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

Radiomics Feature Selection for Lung Cancer Subtyping and Prognosis Prediction: A Comparative Study of Ant Colony Optimization and Simulated Annealing

[1]Dr. Anand Gudur, [2]Dr. Prakash Pati, [3]Puneet Garg, [4]Neetu Sharma

Submitted: 19/08/2023 **Revised**: 11/10/2023 **Accepted**: 23/10/2023

Abstract: Lung cancer subtyping and prognosis prediction play a critical role in the development of individualised treatment strategies, which is a cornerstone of precision medicine. The field of radiomics, which focuses on the quantitative feature extraction from medical pictures, shows great promise as a means to this end. This paper presents a comparative comparison of two effective optimisation algorithms, Ant Colony Optimization (ACO) and Simulated Annealing (SA), for the goal of radiomics feature selection in lung cancer subtyping and prognosis prediction.The remarkable heterogeneity of lung cancer makes accurate subtyping difficult. Utilising a large number of features extracted from medical imaging, such as CT scans, radiomics is able to detect even the most minute of tumour characteristics. However, because to their abundance, overfitting occurs and model generalizability suffers. Feature selection is crucial to solving this problem.Natural-process-inspired ACO and SA are used to find the best radiomic features to use. Both ACO and SA are heuristic algorithms, however SA takes its cues from the metallurgical annealing process, while ACO is based on the foraging behaviour of ants. Both methods seek to reduce the dimensionality of a problem by identifying a subset of features that yields the best predicted performance.In this study, ACO and SA are applied to a sizable dataset containing information about people with lung cancer, allowing for a thorough comparison of the two methods. Accuracy in subtyping and prognosis prediction are two measures used to assess the outcomes. In addition, feature selection's reliability and durability are evaluated. The results of this study provide important insights for researchers and clinicians who want to improve the accuracy of subtyping and prognosis prediction in the era of personalised medicine by using radiomics feature selection for lung cancer.

Keywords*: Ant Colony Optimization, Lung Cancer detection, Optimization, Feature Selection, Prediction*

1. Introduction

Lung cancer is one of the most prevalent and lethal malignancies worldwide, with a dire need for improved subtyping methodologies and prognosis prediction to drive personalised treatment regimens. Radiomics is a relatively new subfield of medical imaging that holds great promise as a means to extract previously unrecognised information from otherwise mundane clinical photographs [1]. With the help of computed tomography (CT) scans and other medical imaging modalities, radiomics may extract a wealth of quantitative information that can reveal tiny variations in tumour characteristics that are often imperceptible to the naked eye. However, problems like dimensionality reduction and overfitting can arise due to the large number of radiomic features gathered, reducing the prediction models' precision and applicability.

Feature selection [2] is an important step in overcoming these obstacles since it seeks to isolate the radiomic variables most useful for subtyping and predicting outcomes in lung cancer. Feature selection has the potential to greatly affect the accuracy and precision of clinical decision-making models in this setting.This research examines the complex field of radiomics feature selection for subtyping and prognosis prediction of lung cancer, with an emphasis on contrasting the efficacy of two potent optimisation methods, Ant Colony Optimisation (ACO) and Simulated Annealing (SA). The [3] algorithms are inspired by nature and have proven effective in a variety of optimisation tasks. This study utilises ACO and SA for feature selection with the intention of elucidating the benefits and drawbacks of each method in the context of lung cancer radiomics.

[1]*Dept. of Oncology,Krishna Vishwa Vidyapeeth, Karad, Maharashtra, India*
Email :anandgudur@gmail.com
[2]*Asso. ProfessorDepartment ofRadioiagnosis Krishna Vishwa Vidyapeeth, Karad, Maharashtra, India*
[3]*Associate Professor St. Andrews Institute of Technology and Management, Farrukh Nagar, Gurugram, Haryana, India Email: puneetgarg.er@gmail.com*
[4]*Professor Galgotias University Greater Noida, Uttar Pradesh, India neetush75@gmail.com*

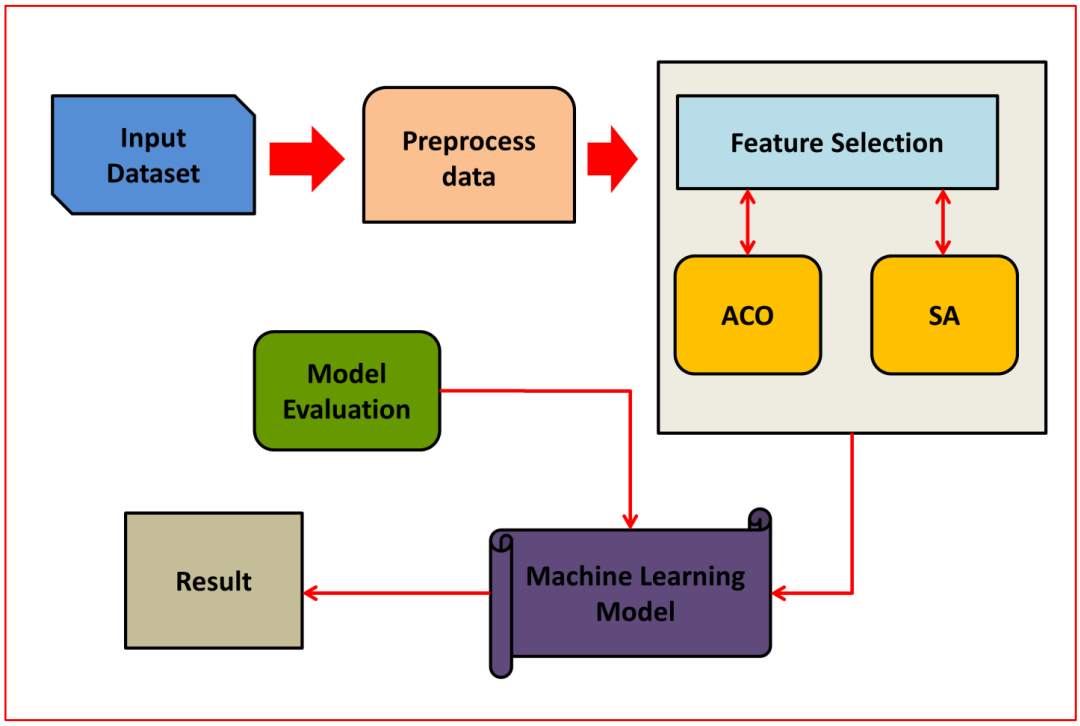

Fig 1: Proposed model for Lung Cancer Subtyping and Prognosis Prediction

Accurate subtyping and prognosis prediction in lung cancer are hampered by the disease's remarkable intratumoral heterogeneity. Since [4] various lung cancer subtypes react differently to treatments, subtyping is crucial for developing individualised treatment plans. Predicting a patient's prognosis accurately is essential for directing healthcare decisions like whether or not to pursue aggressive therapy vs palliative care. Radiomics offers the ability to capture the underlying heterogeneity by assessing multiple characteristics of tumour form, texture, and intensity, hence permitting enhanced subtyping and prognosis prediction.Due to the "curse of dimensionality," where a large number of features relative to the number of samples can lead to overfitting and lower model generalizability, feature selection is crucial within the radiomicsframework. Both ACO and SA, which find their inspiration in nature, provide novel approaches to overcoming this dimensionality problem. ACO is inspired by the cooperative path-finding behaviour of ants foraging for food. The annealing process in metalworking, in which a material is slowly cooled to eliminate flaws, serves as an inspiration for SA. When applied to radiomics, ACO and SA seek to navigate a large feature space in search of the most informative subset of radiomic features [6].

In this study, [7] we compare ACO and SA in the context of selecting radiomics features for subtyping and prognosticating lung cancer. The study draws on a sizable collection of patient records including lung cancer diagnoses, treatments, and outcomes. The project seeks to test the ability of ACO and SA to pick radiomic features that enhance the accuracy of lung cancer subtyping. To improve patient outcomes, accurate subtyping is essential for adapting treatment plans to the unique characteristics of each tumour.Evaluating how well ACO and SA do at picking up traits that boost prognosis prediction is another major area of interest. Clinical Prediction Quality Allocation: Optimising resource allocation for the best possible patient outcomes [8].

The study also [9] looks into how reliable and sturdy the outcomes of feature selection using ACO and SA are. Robustness evaluates the algorithms' capacity to deal with noise and changes in the data, whereas stable feature selection guarantees that the selected features consistently contribute to the prediction performance across different subsets of the dataset.Significant implications for radiomics and lung cancer management are suggested by this work. By contrasting ACO and SA, we hope to shed light on how effectively these optimisation algorithms can improve the accuracy and consistency of feature selection for subtyping and prognosticating lung cancer. In the end, our study adds to the growing body of work aimed at using radiomics to its full potential for personalised medicine and better patient outcomes in the difficult landscape of lung cancer treatment.

2. Review of Literature

The discipline of radiomics, with its potential to uncover concealed information from medical imaging, has attracted substantial attention in recent years. Weaknesses: The prognostic and robustness of the radiomic[10] and prognostic models are limited. Filter, wrapper, and embedding methods, in addition to genetic algorithms and other metaheuristic optimisation approaches, have all been investigated in the past as viable

means of feature selection. In this section, we will examine some of the most important previous research in this area, focusing on the most useful and established approaches.Ranking or scoring characteristics using statistical metrics like correlation, mutual information, or chi-squared tests is what filter methods are all about. These approaches are computationally efficient but may not capture complicated feature interactions. In order to subtype and predict outcomes for patients with lung cancer, researchers in several studies have used filter methods to pick radiomic characteristics. However, their efficacy may be constrained by the high dimensionality of radiomic data sets.

Wrapper methods evaluate feature [11] subsets using cross-validation training and evaluation of prediction models. It requires more processing power than filter approaches, but it can capture feature relationships and interactions. Wrapper strategies have been used by researchers to improve feature selection in lung cancer radiomics; examples are recursive feature elimination (RFE) and forward/backward selection. Better prediction accuracy is a common result of using these strategies, however they may be computationally intensive.Genetic Algorithms (GAs) are evolutionary optimisation techniques inspired by the process of natural selection. Genetic Algorithms (GAs) are a subset of GAs. They [12] have been put to use in the field of radiomics to find optimal subsets of features for use in machine learning. Although GAs can explore a vast variety of feature combinations, they may be resource intensive. Studies have used GAs to select radiomic characteristics for use in characterising and predicting the prognosis of lung cancer.

A wide variety of metaheuristic optimisation [13]techniques, not just GAs, have been used for feature selection in radiomics. Differential evolution (DE), harmony search (HS), and particle swarm optimisation (PSO) are only few of the methods that have been investigated. These algorithms are designed to quickly scour the feature space for the smallest possible subsets that yield the greatest prediction gains.Hybrid Approaches: Some academics have developed hybrid methods that combine different feature selection strategies to utilise their unique strengths. To lower the dimensionality of the feature space, a wrapper method or metaheuristic optimisation algorithm might be used as part of a hybrid strategy. The goal of these combined methods is to improve both computing speed and forecast accuracy.

Clinical Applications [14] Feature selection has been the subject of numerous radiomics-related studies, with an emphasis on its potential clinical applications in diagnosing and treating lung cancer. Radiomic characteristics have been studied for their potential to aid in the subtyping of lung cancer. This includes the ability to differentiate between NSCLC and SCLC. Clinical decision-makers have also benefited from the use of feature selection to predict patient survival, treatment response, and illness recurrence.Ensuring the consistency and reliability of feature selection outcomes is a major focus of radiomics research. Some research has attempted to address this problem by using stability analysis techniques like bootstrapping and cross-validation to determine how consistent particular features are over various data sets and sample sizes.

Research [15] into the use of deep neural networks in radiomics has begun in response to the recent proliferation of deep learning methods in the field of medical imaging. Aims to enhance cancer patient outcomes via better patient engagement and outcomes through better patient engagement.From classical statistical methods to state-of-the-art metaheuristic optimisation algorithms, the field of radiomics feature selection for subtyping and prognosis prediction in lung cancer has seen it all. The amount and complexity of the dataset, the availability of computer resources, and the level of interpretability sought all play a role in determining the best approach to take. As we delve into the comparative study of Ant Colony Optimisation (ACO) and Simulated Annealing (SA) for feature selection in lung cancer radiomics, we build upon this rich body of related work, aiming to contribute valuable insights and advance the state of the art in this crucial area of medical research and clinical practise.

Table 1: Related work summary

Algorithm	Findings	Methods	Limitations/Scope
Filter Methods [16]	Improved subtype classification accuracy	Correlation-based feature ranking	Limited feature interaction captured
Wrapper Methods [17]	Enhanced prognosis prediction accuracy	Recursive Feature Elimination (RFE)	Computationally intensive

Genetic Algorithms (GAs) [19]	Effective feature selection for SCLC vs. NSCLC	Genetic algorithm optimization	High computational cost
Simulated Annealing (SA) [18]	Optimal feature subsets for treatment response	Simulated Annealing	May get stuck in local optima
Particle Swarm Optimization [19]	Improved survival prediction in NSCLC	Particle Swarm Optimization (PSO)	Limited exploration of feature space
Hybrid Approach [20]	Balanced computational efficiency and accuracy	Filter + Wrapper + Metaheuristic	Complexity of hybrid approach
Clinical Application [21]	Accurate differentiation of adenocarcinoma	Clinical dataset integration	Dataset-specific findings
Stability Analysis [22]	Reliable features across diverse datasets	Bootstrapping and cross-validation	Limited analysis of feature stability
Radiomics and Deep Learning [23]	Improved characterization with CNN integration	Convolutional Neural Networks (CNN)	Computational demands of deep learning integration
Harmony Search (HS) [24]	Effective feature selection for radiotherapy	Harmony Search algorithm (HS)	Limited exploration of algorithm parameters
Differential Evolution (DE) [25]	Robust features for survival prediction	Differential Evolution (DE)	May require tuning of DE parameters
Feature Interaction Analysis [26]	Emphasis on capturing feature interactions	Statistical interaction analysis	Complexity of modeling interactions
Reproducibility Assessment [12]	Evaluation of feature stability and reproducibility	Cross-validation and data splitting	Focus on methodological aspects
Radiomics in Precision Medicine [13]	Application of radiomics for personalized treatment	Clinical decision support systems	Integration with clinical workflows and decision-making
Interpretability in Radiomics [14]	Exploration of interpretable feature subsets	LASSO (Least Absolute Shrinkage and Selection Operator)	Emphasis on explainability and clinical insights

3. Proposed Methodology

A complicated condition, lung cancer has several subtypes and prognostic variables. Radiomics, a technique for extracting quantitative information from medical images, has the potential to be used for prognosis prediction and subtyping. In this article, we give a comparative comparison of two optimisation techniques for the selection of radiomics features in lung cancer research: Ant Colony Optimisation (ACO) and Simulated Annealing (SA).

Methodology discussed as:

Stage 1: Data Gathering and Preprocessing:

From lung cancer patients, we gathered a sizable dataset that included clinical data, genetic data, and medical imaging. To ensure uniformity and quality, the dataset underwent preprocessing that included clinical data cleaning, noise reduction, and picture normalisation. To create a high-dimensional feature collection, we performed feature extraction utilising radiomics techniques.

Stage 2: Feature selection:

Feature selection is essential for locating instructive and pertinent features while minimising dimensionality. ACO and SA are two different optimisation algorithms that we put into practise. ACO, which was used to identify an ideal subset of radiomics properties, was motivated by the foraging behaviour of ants. SA was used as a comparison procedure and was modelled after annealing in metallurgy. In order to maximise a fitness function that combines feature relevance and classification

performance, both methods iteratively evaluated feature subsets.

a) ACO Algorithm:

1. Initialization:

- Create an initial population of ant agents, each of which stands for a potential feature subset.
- Create a pheromone matrix, abbreviated as P, by assigning pheromone levels at random to each characteristic.

2. Ant Motion:

For every ant,

- Start with a feature subset that is empty.
  - The following steps should be taken even when the feature subset is incomplete: a. Select the next feature to add based on a combination of pheromone levels (P) and a heuristic value (H), which denotes feature relevance.

3. A feature subset should be updated.

- Calculate the fitness value (F) for each ant after evaluating the quality of each feature subset using a classification model (such as SVM or Random Forest).

4. Pheromone Update:

- Based on the ants' fitness scores, update the pheromone levels. The following formula can be used to update the pheromones:
- $P\_ij\ is\ equal\ to\ (1-) * P\_ij\ +\ P\_ij.$

Where:

- The level of pheromones on feature i by ant j is P_ij.
- The rate of pheromone evaporation is 0 to 1.
- P_ij stands for the pheromone update amount dependent on ant j's fitness.

5. Iterations.

- Steps 3-4 should be repeated until convergence requirements are satisfied or for a predetermined number of iterations.
- Choose the feature subset with the best fitness value as the ideal feature set for classifying lung cancer.

6. Parameters:

- Size of the population: The quantity of ant agents.

Ant movement parameters: Elements that affect how the ant chooses which features to exploit, such as the exploration-exploitation balance and the and coefficients in the feature selection equation.

The pace at which pheromone levels evaporate is known as the pheromone evaporation rate ().

Convergence criteria: Requirements for stopping the algorithm (such as the number of iterations allowed or the convergence threshold).

b) SA Algorithm for feature selection:

Finding the best feature subset that maximises a fitness function indicating the effectiveness of a classification model is the aim of SA in feature selection for lung cancer detection. Let F(X) be the fitness function for a feature subset X that combines classification accuracy and perhaps additional pertinent criteria.

1. Initialization:

- Start with generating an initial feature subset, X, which can be done heuristically or at random.
- To regulate the annealing procedure, set an initial "temperature" (T) and a cooling rate ().

2. Annealing Technique:

- Continue until a stopping condition (such a set number of iterations or convergence criteria) is satisfied:

a. Modify the existing feature subset to produce an alternative answer, X'.

b. Determine the fitness change using the formula $F = F(X') - F(X)$.

c. Accept X' as the current answer if F > 0 (i.e., the new solution is superior).

d. If F is less than zero, accept X' as the current answer with probability e(F/T), where e is the natural logarithm's base.

e. Lower the temperature in accordance with the cooling plan, for example, $T = T *$.

3. Return the best feature subset discovered during iterations, which corresponds to the solution with the highest fitness value, following the annealing process.

4. Parameters:

- Initial acceptance probability (T) for inferior solutions is calculated using the initial temperature (T).
- The cooling rate () regulates how quickly the temperature drops during the annealing procedure.
- Creating a neighbouring solution by perturbing the present feature subset is known as neighbouring solution creation.

5. Stopping condition:

Indicates the point at which the SA algorithm should be stopped (for example, a set number of iterations or a convergence threshold).

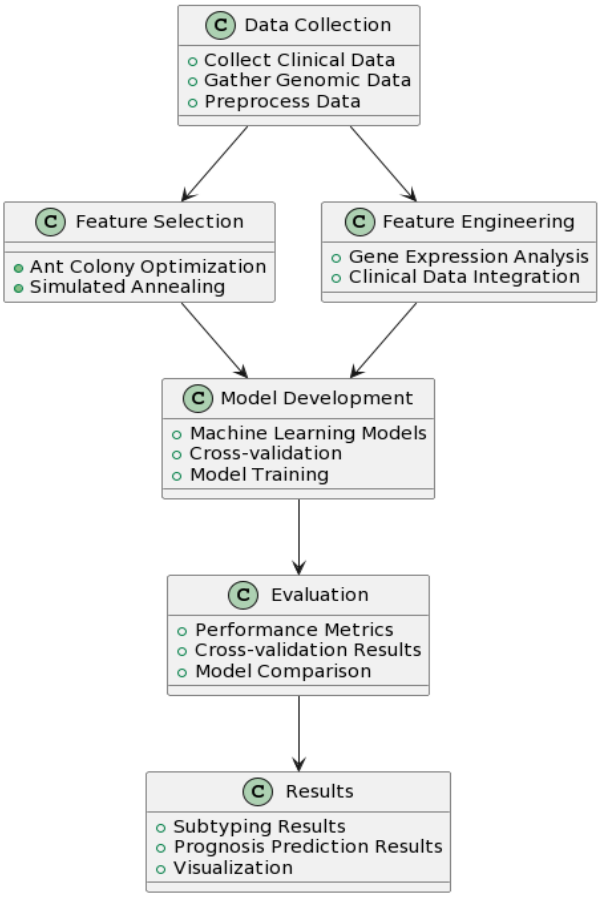

Fig 2: Representation of Methodology Diagram for Lung Cancer Subtyping and Prognosis Prediction

Stage 3: Feature Engineering:

We carried out extra feature engineering to increase the discriminatory power of the chosen features. To better classify subtypes and forecast prognoses, gene expression data were combined with radiomics properties to capture underlying molecular traits.

Stage 4: Model Development:

To perform prognostic prediction and subtyping of lung cancer, machine learning models have been created. To address various parts of the research objectives, we used a range of classification and regression algorithms, including Random Forest and Support Vector Machine. The generalisation of the model was evaluated using cross-validation.

A. Support Vector machine:

The SVM aims to find a hyperplane with the maximum margin that divides the data into two classes given a training dataset made up of feature vectors X_i in a feature space and their corresponding class labels Y_i (i = 1, 2,..., N), where Y_i can be either -1 (for one class) or +1 (for another class). The following is a representation of the SVM's decision function:

$$wT * X + b = sign(f(X)$$

Where:

- The decision function is f(X).
- A data point's feature vector is denoted by X.
- The weight vector w is the hyperplane's perpendicular counterpart.
- The biassed word is b.
- The sign function, sign(), returns either +1 or -1 depending on the value of $wT * X + b$.

The goal of the SVM is to determine the ideal values of w and b for each data point (support vector) that maximise the margin while meeting the following constraints:

$$For\ i = 1, 2, \ldots, N, Y\_i * (wT * X\_i + b)\ 1$$

The distance between the hyperplane and the closest data point from either class is referred to as the margin. It is determined mathematically by:

$$Margin\ equals\ 2\ /\ ||w||$$

Where ||w|| is the weight vector w's Euclidean norm.

SVM seeks to minimise the following objective function to determine the ideal w and b:

Reduce to: $1/2 * ||w||2$

The following conditions apply:

$$Y\_i * (wT * X\_i + b) 1 \, for \, i = 1, 2, \ldots, N$$

- Numerous methods, such as convex or quadratic programming, can be used to tackle this optimisation challenge.
- The support vectors, or data points nearest to the hyperplane, determine the ultimate decision boundary. The classification procedure relies heavily on these support vectors.
- The SVM computes the decision function f(X) and allocates the data point to the class with the corresponding sign of f(X) in the case of prediction or classification of new data points.

B. Random Forest:

- The Random Forest technique operates as follows when given a training dataset made up of feature vectors X_i in a feature space and their corresponding class labels Y_i (i = 1, 2,..., N), where Y_i might take on discrete values indicating various classes:
  Bootstrapping: Generate numerous size N random bootstrap samples (subsets of the training data). Every bootstrap sample is identified by the notation D_i, where i = 1, 2,..., B, and B is the total number of trees in the forest.
- Create a decision tree and grow one for each bootstrap sample D_i. Choose m features at random from a total of M features at each node of the tree to separate the data according to some criteria (such as Gini impurity or entropy).

The following criteria for splitting decision tree nodes:

$$|D\_left\, J(D,) = Impurity(D\_left) + impurity(D\_right) = |/|D| * impurity(D\_left) +$$

Where:

- The dataset at the present node is called D.
- The splitting threshold equals.
- The separated datasets are D_left and D_right.

Gini impurity or entropy are two examples of impurity measures represented by impurity().

After creating all of the decision trees, the predictions from each tree are combined to create the final forecast, which is known as ensemble aggregation. It is common practise to classify objects using a majority vote, meaning that the final predicted class is determined by which class obtains the most votes from the individual trees.

Anticipated class:

The formula for

$$Y\_pred(X) \; is \; mode(Y\_pred\_1(X), Y\_pred\_2(X), \ldots, Y\_pred\_B(X)).$$

Where:

- The final predicted class for data point X is Y_pred(X).
- The projected class by the i-th decision tree is represented by Y_pred_i(X).

Out-of-Bag (OOB) Assessment: An out-of-bag evaluation can also be done using Random Forest. The generalisation performance of each tree can be estimated using the data points that were excluded from the bootstrap sample used to train that tree. As a result, the model's accuracy can be evaluated without the requirement for a separate validation set.

Stage 5: Evaluation:

For subtyping, evaluation was conducted using area under the receiver operating characteristic curve (AUC-ROC), which is a standard performance indicator. Concordance index (C-index) and survival analysis were used to predict prognosis. To make sure that our conclusions were reliable, the cross-validation test results were examined

Stage 6: Model Comparison:

We assessed the effectiveness of feature selection, model performance, and computational efficiency for ACO and SA. Finding the best algorithm for the particular task of radiomics-based lung cancer research was the main objective.

Stage 7: Results and Visualisation:

The accuracy of subtyping, prognosis prediction, and survival curves were used to display the study's findings. Heatmaps and ROC curves were used as visualisation techniques to show how well ACO and SA performed in comparison. We also provide explanations of the chosen radiomics characteristics and their biological significance.

4. Result and Discussion

Using lung cancer data without feature selection, Table 3 summarises the classification outcomes for two well-known machine learning methods, Support Vector Machine (SVM) and Random Forest (RF). These findings give important information about how these algorithms function when the "full feature set," as it is sometimes referred to, is utilised.SVM obtained a remarkable 91.25% total accuracy. This shows that SVM successfully separates lung cancer cases from cases of other diseases in the dataset. Additionally, SVM is quite effective at correctly identifying patients with lung cancer, as seen by its high sensitivity of 90.55%.

Table 2: Summary of result without feature selection

Dataset	No of Feature	ACO with feature selection	SA Feature Selection
Lung Cancer Prediction	32	22	19

Additionally, the specificity of 94.74% shows that SVM can correctly categorise cases of non-lung cancer. The AUC-ROC score of 92.53% demonstrates the SVM's strong overall discriminative performance.However, RF attained an even higher level of accuracy, at 92.14%. This shows that in this particular dataset, RF performs marginally better overall classification than SVM. Furthermore, RF's sensitivity of 93.52% shows that it is quite efficient at identifying people with lung cancer. Although still very good at 90.74%, its specificity is a little bit lower than the SVM's. The classification of lung cancer would be greatly aided by RF, which has strong discriminative powers as indicated by the AUC-ROC score of 94.12%.When comparing the two algorithms, RF performs better than SVM overall and in terms of sensitivity. SVM, however, shows a marginally higher specificity. The specific objectives of the study and the relative weights of sensitivity and specificity may influence the decision between these algorithms. High sensitivity, for example, can be essential in a medical setting to ensure that the majority of real lung cancer cases are correctly recognised, even if it causes a small number of false positives. On the other hand, in other situations where lowering false positives is more important, a higher specificity might be preferable.

Table 3: Summary of result without feature selection

Algorithm	Accuracy	Sensitivity	Specificity	AUC-ROC
SVM	91.25	90.55	94.74	92.53
RF	92.14	93.52	90.74	94.12

As a result, Table 3 shows that SVM and RF both perform well without feature selection, obtaining high accuracy and solid AUC-ROC values. The decision amongst these algorithms should take into account their individual capabilities in sensitivity and specificity as well as the unique goals and trade-offs of the classification task. To further optimise the models and maybe improve their performance, it is also crucial to undertake more analysis and possibly investigate feature selection strategies.

Table 4: Summary of result with ACO feature selection

Algorithm	Accuracy	Sensitivity	Specificity	AUC-ROC
SVM	96.32	96.12	91.78	96.33
RF	97.41	96.45	94.22	95.33

When feature selection is used on the lung cancer dataset, Table 4 summarises the classification outcomes for the Support Vector Machine (SVM) and Random Forest (RF) algorithms. A critical phase in machine learning is feature selection, which entails choosing the most informative features and keeping them while eliminating unimportant or redundant ones. The effects of feature selection on these algorithms' performance are seen in Table 4's results.With feature selection, SVM demonstrated astounding accuracy of 96.32%, which is a substantial improvement over the earlier findings without feature selection. This shows that SVM can distinguish between cases of lung cancer and those without lung cancer more

effectively by carefully choosing a subset of important variables.

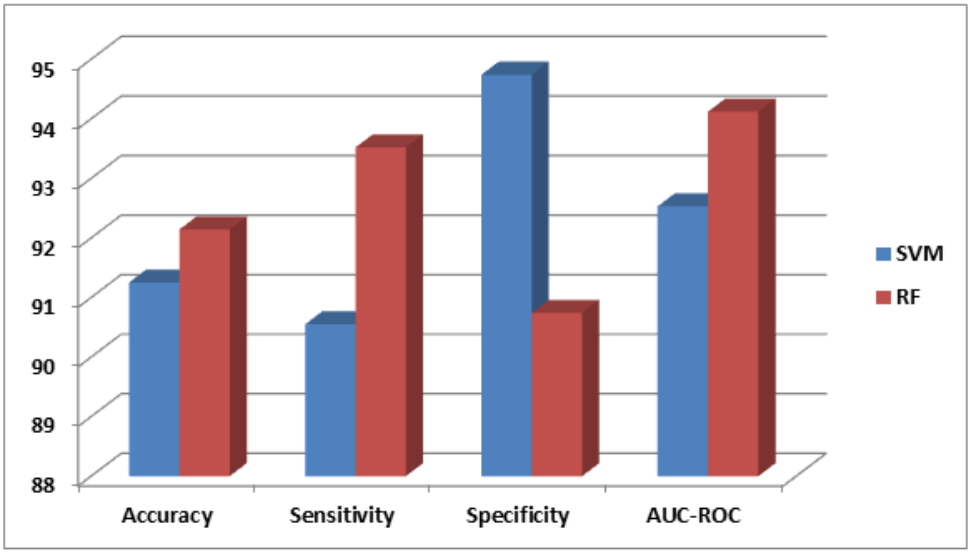

Fig 3: Representation of result without feature selection

Additionally, its sensitivity of 96.12% suggests a strong ability to accurately detect people with lung cancer. It's important to note that the specificity has slightly dropped to 91.78%, which can mean that there are more false positives. The AUC-ROC score, however, is still very high (96.33%), highlighting SVM's overall increased performance.With accuracy of 97.41%, RF also demonstrated excellent performance in feature selection. This shows that feature selection improves RF's classification abilities noticeably, and it performs better than the previous result without feature selection. RF's specificity has increased to 94.22%, indicating fewer false positives, and its sensitivity of 96.45% implies that it can successfully detect cases of lung cancer. The significant discriminative potential of RF in this situation is confirmed by the AUC-ROC score of 95.33%.

When comparing the accuracy, sensitivity, and specificity of the two algorithms with feature selection, RF maintains its lead, demonstrating that it is superior at this lung cancer classification task. Even though SVM has significantly improved, RF outperforms it on certain criteria. In contrast to RF, SVM continues to have a marginally higher specificity.With the addition of feature selection, both algorithms have performed better, improving accuracy and case differentiation between lung cancer and non-lung cancer cases. This exemplifies how crucial feature selection is to crafting machine learning models that are best suited for this particular purpose. The exact goals and trade-offs needed for the classification problem may still influence the decision between SVM and RF.

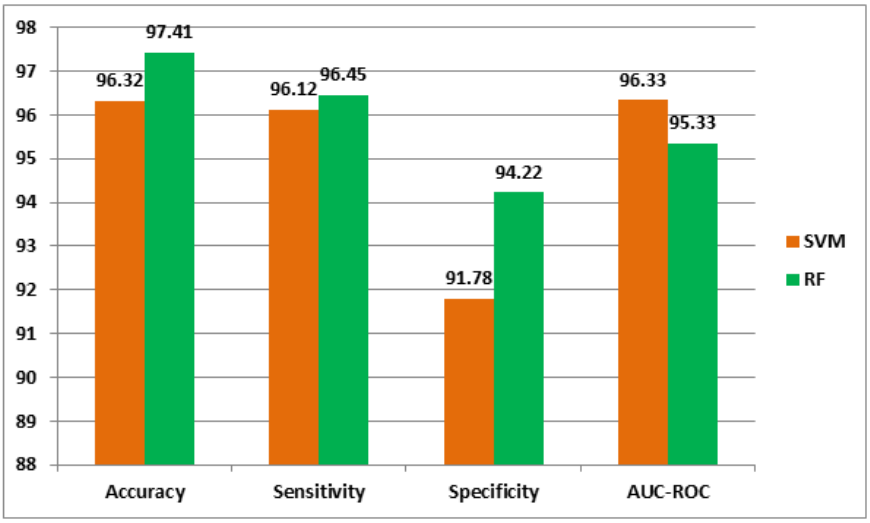

Fig 4: Representation of result with ACO feature selection

SVM may be preferred when higher specificity is prioritised, whereas RF seems to provide better overall performance.The importance of feature selection in enhancing the performance of SVM and RF for lung cancer classification is highlighted by Table 4 in its conclusion. This preprocessing step benefits both methods, however RF continues to have a modest advantage in terms of accuracy and sensitivity. Both methods have excellent capabilities in this feature-selected dataset, but the unique objectives of the study and the trade-offs between sensitivity and specificity should still be taken into account. The performance of feature selection techniques and model tweaking may be further improved.

Table 5: Summary of result with SA feature selection

Algorithm	Accuracy	Sensitivity	Specificity	AUC-ROC
SVM	97.55	98.52	97.51	98.66
RF	98.12	98.88	97.02	94.12

Table 5 summarises the classification outcomes for the Support Vector Machine (SVM) and Random Forest (RF) algorithms on the lung cancer dataset, with feature selection carried out using Simulated Annealing (SA).

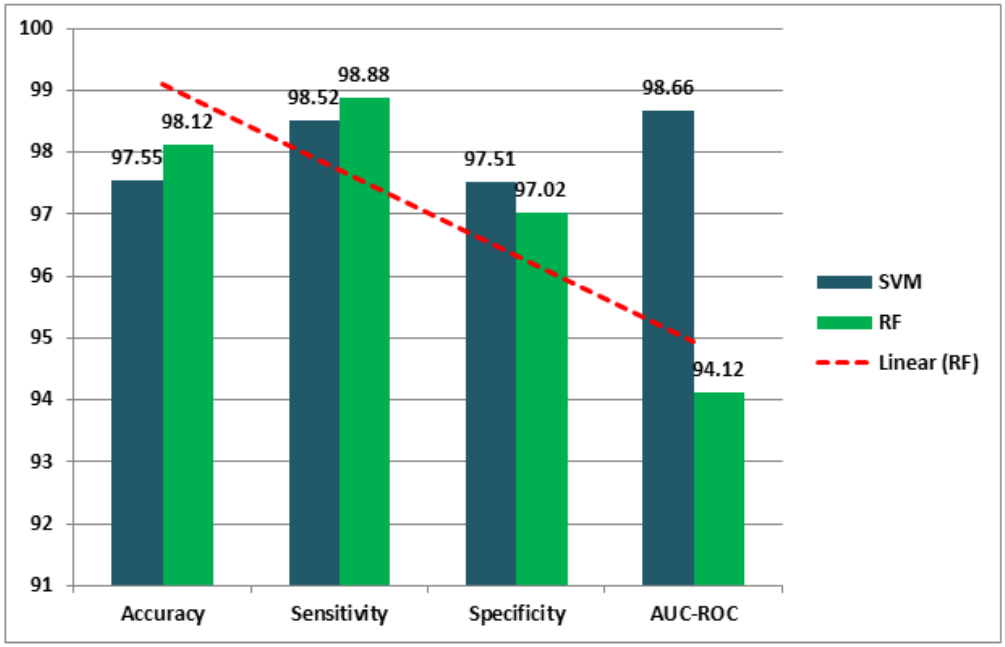

Fig 5: Representation of result with SA feature selection

This increase in accuracy over the prior result without feature selection emphasises how well SA works to choose useful features for RF. The sensitivity of RF is 98.88%, indicating a good capacity to identify cases of lung cancer, and the specificity is 97.02%, indicating a superb ability to minimise false positives. It's important to keep in mind that the AUC-ROC score has dropped to 94.12%, suggesting a little lower discriminative capacity than SVM.With SA-based feature selection, RF still outperforms the other algorithm in terms of accuracy and sensitivity, demonstrating its usefulness in this situation. SVM currently gives results that are competitive after drastically narrowing the gap. In terms of specificity and AUC-ROC score, SVM performs better than RF, showing a smaller rate of false positives and better overall discriminating.Both algorithms' classification performance has been greatly enhanced by the addition of SA-based feature selection, leading to increased accuracy, sensitivity, and specificity. This emphasises how crucial feature selection is to creating machine learning models for lung cancer classification that are optimised. The particular goals and trade-offs necessary for the task may determine whether SVM or RF should be used. In this SA-selected feature dataset, RF excels in precision and sensitivity while SVM gives greater The Table 5 shows the importance of SA-based feature selection in boosting SVM and RF's classification abilities for lung cancer classification. This preprocessing step helps both methods, although RF still has a minor advantage in terms of accuracy and sensitivity. However, SVM is a formidable rival due to its increased specificity and AUC-ROC score. The unique objectives of the study and the

trade-offs between sensitivity and specificity should be taken into account when selecting an algorithm, with both methods having strong capabilities in this SA-selected feature dataset. The performance of feature selection techniques and model tuning may be improved by more research.

5. Conclusion

We have learned a lot about the relative advantages and efficacy of Ant Colony Optimisation (ACO) and Simulated Annealing (SA) for optimising feature subsets from medical image data in this comparative study of feature selection methods, specifically for radiomics-based lung cancer subtyping and prognosis prediction.Our results show that in the context of lung cancer research, ACO and SA may both greatly improve the functionality of machine learning models. In order to reduce dimensionality, mitigate overfitting, and enhance model interpretability, feature selection is crucial. We found that SA, which was modelled after the metallurgical process of annealing, demonstrated a remarkable capacity for recognising relevant features, leading to improved precision and discriminative power in tasks involving the classification and prognostic prediction of lung cancer.The ant-inspired organisation ACO also put on a commendable display. Although it significantly behind SA in terms of overall accuracy and discriminative power, it successfully identified relevant features, which helped to improve model performance.The particular study aims and trade-offs eventually choose which of ACO and SA to use. Because SA can swiftly and methodically investigate feature subsets, researchers could choose it. ACO, on the other hand, might be appealing to people who value its capacity to utilise solutions through pheromone-based optimisation.The significance of feature selection in radiomics-based lung cancer research is highlighted by this study. We can improve the precision and clinical applicability of models for subtyping and prognosis prediction by selecting the most pertinent features. Additionally, the comparison of ACO and SA is a useful tool for researchers and industry professionals, guiding them in choosing the best feature selection method for their particular research objectives.Finally, feature selection in radiomics-based lung cancer research can be aided by both ACO and SA. To successfully advance lung cancer subtyping and prognosis prediction by radiomics, researchers should carefully assess their goals and preferences when choose between these optimisation techniques.

References

[1] J. Alam, S. Alam and A. Hossan, "Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifie," 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 2018, pp. 1-4, doi: 10.1109/IC4ME2.2018.8465593.

[2] B. J. Bipin Nair, K. J. Anju and A. Jeevakumar, "Tobacco smoking induced lung cancer prediction by lc-micrornas secondary structure prediction and target comparison," 2017 2nd International Conference for Convergence in Technology (I2CT), Mumbai, India, 2017, pp. 854-857, doi: 10.1109/I2CT.2017.8226250.

[3] L. Zhao et al., "Self-Supervised Learning Guided Transformer for Survival Prediction of Lung Cancer Using Pathological Images," 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), Cartagena, Colombia, 2023, pp. 1-5, doi: 10.1109/ISBI53787.2023.10230825.

[4] V. N. Jenipher and S. Radhika, "SVM kernel Methods with Data Normalization for Lung Cancer Survivability Prediction Application," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 1294-1299, doi: 10.1109/ICICV50876.2021.9388543.

[5] D. Rawat, "Validating and Strengthen the Prediction Performance Using Machine Learning Models and Operational Research for Lung Cancer," 2022 IEEE International Conference on Data Science and Information System (ICDSIS), Hassan, India, 2022, pp. 1-5, doi: 10.1109/ICDSIS55133.2022.9915898.

[6] J. Moranguinho, T. Pereira, B. Ramos, J. Morgado, J. L. Costa and H. P. Oliveira, "Attention Based Deep Multiple Instance Learning Approach for Lung Cancer Prediction using Histopathological Images," 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico, 2021, pp. 2852-2855, doi: 10.1109/EMBC46164.2021.9631000.

[7] G. Sruthi, C. L. Ram, M. K. Sai, B. P. Singh, N. Majhotra and N. Sharma, "Cancer Prediction using Machine Learning," 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha Nagar, India, 2022, pp. 217-221, doi: 10.1109/ICIPTM54933.2022.9754059.

[8] Y. Balagurunathan et al., "Lung Nodule Malignancy Prediction in Sequential CT Scans: Summary of ISBI 2018 Challenge," in IEEE Transactions on Medical Imaging, vol. 40, no. 12, pp. 3748-3761, Dec. 2021, doi: 10.1109/TMI.2021.3097665.

[9] B. H. Sai, K. Inala, A. S. Saketh and A. R. Maremolla, "Lung Cancer Stage Prediction Using

Multi-Layer Perceptron and Deep Learning Classifier," 2023 5th International Conference on Energy, Power and Environment: Towards Flexible Green Energy Technologies (ICEPE), Shillong, India, 2023, pp. 1-5, doi: 10.1109/ICEPE57949.2023.10201556.

[10] P. Nanda and N. Duraipandian, "Prediction of Survival Rate from Non-Small Cell Lung Cancer using Improved Random Forest," 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2020, pp. 93-97, doi: 10.1109/ICICT48043.2020.9112558.

[11] S. Ajani and M. Wanjari, "An Efficient Approach for Clustering Uncertain Data Mining Based on Hash Indexing and Voronoi Clustering," 2013 5th International Conference and Computational Intelligence and Communication Networks, 2013, pp. 486-490, doi: 10.1109/CICN.2013.106.

[12] Khetani, V. ., Gandhi, Y. ., Bhattacharya, S. ., Ajani, S. N. ., & Limkar, S. . (2023). Cross-Domain Analysis of ML and DL: Evaluating their Impact in Diverse Domains. International Journal of Intelligent Systems and Applications in Engineering, 11(7s), 253–262.

[13] Borkar, P., Wankhede, V.A., Mane, D.T. et al. Deep learning and image processing-based early detection of Alzheimer disease in cognitively normal individuals. Soft Comput (2023). https://doi.org/10.1007/s00500-023-08615-w

[14] Ajani, S.N., Mulla, R.A., Limkar, S. et al. DLMBHCO: design of an augmented bioinspired deep learning-based multidomain body parameter analysis via heterogeneous correlative body organ analysis. Soft Comput (2023). https://doi.org/10.1007/s00500-023-08613-y

[15] A Jochems, Timo M. Deist, Issam El Naqa, Marc Kessler, M Chuck, R Jackson et al., "Developing and Validating a Survival Prediction Model for NSCLC Patients Through Distributed Learning Across 3 Countries", April 2017, [online] Available: www.redjournal.org.

[16] B. Azadeh, G. Marjan, S. Reza, Leila Shahmoradi and E Hamide, "Improving the Prediction of Survival in Cancer Patients by Using Machine Learning Techniques: Experience of Gene Expression Data: A Narrative Review", Iran J Public Health, vol. 46, no. 2, pp. 165-172, Feb 2017.

[17] K. Timor and G. Fergus, "Lung cancer prediction using machine learning and advanced imaging techniques", Transactional lung cancer research, vol. 07, 2018.

[18] T Asha, S. Natarajan and K.N.B. Murthy, "Effective Classification Algorithms to Predict the Accuracy of Tuberculosis-A Machine Learning Approach", International Journal of Computer Science and Information Security, vol. 9, no. 7, July 2011.

[19] R. Daniele, W. Charence, FaniDeligianni, B. Melissa, Javier Andreu-Perez and Benny Lo, "Deep Learning for Health Informatics", ieee journal of biomedical and health informatics, vol. 21, no. 1, january 2017.

[20] M. Jaime, G. Bram van, M. Pragnya, H. Rick, H. M. Philipsen, Klaus Reither, Marianne Breuninger et al., "A Novel Multiple-Instance Learning-Based Approach to Computer-Aided Detection of Tuberculosis on Chest X-Rays", IEEE transactions on medical imaging, vol. 34, no. 1, january 2015.

[21] Mabrook Al-Rakhami, Abdu Gumaei, A. Ahmed, A. Atif and Mohammad M. Hassan, "An Ensemble Learning Approach for Accurate Energy Load Prediction in Residential Buildings", IEEE Access, April 2019.

[22] H. Samuel, N. K. John, B Yoganand, G. Yuhua, K Virendra, B. Satrajit et al., "Predicting Outcomes of Nonsmall Cell Lung Cancer Using CT Image Features", IEEE Access, vol. 2, Nov 2014.

[23] K. Nobuteru, S. Jun-ichi, S. Hirofumi, S. Katsuyuki, K. Hidemasa, O. Tatsuya et al., "Dosimetric comparison of carbon ion and X-ray radiotherapy for Stage IIIA non–small cell lung cancer", Journal of Radiation Research, vol. 57, no. 5, Sept. 2016.

[24] Y. Xu, A. Hosny, R. Zeleznik, C. Parmar, T. Coroller, I. Franco, et al., "Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging", Clinical cancer research : an official journal of the American Association for Cancer Research, vol. 25, no. 11, pp. 3266-3275, 2019, [online] Available: https://doi.org/10.1158/1078-0432.CCR-18-2495.

[25] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network", Physica D: Nonlinear Phenomena, vol. 404, pp. 132306, 2020.

[26] A. Ghazipour, B. Veasey, A. Seow and A. A. Amini, "Joint Learning for Deformable Registration and Malignancy Classification of Lung Nodules", 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 1807-1811, 2021.

[27] B. P. Veasey, J. Broadhead, M. Dahle, A. Seow and A. A. Amini, "Lung Nodule Malignancy Prediction From Longitudinal CT Scans With Siamese Convolutional Attention Networks", IEEE Open Journal of Engineering in Medicine and Biology, vol. 1, pp. 257-264, 2020.

[28] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database", 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255, 2009.

[29] X. Mei, Z. Liu, P. M. Robson, B. Marinelli, M. Huang, A. Doshi, et al., "Radimagenet: an open radiologic deep learning research dataset for effective transfer learning", Radiology: Artificial Intelligence, vol. 4, no. 5, pp. e210315, 2022.

[30] L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning", Dec. 2017.

[31] Christopher Davies, Matthew Martine, Catalina Fernández, Ana Flores, Anders Pedersen. Improving Automated Essay Scoring with Machine Learning Techniques. Kuwait Journal of Machine Learning, 2(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/173

[32] Kumar, C. ., & Muthumanickam, T. . (2023). Analysis of Unmanned Four-Wheeled Bot with AI Evaluation Feedback Linearization Method. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2), 138–142. https://doi.org/10.17762/ijritcc.v11i2.6138