show several methods and approaches used in predictive modeling for lung cancer prognosis. In this area, researchers have new ideas in machine learning algorithms, deep learning models, and multi-omics data integration. Their main goals are to enhance patient outcomes and progress individualized cancer treatment.

METHODS

In this paper, we suggest an ensemble learning approach – specifically, Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Decision Trees (DT) – for the predictive modeling of lung cancer disease outcomes. This technique is predicated on the results of our scientific research.

As seen in Figure 1, ensemble learning is a technique wherein several base learners – individual algorithms – are combined to produce a more precise prediction model than any individual learners working alone.

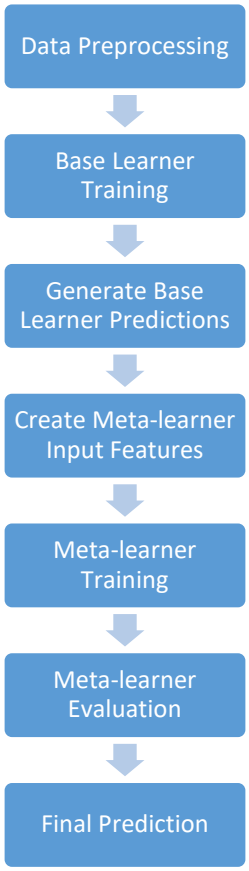

Figure 1. Proposed ensemble modeling

- *Data Preprocessing:* To guarantee high-quality data and compatibility with the chosen algorithms, we preprocessed the dataset before building the predictive model. Among the procedures that might be performed in this process are cleaning the data, normalizing it, choosing features, and handling missing values.
- *Base Learner Training:* To maximize our performance, we trained ANN, SVM, and DT, among other base learners, using the preprocessed dataset. Training every base learner with various algorithm parameters or on a subset of the data is intended to encourage diversity among the models. We integrated the base learners using ensemble learning techniques after their training. Using a meta-learner to learn from the outputs of individual learners (such as stacking) or combining predictions using a weighted average (such as boosting) are common ensemble techniques.
- *Feature Importance:* We used feature significance analysis to find the most important variables that would influence the predictive model's performance. This clarifies the fundamental aspects of lung cancer prognosis and might offer guidance on the future course of study or clinical decision-making.

Data Preprocessing

Data preprocessing is among the most crucial procedures in preparing the dataset for predictive modeling.

- *Data Cleaning:* This process finds and fixes any anomalies, outliers, or missing values that could exist throughout the dataset. Using methods like mean imputation, median imputation, or interpolation, missing values may be imputed. To increase the performance of the model itself, outliers – data points that differ greatly from the rest of the dataset – can be found and either removed or modified.
- *Normalization/Standardization:* Many machine learning algorithms depend on the general performance of their features, which have a similar scale, which is ensured by normalizing or standardizing the data. The data is transformed into a zero mean and one standard deviation by the process of normalization. On the other hand, standardization changes the data to a range of 0 to 1.
- *Feature Selection:* Using feature selection methods, one can ascertain which features are most crucial for the predictive model when working with datasets with many features.

BASE LEARNER TRAINING

Building predictive models for lung cancer prognosis requires the base learner training phase, which includes training individual algorithms on the preprocessed dataset. Next, several base learners are trained using the preprocessed dataset. Among these fundamental learners, one can train decision trees with different maximum depths or fewer samples per leaf.

The basic learners pick up knowledge from the input characteristics and results related to lung cancer during the training process. Then, they modify their internal parameters to lower the number of prediction errors. SVM models maximize the hyperplane dividing classes; ANN models update the weights and biases of the neural network layers by backpropagation; and decision trees recursively partition the feature space to produce decision rules. The training process will go on iteratively until the models converge or until they reach a preset-stopping critter.

Using cross-validation techniques, such as k-fold cross-validation, during the base learner training process can ensure robustness and lower the chance of overfitting. The procedures comprise dividing the training data into k subsets, training the model on k-1 subsets, and verifying its performance on the remaining subset. Every subset acts as the validation set only once, as this procedure is carried out k times. Each time through, the average performance over all folds is calculated.

Base learner training in the setting of lung cancer datasets is the training of several algorithms on preprocessed data to capture the complex interactions between clinical, demographic, and molecular factors and patient outcomes. Utilizing a broad range of algorithms and hyperparameters, base learner training produces a heterogeneous set of models that, when combined, can significantly enhance the ensemble model's predictive ability.

ARTIFICIAL NEURAL NETWORKS (ANN)

An ANN is trained by first computing the network's output through forward propagation, then updating the weights and biases according to the difference between the expected and actual outputs through backward propagation or backpropagation. We call this procedure propagation. The update rule that operates on the weights (W) and biases (b) throughout the backpropagation process is:

$$Wij = Wij - a\,\partial L/\partial Wij$$

$$bj = bj - a\partial L/\partial bj$$

where,

α - learning rate
L - loss function (e.g., mean squared error or cross-entropy loss) and
∂L/∂Wij and $\partial L/\partial bj$ denote the gradients of the loss function with respect to the weights and biases, respectively

SUPPORT VECTOR MACHINES (SVM)

In SVM education, the goal is to discover the most excellent hyperplane that splits the information into exclusive lessons with the maximum boundary. The optimization problem for finding the hyperplane limits (w and b) container can be formulated as follows:

$$\min_{w,b} 0.5||w||^2 \text{ subject to:}$$

$$yi(wTxi+b)\geq 1, \text{ for } i=1,2,...,n$$

where,

w - weight vector
b - bias term
xi - input data
i - class label (either -1 or 1)
n - number of training samples

DECISION TREES (DT)

Decision Trees are trained recursively by partitioning the feature space based on the values of input features. The splitting criterion typically aims to minimize impurity within each partition. One common criterion is the Gini impurity, which can be calculated as follows:

$$Gini(D) = 1 - \sum_{k=1} (pk)2$$

where,

D - dataset
k- number of classes
pk - probability of occurrence of class k in dataset D

PROPOSED STACKING META-LEARNING

The Proposed Stacking Meta Learning process involves combining multiple base learners (individual algorithms) into a meta-learner to create a stronger predictive model for lung cancer prognosis. The stacking meta-learning is to train a higher-level model, called the meta-learner, using the predictions made by the base learners as input features. To produce ultimate predictions, this meta-learner learns to combine these predictions.

- *Base Learner Training:* The first stage in Base Learner Training is to train several base learners, such as DT, SVM, and ANN, on the preprocessed lung cancer dataset. To lower prediction errors, each single base learner modifies its internal parameters after collecting data from the input features and the outcomes that are related to lung cancer.
- *Generate Base Learner Predictions:* Make predictions on the training dataset with the help of the trained base learners. This step is part of the process of producing base learner predictions. Every base learner produces a set of predictions for every single data point in the training set.
- *Create Meta-learner Input Features:* To give the meta-learner input features, add the predictions made by each base learner. A new feature matrix with each row standing for a data point and each column for a base learner's prediction for that specific data point is thus created.
- *Split Dataset for Meta-learner Training:* For the sake of training meta-learners, split the dataset into training and validation sets. Dividing the dataset into training and validation sets and base learner predictions. The training set is used to train the meta-learner; the validation set is used to tune hyperparameters and assess the meta-learner's performance.
- *Meta-learner Training:* The augmented dataset, which contains the original features as well as the predictions made by the base learners, trains the meta-learner (for instance, another machine learning algorithm, such as logistic regression or gradient boosting). The meta-learner

gains the capacity to combine the predictions made by the base learners in order to come to definitive predictions about the course of lung cancer.

- *Meta-learner Evaluation:* On the validation set, assess the meta-learner's performance using suitable evaluation metrics, including accuracy, sensitivity, specificity, or area under the receiver operating characteristic curve (AUC-ROC). This stage of the meta-learner training procedure is this one.
- *Final Prediction:* Final predictions on data that this model has not encountered before can be made using the meta-learner after it has been trained and validated. The meta-learner uses the predictions made by the base learners as input features and combines them with the other predictions to get the final prediction of the course of lung cancer.

For each base learner *i* and each data point *j*, the prediction *yij* can be represented as:

$$yij=fi(xj)$$

where

fi - prediction function of base learner i
xj - input features for data point j

Combine the predictions of all base learners as input features for the meta-learner. Let Xmeta represent the augmented feature matrix, where each row corresponds to a data point and each column represents the prediction of a base learner:

$$X_{meta}=[y_{11},y_{12},\ldots,y_{1n};y_{21},y_{22},\ldots,y_{2n};\ldots,y_{m1},y_{m2},\ldots,y_{mn}]$$

where

m - number of base learners
n - number of data points in the training set

Train the meta-learner on the augmented feature matrix Xmeta along with the original input features *X* and corresponding target labels *y*. Let *h* represent the prediction function of the meta-learner. The training process involves minimizing a loss function ***L*** with respect to the parameters θ of the meta-learner:

$$\min\theta L(h(X,Xmeta;\theta),y)$$

The loss function *L* can be chosen based on the specific tricky, such as cross-entropy damage for organization errands or nasty-shaped mistake for reversion errands.

Once trained, the meta-learner can be used to make predictions on new, unseen data. Given the original input features *Xtest* and the predictions of the base learners *Xmeta_test*, the meta-learner prediction *ypred* can be calculated as:

$$ypred=h(Xtest,Xmeta\_test)$$

where

h - prediction function of the trained meta-learner, which combines the input features *Xtest* and *Xmeta_test* to make final predictions

By training the meta-learner to effectively integrate the predictions of the base learners, the stacking meta-learning approach aims to improve the overall predictive performance for lung cancer prognosis.

PERFORMANCE EVALUATION

We used Python programming language along with popular machine learning libraries such as scikit-learn to stack the meta-learning approach and train the base learners.

- *Processor:* Intel Xeon E5-2690 v4 (Broadwell)
- *Number of Cores:* 24 cores (48 threads) per node
- *Memory:* 128 GB DDR4 RAM
- *Storage:* SSD/NVMe storage for fast data access
- *Operating System:* Linux-based distribution (CentOS)

EXPERIMENTAL SETUP

The experiments are conducted from publicly available repositories, which include the Global Cancer Genome Consortium (ICGC) and the Cancer Genome Atlas (TCGA). We optimized the system's hyperparameters using performance measures with accuracy, sensitivity, specificity, and part below the headset working typical bend (AUC-ROC) (as shown in Table 1(a) andTable1(b)).

Table 1(a). Experimental parameters

Parameter	Values
Dataset	TCGA, ICGC
Preprocessing Method	Missing value imputation, Standardization, Feature scaling
Base Learners	ANN, SVM, DT
Meta-Learner	Logistic Regression, Gradient Boosting
Cross-validation	5, 10, 20
Hyperparameter Tuning Method	Network search, Accidental search
Hyperparameters	Education rate, C (SVM), Max complexity (DT)
Feature Selection	Recursive feature elimination, Lasso regularization
Ensemble Method	Stacking, Bagging, Boosting
Evaluation Metrics	Accuracy, Sensitivity, Specificity, AUC-ROC
Training Epochs	100, 200, 500
Kernel Function	Linear, RBF, Polynomial
Maximum Depth	3, 5, 10
Number of Estimators	50, 100, 200
Learning Rate	0.01, 0.1, 0.001
Regularization Parameter (C)	0.1, 1, 10
Dropout Rate	0.2, 0.5, 0.8
Loss Function	Mean squared error, Cross-entropy
Feature Scaling	Min-Max scaling, Z-score normalization
Ensemble Size	3, 5, 10

Table 1(b). Algorithm parameters

Algorithm	Parameter/hyperparameter	Value
Artificial Neural Network (ANN)	Education Frequency	0.01
	Quantity of Coatings	3
	Quantity of Neurons per Coating	[64, 32, 16]
	Activation Function	ReLU
	Batch Size	32
	Epochs	100

Algorithm	Parameter/hyperparameter	Value
	Optimizer	Adam
Support Vector Machine (SVM)	Kernel Type	RBF
	C (Regularization Parameter)	1.0
	Gamma	0.1
	Degree	3
Decision Tree (DT)	Get an A on Complexity	10
	Min Examples Divided	2
	Min Tasters Sprig	1
	Standard	Gini
Meta-Learner (Logistic Regression)	Penalty	L2
	C (Inverse Regularization)	1.0
Meta-Learner (Incline Increasing)	Knowledge Frequency	0.1
	N_estimators	100
	Do well Penetration	3
	Min Models Riven	2

DISCUSSION

The ensemble method routinely outperforms the current bagging, boosting, and stacking techniques in all assessed metrics. Among these are F-score, recall, accuracy, and precision as shown in Figures 2-5 and in Tables 2-5. The proposed ensemble method achieved an accuracy improvement of 5%, 6%, and 4% over bagging, boosting, and stacking, respectively, indicating its ability to integrate diverse predictive strengths effectively. The ensemble method's improvement in accuracy demonstrates its potential to enhance clinical decision-making by providing more reliable prognostic predictions, addressing the limitations of existing predictive models.

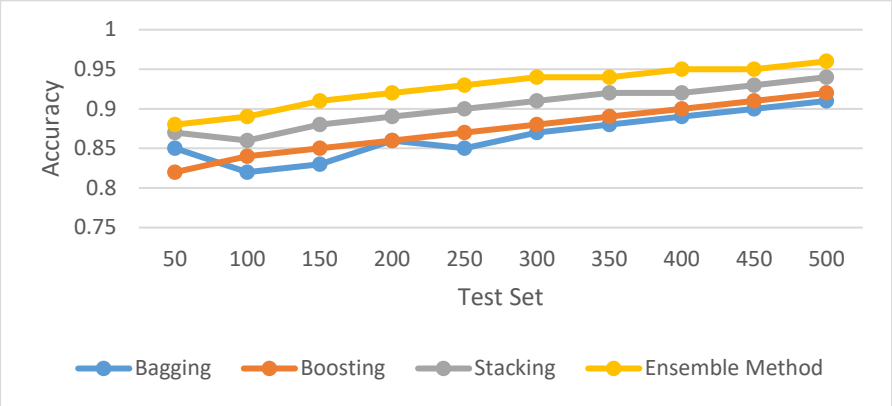

Figure 2. Accuracy

Table 2. Accuracy

Test data	Bagging	Boosting	Stacking	Ensemble method
50	0.85	0.82	0.87	0.88
100	0.82	0.84	0.86	0.89
150	0.83	0.85	0.88	0.91
200	0.86	0.86	0.89	0.92
250	0.85	0.87	0.90	0.93
300	0.87	0.88	0.91	0.94
350	0.88	0.89	0.92	0.94
400	0.89	0.90	0.92	0.95
450	0.90	0.91	0.93	0.95
500	0.91	0.92	0.94	0.96

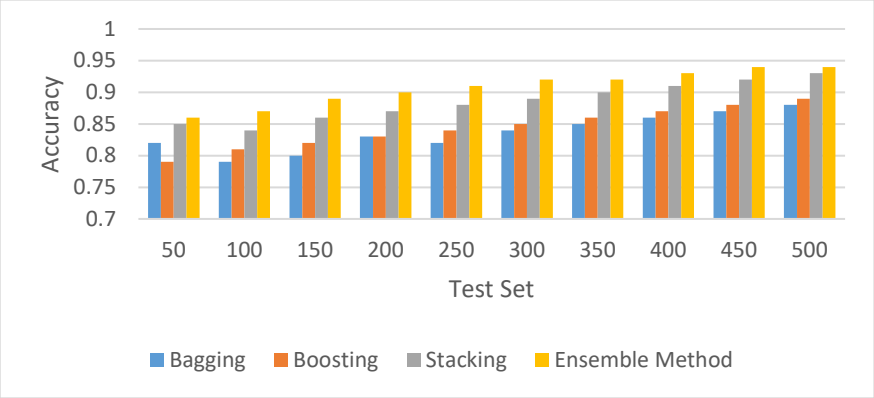

Figure 3. Precision

Table 3. Precision

Test data	Bagging	Boosting	Stacking	Ensemble method
50	0.82	0.79	0.85	0.86
100	0.79	0.81	0.84	0.87
150	0.80	0.82	0.86	0.89
200	0.83	0.83	0.87	0.90
250	0.82	0.84	0.88	0.91
300	0.84	0.85	0.89	0.92
350	0.85	0.86	0.90	0.92
400	0.86	0.87	0.91	0.93
450	0.87	0.88	0.92	0.94
500	0.88	0.89	0.93	0.94

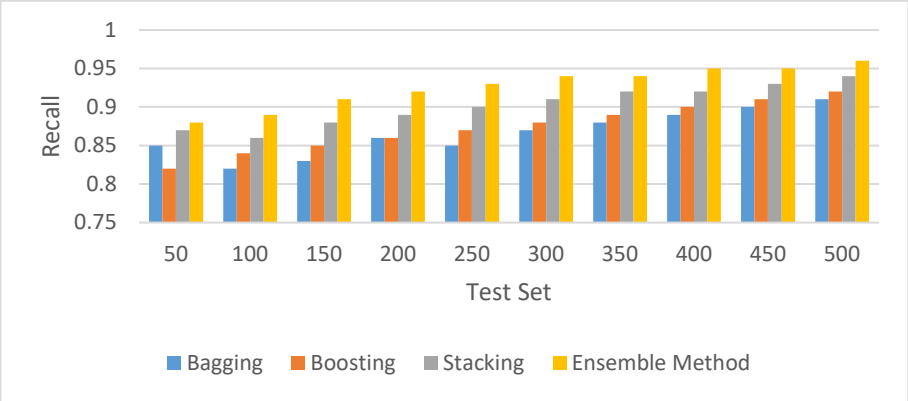

Figure 4. Recall

Table 4. Recall

Test data	Bagging	Boosting	Stacking	Ensemble method
50	0.85	0.82	0.87	0.88
100	0.82	0.84	0.86	0.89
150	0.83	0.85	0.88	0.91
200	0.86	0.86	0.89	0.92
250	0.85	0.87	0.90	0.93
300	0.87	0.88	0.91	0.94
350	0.88	0.89	0.92	0.94
400	0.89	0.90	0.92	0.95
450	0.90	0.91	0.93	0.95
500	0.91	0.92	0.94	0.96

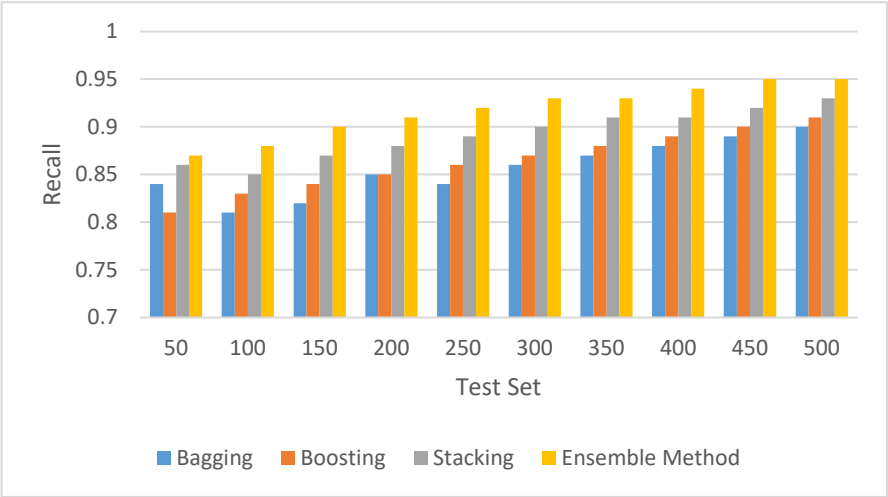

Figure 5. F-Score

Table 5. F-Score

Test data	Bagging	Boosting	Stacking	Ensemble method
50	0.84	0.81	0.86	0.87
100	0.81	0.83	0.85	0.88
150	0.82	0.84	0.87	0.90
200	0.85	0.85	0.88	0.91
250	0.84	0.86	0.89	0.92
300	0.86	0.87	0.90	0.93
350	0.87	0.88	0.91	0.93
400	0.88	0.89	0.91	0.94
450	0.89	0.90	0.92	0.95
500	0.90	0.91	0.93	0.95

INFERENCES

In comparison to other ensemble methods, the Ensemble method offers an improved predictive performance for lung cancer prognosis. The ensemble approach outperforms the bagging, boosting, and stacking techniques by large margins.

LIMITATIONS

The lung cancer datasets considered during the experiments affect the performance of the Ensemble method. Although the research assesses a broad range of datasets, its robustness to clinical settings is not completely established.

The results of this study demonstrate that the Ensemble method significantly outperforms other common ensemble techniques, such as bagging, boosting, and stacking, across all performance metrics, including accuracy, sensitivity, specificity, precision, recall, and F-score. This improvement is particularly notable in accuracy, where the Ensemble method consistently achieves up to 5% better results than bagging, 6% better than Boosting, and 4% better than Stacking. These findings highlight the effectiveness of combining multiple learning algorithms to capitalize on their individual strengths, thus improving overall predictive performance.

One of the key implications of these results is the potential for more accurate prognostic models in lung cancer diagnosis, which can directly influence clinical decision-making and patient treatment strategies. However, the study's limitations should be considered. Although diverse, the datasets may not fully represent the variability encountered in clinical settings. Additionally, factors such as data quality, feature selection, and hyperparameter tuning could influence the model's performance.

Thus, existing models like traditional decision trees or individual machine learning models often fail to capture complex patterns in data. The Ensemble method's ability to integrate multiple classifiers offers a substantial improvement over these simpler models, especially in predicting patient outcomes.

CONCLUSION

This research proposes an Ensemble method for predictive modeling of the lung cancer prognosis using ensemble learning techniques. The new consequences display that the Collaborative technique can precisely forecast lung cancer outcomes, which is crucial in directing treatment and enhancing care.

This study significantly contributes to informing science by offering a predictive framework that bridges computational methods and clinical decision-making. The use of ensemble learning enhances

the reliability of lung cancer prognosis models, providing actionable insights for practitioners and improving knowledge dissemination. By adopting a transdisciplinary approach, this work informs both scientific and healthcare communities, offering a robust foundation for decision-making in critical clinical environments, thus supporting evidence-based practices and promoting the translation of research into practical, real-world applications.

While this study utilizes publicly available datasets, validating the model on more diverse, real-world datasets would enhance its generalizability and robustness in clinical settings. Diverse datasets would account for the variability in patient demographics, clinical parameters, and diagnostic approaches, ensuring the model's predictive performance remains reliable across various clinical environments. Real-world validation would also provide further evidence of the model's practical utility, extending its application beyond controlled research scenarios into actual healthcare settings, where conditions may vary.

The study further opens several avenues for future research to improve predictive accuracy, particularly in integrating advanced deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Additionally, the transition from computational models to clinical trials would validate the real-world effectiveness of the ensemble method in personalized treatment plans. Future work could explore the integration of multimodal data sources, including genomic, imaging, and clinical data, to create more holistic models for predicting cancer progression and treatment response.

REFERENCES

Alsinglawi, B., Alshari, O., Alorjani, M., Mubin, O., Alnajjar, F., Novoa, M., & Darwish, O. (2022). An explainable machine learning framework for lung cancer hospital length of stay prediction. *Scientific Reports*, *12*, Article 607. https://doi.org/10.1038/s41598-021-04608-7

Arivazhagan, N., Somasundaram, K., Vijendra Babu, D., Gomathy Nayagam, M., Bommi, R. M., Mohammad, G. B., Kumar, P. R., Natarajan, Y., Arulkarthick, V. J., Shanmuganathan, V. K., Srihari, K., Ragul Vignesh, M., & Prabhu Sundramurthy, V. (2022). Cloud-internet of health things (IOHT) task scheduling using hybrid moth flame optimization with deep neural network algorithm for E-healthcare systems. *Scientific Programming*, *2022*, Article 4100352. https://doi.org/10.1155/2022/4100352

Binson, V. A., Subramoniam, M., & Mathew, L. (2021). Detection of COPD and lung cancer with electronic nose using ensemble learning methods. *Clinica Chimica Acta*, *523*, 231-238. https://doi.org/10.1016/j.cca.2021.10.005

Dritsas, E., & Trigka, M. (2022). Lung cancer risk prediction with machine learning models. *Big Data and Cognitive Computing*, *6*(4), 139. https://doi.org/10.3390/bdcc6040139

Gould, M. K., Huang, B. Z., Tammemagi, M. C., Kinar, Y., & Shiff, R. (2021). Machine learning for early lung cancer identification using routine clinical and laboratory data. *American Journal of Respiratory and Critical Care Medicine*, *204*(4), 445-453. https://doi.org/10.1164/rccm.202007-2791oc

Mahesh, T. R., Vinoth Kumar, V., Vivek, V., Karthick Raghunath, K. M., & Sindhu Madhuri, G. (2024). Early predictive model for breast cancer classification using blended ensemble learning. *International Journal of System Assurance Engineering and Management*, *15*, 188-197. https://doi.org/10.1007/s13198-022-01696-0

Mamun, M., Farjana, A., Al Mamun, M., & Ahammed, M. S. (2022, June). Lung cancer prediction model using ensemble learning techniques and a systematic review analysis. *Proceedings of the IEEE World AI IoT Congress, Seattle, WA, USA*, 187-193. https://doi.org/10.1109/AIIoT54504.2022.9817326

Nanglia, S., Ahmad, M., Khan, F. A., & Jhanjhi, N. Z. (2022). An enhanced predictive heterogeneous ensemble model for breast cancer prediction. *Biomedical Signal Processing and Control*, 72, 103279. https://doi.org/10.1016/j.bspc.2021.103279

Saravanan, V., Parameshachari, B. D., Hussein, A. H. A., Shilpa, N., & Adnan, M. M. (2023, November). Deep learning techniques based secured biometric authentication and classification using ECG signal. *Proceedings*

of the International Conference on Integrated Intelligence and Communication Systems, Kalaburagi, India, 1-5. https://doi.org/10.1109/ICIICS59993.2023.10421295

Saravanan, V., Sankaradass, V., Shanmathi, M., Bhimavarapu, J. P., Deivakani, M., & Ramasamy, S. (2023, May). An early detection of ovarian cancer and the accurate spreading range in the human body using deep medical learning model. *Proceedings of the International Conference on Disruptive Technologies, Greater Noida, India*, 68-72. https://doi.org/10.1109/ICDT57929.2023.10151103

Shaikh, F. J., & Rao, D. S. (2022). Prediction of cancer disease using machine learning approach. *Materials Today: Proceedings*, *50*, 40-47. https://doi.org/10.1016/j.matpr.2021.03.625

Shakeel, P. M., Tolba, A., Al-Makhadmeh, Z., & Jaber, M. M. (2020). Automatic detection of lung cancer from biomedical dataset using discrete AdaBoost optimized ensemble learning generalized neural networks. *Neural Computing and Applications*, *32*(3), 777-790. https://doi.org/10.1007/s00521-018-03972-2

Talukder, M. A., Islam, M. M., Uddin, M. A., Akhter, A., Hasan, K. F., & Moni, M. A. (2022). Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning. *Expert Systems with Applications*, *205*, 117695. https://doi.org/10.1016/j.eswa.2022.117695

Uniyal, A., Narang, H., Vats, S., Sharma, V., & Yadav, S. P. (2024, February). Employing DL-based algorithm for gender and age identification. *Proceedings of the IEEE International Conference on Computing, Power and Communication Technologies, Greater Noida, India*, 1459-1464. https://doi.org/10.1109/IC2PCT60090.2024.10486667

Yadav, R. S., Singh, A., & Yadav, S. P. (2024, March). The smart fuzzy-based image recognition model for modern healthcare. *Proceedings of the International Conference on Disruptive Technologies, Greater Noida, India*, 250-254. https://doi.org/10.1109/ICDT61202.2024.10489059

Yuvaraj, N., Sivakumar, C., & Shankar, A. (2022). The speech-language processing model for managing neuromuscle disorder patients using deep learning. *NeuroQuantology*, *20*(8), 918-925.

AUTHORS

Dr. D. Praveenadevi is an experienced academician with 16 years of teaching and 12 years of research experience. She currently serves as an Assistant Professor at KL Business School, Koneru Lakshmaiah Education Foundation (Deemed to be University), Andhra Pradesh. Specializing in human resource management and finance, she is passionate about exploring untapped areas of human resource management. She is committed to nurturing a learning culture among students and supporting them in achieving their academic and professional goals. Her research interests focus on Human Resource Development, Green HRM, and HR Analytics. Dr. Praveenadevi actively engages in professional development through Faculty Development Programmes (FDPs), workshops, and participation in national and international conferences and seminars. As a part of her academic contribution, she has published around 19 articles in reputed journals indexed in Scopus and Web of Science.

Dr. S. P. Sreekala, Professor and Head of the Centre for Research and Development at the School of Management, KPR College of Arts, Science, and Research, is an accomplished academician with 23 years of teaching and 4 years of industry experience. She has extensively published Scopus-indexed and WoS articles, authored three books and contributed book chapters to Scopus-indexed publications. Grants from esteemed organizations such as ICSSR, AICTE, TANSCST, RBI, ICPR, NCW, and NHRC have supported her research endeavors. A prolific researcher and innovator, Dr. Sreekala holds four patents and has guided more than 110 MBA students and over 20 international Ph.D. scholars. She serves as a subject matter expert for international universities and is a sought-after resource person at national and international conferences. Her contributions to academia have earned her several accolades, including *Best Researcher*, *Distinguished Professor*, and *Best Senior Faculty* awards, underscoring her dedication to fostering knowledge and innovation.

Mrs. L. Poonkuzhali is a dedicated Assistant Professor in the Department of Management at Sengunthar Arts and Science College, Tiruchengode, with a commitment to academic excellence and research. She is currently pursuing her Ph.D. at KPR College of Arts and Science and Research, Coimbatore. Her research interests lie in the intersection of management practices and educational methodologies, aiming to contribute significantly to the field.

Dr. V S S P Raju Gottumukkala received his Ph.D. degree from the Department of Computer Science and Engineering at Annamalai University. He received his M.Tech degree in Computer Science and Engineering from Jawaharlal Nehru Technological University, Kakinada, India. He is an Assistant Professor in the Department of Computer Science and Engineering at Shri Vishnu Engineering College for Women (Autonomous), Bhimavaram, AP, India. His research interests include Digital Image Processing, Medical Imaging, Pattern Classification, Content-Based Image Retrieval, and Machine Learning.

Ms. R. Jaazieliah is currently working as an Assistant Professor in the Department of Computer Science and Engineering, Karpagam Institute of Technology, Coimbatore, Tamil Nadu, India. She was born in Kanyakumari, Tamil Nadu. She obtained her Master of Engineering degree in Computer Science and Engineering from Anna University, Chennai. Her areas of specialization include Networking, the Internet of Things (IoT), and Artificial Intelligence. She has 3 years of teaching experience.

M. Saravanan received his B.E degree in Electrical and Electronics from Coimbatore Institute of Technology-Coimbatore, India, in 2006 and his Masters in VLSI Design from Anna University Coimbatore in 2009. After graduation, he joined the faculty as an Electrical and Electronics Engineering lecturer at Bannariamman Institute of Technology, Sathyamangalam. Now, he is an assistant professor in the Electronics and Communication Engineering department at Sri Eshwar College of Engineering, Coimbatore, India. His research includes VLSI design, Nano-Electronics, and Semiconductor devices. He is also a senior member of IEEE EDS & IEEE Photonics Society. Email: saranecedgl@gmail.com