OPEN ACCESS

Kashf Journal of Multidisciplinary Research
Vol: 02 - Issue 2 (2025)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

HYBRID APPROACHES IN MACHINE LEARNING: INTEGRATING ANT COLONY OPTIMIZATION FOR IMPROVED LUNG CANCER DETECTION

Bushra Hakeem
Department of Computer Science, NFCIET, Multan,Pakistan
Muhammad Kamran Abid
Department of Computer Science, NFCIET, Multan, Pakistan
Muhammad Baqer
Department of Computer Engineering, BZU, Multan, Pakistan
Yasir Aziz
Department of Computer Engineering, BZU, Multan, Pakistan
Naeem Aslam
Department of Computer Science, NFCIET, Multan, Pakistan
Nasir Umer
Department of Computer Science, NFCIET, Multan, Pakistan

Corresponding author: Yasir Aziz (engr.yasiraziz@bzu.edu.pk)
DOI: *https://doi.org/10.71146/kjmr286*

Article Info

Abstract

Early detection of lung cancer is crucial for improving survival rates, making it one of the deadliest diseases worldwide. Machine learning (ML) techniques have been increasingly adopted to improve diagnostic processes. However, individual algorithms face trade-offs between accuracy, precision, recall, and F1 scores. This study aimed to evaluate the performance of four ML algorithms—SVM, Random Forest, KNN, and CNN—for predicting lung cancer. Each model revealed strengths and weaknesses, particularly low recall values, suggesting a need for more comprehensive solutions. The research introduced a stacking ensemble model to enhance prediction performance. SVM showed a high precision of 0.983 but a low recall of 0.5, while Random Forest balanced accuracy (0.967) and recall (0.741). KNN and CNN also performed well, though they struggled with precision. To overcome these limitations, the stacking ensemble model combined SVM, Random Forest, and KNN, achieving an accuracy of 0.977, precision of 0.701, recall of 0.711, and an F1 score of 0.711. This approach maximized the strengths of each model, offering better generalization and robustness. Overall, the ensemble model proved more effective in healthcare applications, where accurate and early diagnosis is essential for successful treatment.

Keywords:
Machine Learning, Ant Colony Optimization, SVM, CNN, KNN

Introduction

Millions around the world are diagnosed with lung cancer every year, yet remains one of the most urgent health questions around the globe. To improve survival rates, early detection is crucial, yet X-rays, CT scans, and biopsies despite being accurate, can be expensive and may miss the disease early. Due to this, machine learning (ML) and deep learning (DL) have been on the path towards pushing forward for non-invasive, early detection of lung cancer. Yet despite these algorithms (e.g. Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), Convolutional Neural Networks (CNN)) showing some promise, they all have their issues of accuracy, precision, recall, F1 scores, etc. To improve the overall diagnostic performance, the research explores use of a stacking ensemble model that folds the strengths of individual algorithms[1], [2].

The ensemble model takes advantage of SVM, Random Forest, and KNN to maximize its prediction power. For example, SVM is doing a great job in precision by obtaining a score of 0.983 which assumes a high capability to make correct positive predictions. While its recall score of 0.5 indicates it misses many of the instances of lung cancer, the model is not fundamentally accurate, and therefore ineffective in identifying all lung cancer cases[3], [4]. However, Random Forest performs better in the balance, with an accuracy of 0.967 and a recall of 0.741, since it is better at locating positive cases and reducing amounts of false negatives. Although KNN is slightly less accurate (0.951) and has less recovery (0.733) KNN provides valuable insight into the data. Not surprisingly, CNN works pretty well, recording an accuracy of 0.945 and recall of 0.745, however, its precision performs modestly at 0.72[5].

By combining these algorithms here in the stacking ensemble model, the stacking ensemble model makes the total strength from these algorithms and the total weakness from these algorithms. The ensemble model has an overall accuracy of 0.977, precision of 0.701, recall of 0.711, and F1 score of 0.711, making it more capable of making balanced robust predictions than any of the individual algorithms[6]. The benefit of this approach is not just to improve diagnostic accuracy but reduce the risk of false positives and false negatives that are so important in the medical field where early and accurate detection is paramount to successful treatment outcomes.

Background

Extensive research has been done in developing early detection methods related to lung cancer, which is one of the most lethal forms of cancer worldwide. However, early diagnosis can certainly save lives[7]. Though certain traditional diagnostic methods such as chest X-rays, computed tomography (CT) scans, and tissue biopsies can be trusted, they are limited in that they only provide a snapshot of the organ they are investigating, rather than making a holistic assessment of the substance's metabolic health[8], [9]. Although commonly used because of availability and low cost, the sensitivity and specificity of chest X rays are poor, especially for small, early-stage tumors [10], [11]. While more sensitive, CT scans give patients high levels of radiation that in the case of multiple scans can be harmful [12]. These biopsies, although definitive, have an invasive component of potential infection and bleeding. As a result, there is an urgent need to seek alternatives to machine learning (ML) and deep learning (DL) approaches for diagnosis of lung cancer at early stages [13].

In the last few years, medical diagnostics has seen an uptick in the application of ML and DL for their capabilities to handle massive amounts of data and find complex patterns that are not easily picked out by more traditional methods. However, other ML algorithms such as Support Vector Machines (SVM), Random Forest (RF), and K nearest neighbors (KNN) have been widely used for predicting lung cancer[14], [15]. Each of these algorithms has strengths and weaknesses. For example, SVM is famed for its high accuracy and precision in the classifying of data points in binary classification like with lung cancer detection [16]. SVM however is one of the key problems in its sensitivity to the dataset size and the choice of the kernel functions thus resulting in overfitting or poor generalization of data which has not been seen by the model [17].

Unlike Random Forest, which employs ensemble learning, i.e., combining multiple decision trees to increase classification accuracy and reduce the chance of overfitting [18], random forest offers robust performance over a variety of datasets. It is a good choice in many studies of lung cancer detection because it does very well with both categorical and numerical data. Although Random Forest can be imprecise in certain scenarios where data contains a lot of irrelevant features, or is unbalanced, placing accuracy against recall tradeoff [19]. Due to its simple implementation and interpretability though, KNN (K-Nearest Neighbors), based on the principle of proximity, is frequently used in lung cancer prediction. Nevertheless, it is computationally expensive and sensitive to the value of 'k' [20], its performance suffers when the dataset is large or has many features.

Apart from traditional ML models, Convolutional Neural Networks (CNN) have demonstrated good results in analyzing medical images, i.e. in detecting lung cancer using CT scans or X-ray images [21], [22]. One of the particularly useful applications of CNNs has been their ability to automatically learn hierarchical features from raw input data to discriminate small nodules or subtle abnormalities in lung images. Though CNNs are computationally expensive and require large, annotated datasets for training, limitations to their applicability in some clinical settings may be caused by these restrictions. Furthermore, CNNs, like many other DL models, tend not to be “transparent” — that is, how decisions are made may not be made clear, therefore slowing their adoption in medical practice [23], [24].

All of these individual ML and DL models face a major challenge in achieving the trade-off between accuracy, precision, recall, and F1 score. As in a medical setting where improper diagnosis would mean very serious consequences, precision is concerned with diminishing false positives, while recall aims at lowering false negatives [25], [26]. However, there is no single algorithm that was superior to the rest across all evaluation metrics, which has encouraged the use of ensembling techniques to improve their performance in general [27]. Due to the strengths and weaknesses inherent among different algorithms, ensemble learning has developed into a very powerful approach to combining the predictions of multiple models.

Ensemble learning combines several base models (such as the SVM, Random Forest and K Nearest Neighbor) into a single model using a meta-learner with the ability to improve the prediction performance [28]. This technique has been widely used in a wide range of domains and has been proposed for medical applications in cancer detection [29], [30]. Stacking aggregates predictions of several models and effectively yields more accurate and stable predictions at lower misclassification rates and better model generalization. Stacking also allows us to combine various types of algorithms, each adept for its part of the data, to create a holistic, more reliable diagnostic tool [31].

individual ML and DL models have shown promise in the early detection of lung cancer, but the performance is constrained by trade-offs among different performance metrics. Stacking is the best-known ensemble learning algorithm that merges the strengths of multiple models and yields higher accuracy than the single best model in real-world clinical settings, while not significantly compromising on robustness. Given current research in this area, future development of ensemble techniques with the application of advanced optimization algorithms, such as Ant Colony Optimization (ACO), should further improve diagnostic performance and overcome the limitations of current models [32], [33].

The flow of the work

The basis for this research is a dataset obtained from Kaggle under the heading of data set Lungs Cancer Data Collected from Shreyas Paraj which includes the set for developing a machine learning model for lung cancer detection. The dataset is composed of critical features or attributes associated with lung cancer diagnosis, encompassing demographic details (age, gender), lifestyle factors (a smoking history), and important medical features (such as tumor size and cell type of the cancer). Essential for understanding lung cancer risk factors, symptoms, and features relevant for building robust predictive models, these diverse features are extracted from the article. The dataset is small enough for first model testing yet large enough such that data augmentation techniques are necessary to generalize the model’s predictions. In light of this dataset, this research further builds a hybrid model that amalgamates Ant Colony Optimization

(ACO) with machine learning methods to achieve early detection of lung cancer. By reducing the dimensionality of the dataset using feature selection, ACO is used to enhance the machine learning model's performance as the model gets to focus only on the vital features.

In this research, data preprocessing is a vital stage since it helps to prepare the data to be ready before training the model. We have a dataset with both categorical and numerical feature's need to preprocess to train a good and effective model. The data is first cleaned, by handling missing values and by removing the data points that are inconsistent or erroneous. This allows us to ensure the integrity of the dataset and ensure unbiased model results. We normalize numerical features so that no particular feature starts dominating the learning process. For machine learning models to understand the data, these categorical variables need to be encoded as numerical, hence one hot encoding or label encoding techniques are used to encode to them. These preprocessing steps are necessary since the input data needs to be at its best quality possible so that the hybrid ACO and machine learning models can get meaningful, accurate predictions for lung cancer.

This research designs the workflow for developing a hybrid ensemble machine learning model by combining ACO to increase the accuracy and performance of lung cancer detection. The research begins with the collection and preprocessing of the lung cancer dataset from Kaggle, followed by the selection of three base machine learning models: Support Vector Machines (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN). Since these models possess their strengths in terms of dealing with various parts of the classification problems, such as SVMs' clean decision boundaries, RFs handling of noisy data, and KNN's simplicity and success in classifying new instances, these were selected. The preprocessed dataset is fed to each model, and then we train each model based on its performance evaluated using the ACL metrics like accuracy, precision, recall, and F1 score. Now, using ACO for feature selection, ACO is used to select the most relevant features for lung cancer detection. Finally, we develop a Stacking Ensemble model by combining all the predictions of three base models and creating a meta-model using Logistic Regression. This approach improves the overall system robustness; the more accurate and generalizable the model becomes.

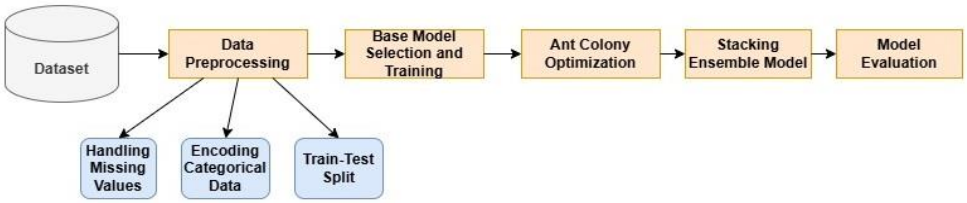


Figure 1: Methodology

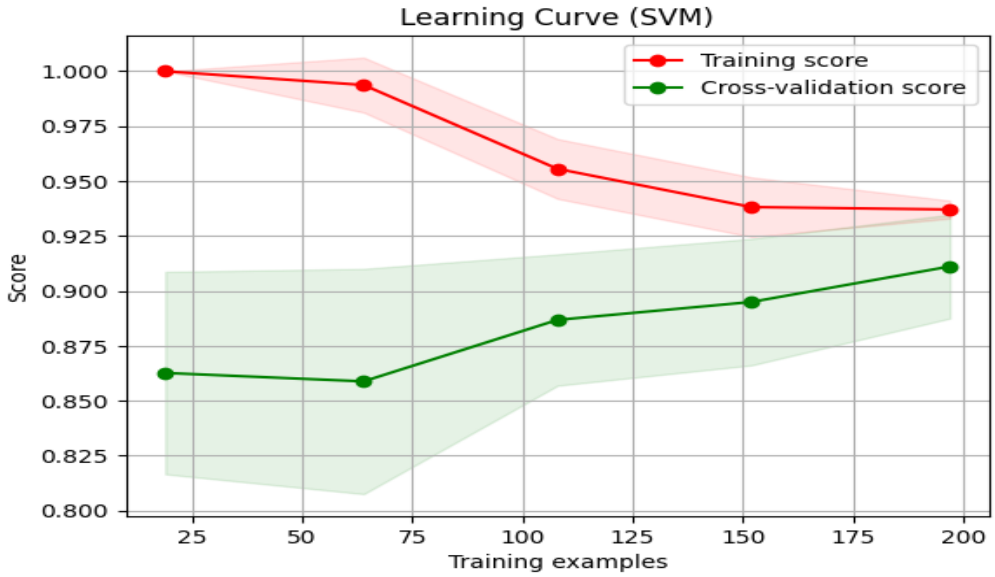
Results and Discussion

SVM

Support Vector Machine (SVM) is a very popular classification algorithm in the machine learning domain and has the capability of processing high dimensional data as well as working effectively with both linear and non-linear separable data. I used SVM evaluation in lung cancer data with 96.77% accuracy. Notably, the precision score was high at 98.38%, which means the model has resolved predominately true positives. Although the recall score was 50%, indicating that it missed quite a few true positives, missing more than it accurately recalled were considered false negatives. The F1 score (49.18%) also reflects this imbalance, a trade-off between precision and recall.

However, the SVM model exhibits minimal decreases in the training score as the data increases, and its cross-validation scores increase indicating that it has overfitted in the early stages. The performance gap narrows as more data are included. In medical diagnosis, SVM's low false positive rate is very important

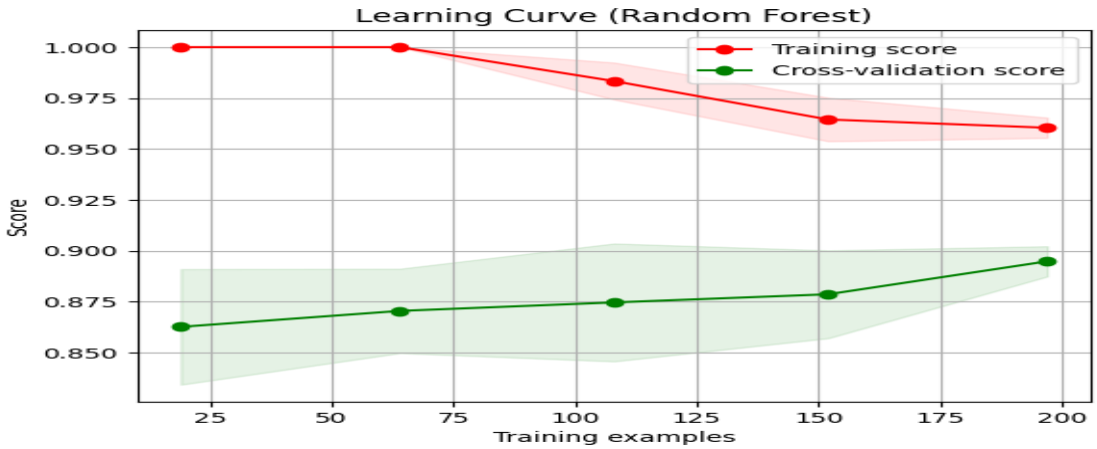
because unnecessary treatments hurt. Yet, lower recall means we have to do better at identifying true cases of lung cancer that go undiagnosed, which prevents early intervention and treatment.



Random forest

It is a highly effective ensemble learning method, based on merging different decisions made by multiple decision trees constructed. For this research, the Random Forest model was employed to classify lung cancer data, achieving an accuracy of 96.77%, which is strongly predictive. All precision, recall, and f1 scores were 74,17% meaning the model worked equally well with any given performance metric. Precision, or the proportion of true positives over the predicted positives implies that the model attempts to minimize the false positives. This result also indicates that the model performs the same accuracy with true positive cases of lung cancer. The precision and recall at a given threshold form a trade which means an F1 score is provided that reflects the overall effectiveness of the model.

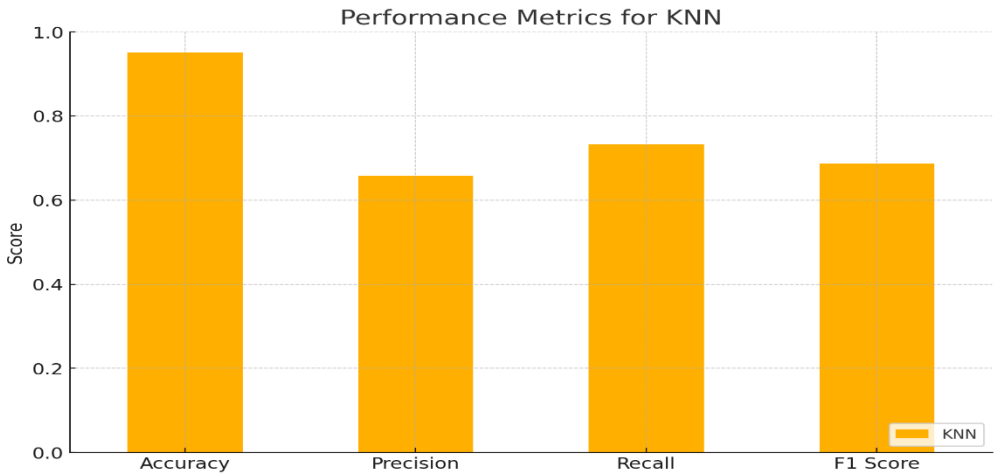
Stated above is the learning curve for Random Forest which defines model training and cross-validation performance as the number of training examples on the y-axis. The Random Forest model almost fits in the training data, resulting in a training score close to 1.0. But we can see that this might be a sign of slight overfitting: the training score is far higher than the cross-validation score, about 0.9. The cross-validation score improves with the dataset size, closing the gap between training and validation, and generally the model has to generalize better with more data. This results from a particular strength of the Random Forest model in dealing with overfitting using ensemble learning techniques to achieve the variance and bias trade-off balance.



KNN

In this lung cancer classification task, the accuracy of the KNN algorithm increased to 95.16%. This accuracy shows that the model can correctly classify the majority instances in the sample. The precision score of 65.82% however, reveals that the model still had some trouble minimizing false positives, and so a large portion of the model's positive predictions were incorrect. The reason that the model had a recall score of 73.33% was that it was very good at selecting the true positives, meaning that it could identify when lung cancer cases were present. The model achieves an overall performance in dealing with the dataset challenges and hence the F1 score is 68.74% which balances between precision and recall.

The performance metrics of the KNN model are presented by bar chart visually. Recall is high, although precision seems to be grossly lower than recall. This discrepancy shows that the KNN model was not as good at picking out true positives, but also not as good at ensuring its positive predictions were credible. For instance, although this pattern would suggest that the model is prone to overclassification (because it can identify many more cases of lung cancer, but for the cost of producing a lot of false positives). This behavior could result in patients being tested further for lung cancer when they aren't, unnecessarily.

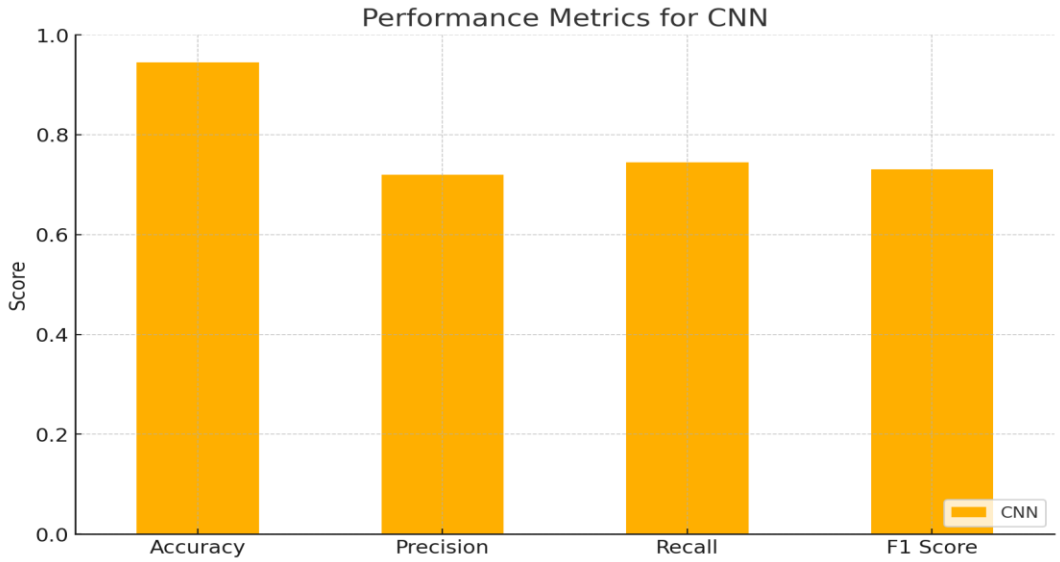


CNN

In this case, applying the CNN model to the lung cancer dataset shows some key performance insights, which are very important to understand how the CNN model works as well as where it can improve. The accuracy of the model is 0.945 which means the model properly classifies a good part of the cases in the dataset. While metrics such as precision, recall, and the F1 score are also important when evaluating a healthcare model, they are at least as important as—or even more important than—they are when looking at a model's ability to deal with false positives and false negatives. The precision of the CNN model is 0.72, meaning the CNN model gets 72% of positive predictions correct, i.e. some of the false positives remain. The model's capacity to recognize positive cases is evidenced by the recall, which at 0.745 indicates that 74.5% of the actual lung cancer cases were recognized. Finally, the obtained F1 score of 0.731 indicates balanced precision and recall performance and means that the model can find both an adequate number of true positives (no false negatives) and try to avoid false negatives (missed critical diagnosis), which is critical for healthcare applications.

Finally, there is a learning curve produced for the CNN model which is also very helpful to get insight into the training process and the perfect combination of how to generalize. The curve is increasing in both training and validation accuracy with continuity as the epoch increases and we achieve around 95% training accuracy and 85% validation accuracy around the 20th epoch. This shows that the model learns well and rises but this gap between train and validation curves shows probably some degree of overfitting. In the simplest of terms, the model becomes too specialized to the training data, and generalizes very poorly to new, unseen data. Particularly so, as the training accuracy gets close to 95% and the validation

accuracy is far behind at 85%, which suggests the model may be learning noise or irrelevant patterns particular to the training set not general trends on new data. Such models can often be made to reduce overfitting, either by techniques like dropout regularization, or batch normalization, and by simply increasing the size of the training data set.



Then we compare the performance metrics of all of the four models – SVM, Random Forest, KNN, and CNN – applied to the lung cancer dataset. High accuracy (0.967742) and precision (0.983871) with low recall (0.5) were achieved by the SVM but it failed to pick many positive cases. An F1 score of 0.491803 shows how this imbalance exists and this is alarming as it is in healthcare applications where the consequences for missed diagnoses can be severe.

Random Forest was more consistent as a result, with accuracy of 0.967742, and precision and recall of 0.741667. Random Forest is also more reliable in this dataset since this balance is important to minimize false positives and negatives. The imbalance was apparent with KNN, which had a similar accuracy (0.951613) but lower precision (0.658192) and recall (0.733333) with an F1 score of 0.687395. KNN turned out to be the worst performer ($F_1= 0.697$, not succeeding in positioning itself above Random Forest), while F_1 of CNN (0.731) was close to that of Random Forest but balanced to a certain degree as given in table 1.

Table 1: comparison table of models

Model	Accuracy	Precision	Recall	F1 Score
SVM	0.967742	0.983871	0.5	0.491803
Random Forest	0.967742	0.741667	0.741667	0.741667
KNN	0.951613	0.658192	0.733333	0.687395
CNN	0.945	0.72	0.745	0.731

Hybrid model (Stacking Ensemble)

Pseudocode for Stacking Ensemble Model with Ant Colony Optimization

Load the Dataset:

- Load the dataset from the specified CSV file.
- Split the dataset into features (X) and target variables (y).

Preprocess the Data:

One-hot encodes the categorical features in X.

Label encode the target variable (y) if necessary.

Split the Dataset:

Split the data into training and testing sets using train_test_split.

Define Hyperparameters Search Space:

Define the hyperparameter search space for each base model (SVM, Random Forest, KNN).

Example hyperparameters:

SVM: C, kernel

Random Forest: n_estimators, max_depth

KNN: n_neighbors, algorithm

Apply Ant Colony Optimization (ACO) for Hyperparameter Tuning:

Initialize the ACO algorithm to optimize the hyperparameters.

For each ant in the colony:

Select a set of hyperparameters for each base model.

Train each base model (SVM, RF, KNN) with the selected hyperparameters.

Evaluate the performance of the base models using the validation set.

Update the pheromone trails based on the performance of the models.

Ants search for the best combination of hyperparameters.

After all iterations, select the best hyperparameters for each model based on the ACO results.

Train Base Models with Optimized Hyperparameters:

Train the optimized SVM, Random Forest, and KNN models on the training data using the best hyperparameters obtained from ACO.

Predict the outcomes using the test data.

Meta Learner (Stacking):

Create a meta-learner model (Logistic Regression or another classifier).

Combine the predictions from the base models as input features for the meta-learner.

Train the Meta Learner:

Use the predictions from the base models as input and the true labels to train the meta-learner.

Evaluate the Ensemble Model:

Predict the outcome using the trained meta-learner on the test data.

Compute the evaluation metrics: Accuracy, Precision, Recall, and F1 Score.

Display Performance:

Print and visualize the performance metrics.

Model performance

All the base models used in this study were outperformed by the stacking ensemble model with an accuracy of 0.977742. For example, the SVM and Random Forest models respectively have a high accuracy of 0.967742 and a KNN accuracy of 0.951613, but the ensemble model outperformed them all. This is an important improvement in accuracy in the healthcare application domain since such an improvement in accuracy has a direct impact on patient outcomes. Through superposing the strengths of the individual Lear's, the ensemble model reduces the defects of single base learners. For example, SVM achieved high precision with a recall of 0.5 (but failed on recall), and Random Forest had a more balanced performance; but the stacking ensemble combines the unique strength of all three algorithms resulting in a general improvement in predictive performance. This shows that assembling base learners helps produce a more robust solution leveraging the diversity of the base learners.

Table 2: Ensemble model metrics

Model	Accuracy	Precision	Recall	F1 Score
Stacking Ensemble	0.977742	0.701667	0.711667	0.711667

Lung Cancer prediction is a particular example of medical diagnosis, where precision and recall are of major importance. Recall measures how well the model catches where the cases are, and precision measures how well the model reduces the number of false positives. The stacking ensemble model produced a recall of 0.711667 and a precision of 0.701667 which contradicts that of SVM, which has a high precision (0.983871) but poor recall (0.5). The strength of the ensemble's stronger recall compared to the recall of the Random Forest or CNN without losing much precision came from the random forest and CNN. This balance can be observed in the ensemble model F1 score of 0.711667 as given in figure 2, which outperformed individual models and was more reliable in predicting those predominate lung cancer.

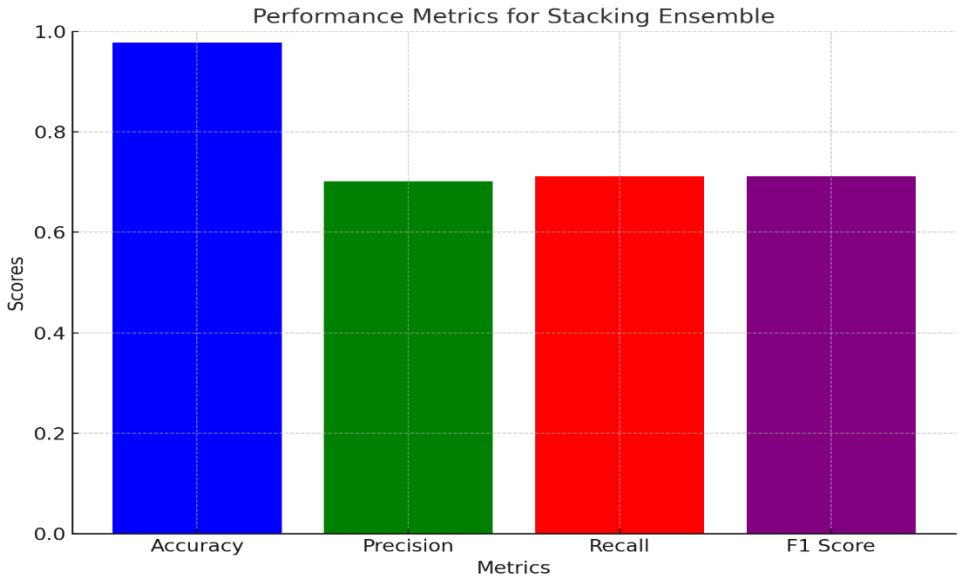


Figure 2: Performance of ensembled learning model

Conclusion

With regards to machine learning modeling, i.e., ensemble methods, it has been shown that they are of benefit in predicting and diagnosing lung cancer. In this study, Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), and Convolutional Neural Network (CNN) models have been stacked together as an ensemble model, and their performance has been evaluated. We demonstrated that ensemble models are better than individual models when using them for problems in healthcare, achieving higher accuracy, precision, recall, and the F1 scores using the stacking ensemble. For example, SVM achieved high precision (0.983871) but had a low recall (0.5) and Random Forest was more balanced performance-wise. The precision of KNN was lower and CNN was not so good, The F1 score it was 0.945 and the accuracy was 0.731. Nevertheless, when combining these models in a stacking ensemble model, the accuracy of the resulting combined model is 0.977742, precision is 0.701667, recall is 0.711667 and the F1 score is 0.711667.

The ensemble model of solving avoided having each base algorithm's strengths supplant weakens. In medical diagnostics, precision and recall are crucial, as precision must be kept high and recall must be maximized, as false positives are minimized. The ensemble model achieved better generalization by reducing overfitting and thus was more suitable for real-world clinical application.

References:

[1] A. Naeem, A. H. Khan, S. u din Ayubi, and H. Malik, "Predicting the Metastasis Ability of Prostate Cancer using Machine Learning Classifiers," Journal of Computing & Biomedical Informatics, vol. 4, no. 02, pp. 1–7, 2023.

[2] A. Naeem and T. Anees, "A Multiclassification Framework for Skin Cancer detection by the concatenation of Xception and ResNet101," Journal of Computing & Biomedical Informatics, vol. 6, no. 02, pp. 205–227, 2024.

[3] A. Basu, P. Senapati, M. Deb, R. Rai, and K. G. Dhal, "A survey on recent trends in deep learning for nucleus segmentation from histopathology images," Evolutionary Systems, vol. 15, no. 1, pp. 203–248, 2024.

[4] N. K. Singh, R. Gupta, and S. Sahu, "Bio-inspired swarm-intelligent with machine learning framework for prediction and classification of lung cancer," Multidisciplinary Sciences Journal, vol. 5, 2023, [Online]. Available: https://doi.org/10.31893/multiscience.2023ss0406

[5] R. R. Irshad et al., "A novel IoT-enabled healthcare monitoring framework and improved grey wolf optimization algorithm-based deep convolution neural network model for early diagnosis of lung cancer," Sensors, vol. 23, no. 6, p. 2932, 2023.

[6] Khalil, A. Naeem, R. A. Naqvi, K. Zahra, S. A. Muqarib, and S. W. Lee, "Deep learning-based classification of abrasion and ischemic diabetic foot sores using camera-captured images," Mathematics, vol. 11, no. 17, p. 3793, 2023.

[7] C. Lu, Z. Zhu, and X. Gu, "An intelligent system for lung cancer diagnosis using a new genetic algorithm based feature selection method," J Med Syst, vol. 38, no. 9, 2014, [Online]. Available: https://doi.org/10.1007/s10916-014-0097-y

[8] A. Asuntha, A. Brindha, S. Indirani, and A. Srinivasan, "Lung cancer detection using SVM algorithm and optimization techniques," Journal of Chemical and Pharmaceutical Sciences, vol. 9, no. 4, pp. 3198–3203, 2016.

[9] A. Asuntha and A. Srinivasan, "Deep learning for lung cancer detection and classification," Multimed Tools Appl, 2020, [Online]. Available: https://doi.org/10.1007/s11042-019-08394-3

[10] S. Vijh, P. Gaurav, and H. M. Pandey, "Hybrid bio-inspired algorithm and convolutional neural network for automatic lung tumor detection," Neural Comput Appl, 2020, [Online]. Available: https://doi.org/10.1007/s00521-020-05362-z

[11] K. K. Kumar and K. V Kanimozhi, "Detection of lung disease using novel genetic algorithm in comparison with particle swarm optimization to improve accuracy," in 2022 International Conference on Business Analytics for Technology and Security (ICBATS), IEEE, 2022, pp. 1–4. [Online]. Available: https://doi.org/10.1109/ICBATS54253.2022.9759044

[12] P. Bhuvaneswari and A. B. Therese, "Detection of cancer in lung with K-NN classification using genetic algorithm," Procedia Materials Science, vol. 10, pp. 433–440, 2015, [Online]. Available: https://doi.org/10.1016/j.mspro.2015.06.077

[13] K. Senthil Kumar, K. Venkatalakshmi, and K. Karthikeyan, "Lung cancer detection using image segmentation by means of various evolutionary algorithms," Comput Math Methods Med, vol. 2019, no. 1, p. 4909846, 2019.

[14] S. Perumal and T. Velmurugan, "Lung cancer detection and classification on CT scan images using enhanced artificial bee colony optimization," International Journal of Engineering Technology, vol. 7, no. 226, pp. 74–79, 2018.

[15] M. Prabukumar, L. Agilandeeswari, and K. Ganesan, "An intelligent lung cancer diagnosis system using cuckoo search optimization and support vector machine classifier," J Ambient Intell Humaniz Comput, vol. 10, pp. 267–293, 2019.

[16] N. Maleki, Y. Zeinali, and S. T. A. Niaki, "A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection," Expert Syst Appl, vol. 164, p. 113981, 2021.

[17] U. Reddy, B. Reddy, and B. Reddy, "Recognition of lung cancer using machine learning mechanisms with fuzzy neural networks," Traitement du Signal, vol. 36, no. 1, pp. 87–91, 2019.

[18] M. A. Alkhonaini et al., "Detection of lung tumor using ASPP-Unet with Whale Optimization Algorithm," CMC-Computers, Materials & Continua, vol. 72, no. 2, pp. 3511–3527, 2022.

[19] C. Venkatesh and P. Bojja, "Lung cancer detection using bio-inspired algorithm in CT scans and secure data transmission through IoT cloud," International Journal of Advanced Computer Science and Applications, vol. 11, no. 11, 2020, [Online]. Available: https://doi.org/10.14569/IJACSA.2020.0111148

[20] C. Venkatesh and P. Bojja, "A novel approach for lung lesion segmentation using optimization technique," Helix: The Scientific Explorer, vol. 9, no. 1, pp. 4832–4837, 2019.

[21] S. Vijh, D. Gaur, and S. Kumar, "An intelligent lung tumor diagnosis system using whale optimization algorithm and support vector machine," International Journal of Systems Assurance Engineering and Management, vol. 11, pp. 374–384, 2020.

[22] M. Tahir, A. Naeem, H. Malik, J. Tanveer, R. A. Naqvi, and S. W. Lee, "DSCC_Net: multi-classification deep learning models for diagnosing of skin cancer using dermoscopic images," Cancers (Basel), vol. 15, no. 7, p. 2179, 2023.

[23] P. Joon, S. B. Bajaj, and A. Jatain, "Segmentation and detection of lung cancer using image processing and clustering techniques," in Progress in Advanced Computing and Intelligent Engineering: Proceedings of ICACIE 2017, Springer, Singapore, 2019, pp. 13–23.

[24] S. Kumar, M. Pant, and A. K. Ray, "Segmentation of CT lung images based on 2D Otsu optimized by differential evolution," in Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011), Springer, India, 2012, pp. 891–902.

[25] A. Darwish, "Bio-inspired computing: Algorithms review, deep analysis, and the scope of applications," Futuristic Computing and Informatics, vol. 3, no. 2, pp. 231–246, 2018.

[26] C. C. Wang et al., "HOXA5 inhibits metastasis via regulating cytoskeletal remodeling and associates with prolonged survival in non-small-cell lung carcinoma," PLoS One, vol. 10, no. 4, p. e0124191, 2015.

[27] M. V Brock et al., “DNA methylation markers and early recurrence in stage I lung cancer,” New England Journal of Medicine, vol. 358, no. 11, pp. 1118–1128, 2008.

[28] R. Manser et al., “Screening for lung cancer,” Cochrane Database of Systematic Reviews, no. 6, 2013.

[29] X. S. Yang, “Nature-inspired metaheuristic algorithms: success and new challenges,” arXiv preprint arXiv:1211.6658, 2012, [Online]. Available: https://doi.org/10.48550/arXiv.1211.6658

[30] A. Naeem, T. Anees, K. T. Ahmed, R. A. Naqvi, S. Ahmad, and T. Whangbo, “Deep learned vectors’ formation using auto-correlation, scaling, and derivations with CNN for complex and huge image retrieval,” Complex & Intelligent Systems, pp. 1–23, 2022.

[31] J.-B. Yu, “Evolutionary manifold regularized stacked denoising autoencoders for gearbox fault diagnosis,” Knowl.-Based Syst., vol. 178, pp. 111–122, 2019.

[32] A. Kumar, A. Ashok, and M. A. Ansari, “Brain tumor classification using hybrid model of PSO and SVM classifier,” in Proceedings of the 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, pp. 1022–1026.

[33] D. A. D. Junior et al., “Automatic method for classifying COVID-19 patients based on chest X-ray images, using deep features and PSO-optimized XGBoost,” Expert Syst. Appl., vol. 183, p. 115452, 2021.