An Efficient Model for Lung Cancer Detection through the Integration of Genetic Algorithm and Machine Learning

Abdulaziz A. Alsulami
Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
aaalsulami10@kau.edu.sa (corresponding author)

Received: 5 October 2024 | Revised: 30 October 2024 | Accepted: 3 November 2024

DOI: https://doi.org/10.48084/etasr.9188

ABSTRACT

Prompt lung cancer detection is essential for patient health. Deep Learning (DL) models have been intensively used for lung cancer screening, as they provide high accuracy in diagnoses. However, DL models require significant computational power, which may not be accessible in all settings. Conventional Machine Learning (ML) models may not produce high prediction accuracy, especially with large data. This study uses a Genetic Algorithm (GA) approach to select optimal features from lung cancer images and reduce their dimensionality. This allows conventional ML models to achieve a high prediction accuracy when classifying medical images while using lower computational power compared with DL models. The proposed model integrates GA along with ML for lung cancer detection. The experimental results show that using GA with a feed-forward neural network classifier achieved high performance, reaching 99.70% classification accuracy.

Keywords-machine learning; image classification; genetic algorithm; deep learning

I. INTRODUCTION

Lung cancer is considered one of the deadliest cancer types worldwide [1, 2]. The USA is expected to experience 2,001,140 new cases of cancer and 611,720 deaths related to cancer in 2024 [1]. This means that lung cancer is responsible for around 20% of all cancer deaths [1]. Smoking and chronic contact with airborne pollutants are the most frequent causes of lung nodules, which are areas of higher density in the lung [3]. Although lung nodules can be considered the first indication of lung cancer, early and precise identification is crucial for treatment [4, 5]. Immunotherapy has recently shown promising results for people with lung cancer [6]. However, the objective response varies significantly from patient to patient [7]. Thus, it is important to accurately detect patients with lung cancer who are sensitive to immunotherapy [8]. To correctly identify a patient, numerous Computed Tomography (CT) scans of the lungs may be necessary [9]. However, this places a burden on physicians that can affect their judgment and result in a wrong conclusion or a lack of diagnosis [10, 11]. Computer-Aided Detection (CAD) techniques could help physicians provide diagnoses with greater accuracy. Several CAD frameworks operate in two phases: reducing the rate of false positives and extracting candidate frames [12, 13]. This approach sends several unclear nodules detected within the initial coarse imaging scan (high false positive rate and high sensitivity) to the next stage for analysis [14, 15]. Other techniques are intensity threshold and shape curvature [13].

In addition, conventional methods have been used to reduce false alarms, including gradient, location, density, shape, size, texture, and both lower and upper human data [16, 17]. There are two issues with traditional CAD techniques. The first issue is general inefficiency, and the other is that the actual circumstance differs considerably, affecting the overall identification outcome and assumption [18-20]. With the advancement of big data and statistical techniques, AI has become more crucial to every aspect of people's lives [21]. AI focuses on developing models based on intelligence that can think and reason to cope with problems, make decisions, and automate tasks [22]. Machine Learning (ML) is a subfield of AI that examines a large number of data samples with a particular objective, to process and then classify them [23]. AI refers to the idea of a "thinking machine", highlighting the ability to formulate independent choices [24]. The fields of precision medicine, personalized medicine, and biomedical research may all be significantly affected by Deep Learning (DL) methods that are based on artificial intelligence and a subset of ML [25]. DL models such as Convolutional Neural Networks (CNNs) are utilized by researchers to classify medical images [26]. Figure 1 illustrates the general architecture of classifying medical images using DL model. First, the images are collected by different medical scanning technologies like X-rays and CT scans. Then, the collected images are preprocessed using several tasks such as feature extraction, image resizing, and image augmentation. Third, the images are ready to be fed into a DL module to perform classification tasks. The model is trained and evaluated on preprocessed images. Finally, each image is classified into a category depending on the medical conditions or diagnostic labels. ML is an approach for learning new information about patients by using computers to analyze genetics, imaging, and other biological data. This is done through mathematical simulation and analysis and the results are utilized in clinical and scientific research. The cross-fertilizing of artificial intelligence and medicine is an increasingly important topic for research [27]. This means that there is a need for using detection method to detect lung cancer and exploring this disease at an earlier phase, which helps in having good treatment and decreases the death rate caused by lung cancer. Medical images contain a large number of features which make the classification by traditional ML methods more challenging. Therefore, there is a need for an efficient feature selection technique to select the most important features to reduce the dimensionality of feature space [28].

This study proposes a novel model that integrates a Genetic Algorithm (GA) approach with an ML model to detect lung cancer. The GA is utilized to select the most relevant features from lung cancer images, which results in reducing the size of the feature space. Using conventional ML models instead of CNNs can address limited computational capabilities in clinics. Moreover, this approach achieves high classification performance on lung cancer images. The contributions of this research are:

- Implements GA to select the optimal subset of features from lung cancer images. This reduces redundant and irrelevant features, helping to reduce the computation process and obtain high classification accuracy with ML models.
- Studies and analyzes the classification performance of common ML models to evaluate the effectiveness of using GA with ML models. Integrating GA with a Feed-Forward Neural Network (FFNN) classifier achieved the highest classification accuracy, reaching 99.70%.
- Compares the classification performance of the proposed model (GA-FFNN) with state-of-the-art classifiers on the same dataset.

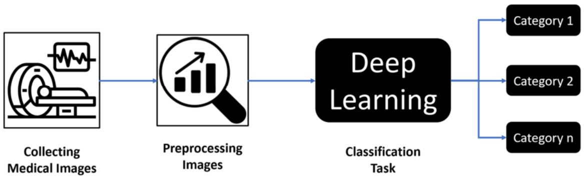

Fig. 1. General pipeline for classifying medical images.

II. RELATED WORKS

In [29], a method called Sybil was presented using Low-Dose Computed Tomography (LDCT) images. Sybil requires LDCT images without the need for radiologist annotations and can run in real-time. Sybil was validated on three datasets: a set of 6282 LDCTs of participants in the National Lung Screening Trial (NLST), a set of 12,280 from Chang Gung Memorial Hospital (CGMH), and a set of 8,821 from Massachusetts General Hospital (MGH). This method achieved prediction rates of 92% on NLST, 86% on MGH, and 94% on CGMH. Authors in [30] proposed the LungNet-SVM algorithm, an enhanced version of the AlexNet architecture, as a framework for an automated cancer nodule identification method for CT scans. It consists of two complete connected layers, three pooling layers, and seven convolutional layers for feature extraction. This model used an SVM to categorize nodules into benign or malignant. The authors evaluated their model using the LUNA16 dataset. The model had an accuracy rate of 97.64%. Their results indicate that their proposed model achieved higher accuracy than the existing methods. Authors in [31] proposed CCDCHNN (Cancer Cell Detection utilizing Hybrid Neural Network) to detect lung cancer in early phases. DL was used on CT images to extract features. The authors stated that feature extraction is important for the early detection of lung cancer. They utilized an advanced 3D CNN to enhance the accuracy of lung cancer detection. Their method was able to differentiate between malignant and benign tumors. In their experiments, they used two publicly available datasets, LIDC/IDRI and LUNA16. The proposed model achieved an accuracy rate of 95% outperforming other algorithms. Authors in [32] suggested a hybrid ML method for lung nodule for diagnosis in the early stages based on clinical images and statistical data used. They used lung nodule segmentation by utilizing Improved Snake Swarm Optimization with a Bat model (ISSO-B). To minimize the dimensionality issue, an approach called Chaotic Atom Search Optimization (CASO) was utilized to choose important features among several features. They developed a hybrid Learning-based Deep Neural Network (L-DNN) classifier to predict and classify nodules depending on hybrid learning. They assessed their method by utilizing two datasets, LIDC-IDRI and FAH-GMU. They compared L-DNN with SVM, Softmax, and ELM, and the results indicated that L-DNN achieved an accuracy of 96.39%, outperforming the others. Authors in [33] introduced a technique used for lung nodule detection and classification with CT images depending on hybrid DL. They presented an algorithm called Chaotic Bird Swarm Optimization (CBSO) for the segmentation of lung nodules utilizing statistical data. They introduced an improved version of the Fish Bee algorithm to extract and select features. They developed the Hybrid Differential Evolution-based Neural Network (HDE-NN) to classify and predict tumor. According to the experimental findings, using computed tomography improves sensitivity and lowers the amount of false positives when identifying lung nodes on CT images. The proposed strategy illustrates how integrating clinical practice with the HDE-NN node identification could provide advantages. They the LIDC-IDRI dataset and de novo scans and achieved an accuracy rate of 96.39% which was higher than that of SVM, Dense softmax CNN, and ELM. Authors in [34] introduced an innovative hybrid deDL model through the pretrained CNN model SqueezeNet and ResNet101 to extract feature mappings from the CT images in the dataset. These two DL CNN architectures outperformed the other examined architectures. To increase the computational effectiveness and performance of the suggested method, the optimal features from both models were extracted utilizing minimize Redundancy and Maximize Relevance (mRMR). Features that did not affect the accuracy were eliminated. SVM and KNN were used to build the optimal feature map and classify the data. The model achieved an accuracy rate of 99.09% with SVM. The experiment's findings demonstrate that the hybrid model did excellently with regard to of accuracy when applied to the IQOTH/NCCD dataset.

Authors in [35] proposed an enhanced ML model to improve the classification of lung cancer phase using CT images. They conducted several preprocessing steps using the IQ-OTHNCCD lung cancer dataset that included CT scans from healthy persons and lung cancer patients at different phases. Their preprocessing included Gaussian blurring, scaling, and normalization. CNN was trained on the preprocesses data. The issue of class imbalance was addressed through the SMOTE technique. The research findings indicated that the model achieved a classification accuracy of 99.64%. Authors in [36] developed a method for detecting lung cancer called Hybrid Lung Cancer Stage Classifier and Diagnosis Model (Hybrid-LCSCDM). The authors tried to simplify the lung cancer diagnose process by classifying the patients to normal, malignant, and benign. They analyzed the CT images using a two-part method including VGG-16 model that used to extract key features in CT of lung that indicates cancer. These extracted features were classified by XGBoost that helps in sorting the scans to three groups. IQ-OTH/NCCD dataset was used for training and evaluating the suggested model. Their model achieved an overall accuracy rate of 98.54%

In [37], an ensemble model was used to enhance the accuracy of lung cancer detection, called the Mitscherlich Function-based Ensemble Network (MENet), which integrated the probabilities of prediction gathered from MobileNetV2, Xception, and InceptionResNetV2. The ensemble method was based on the Mitscherlich work, aggregating the outputs of the aforementioned base classifiers into a fuzzy rank. Two publicly available datasets, LIDC-IDRI and IQ-OTH/NCCD, were used in training and testing, where the proposed ensemble model achieved an accuracy of 99.54%, outperforming other models.

Most of the previous studies did not fully focus on the computational process when designing their methods. This is an important aspect, as it can optimize the use of available resources. In addition, some studies did not pay enough attention to selecting the optimal features of medical images. Selecting optimal features is important since it can reduce computational load while achieving high classification accuracy.

III. METHODOLOGY

A. *Proposed Model Overview*

This study proposes a detection model for lung cancer that classifies CT images into three classes: normal, malignant, and benign. It leverages GA to select optimal features, minimizing redundancy and eliminating irrelevant features. This enhances the classification performance of ML models when performing classification tasks. Therefore, the proposed model integrates GA with an ML model.

Figure 2 illustrates a flow diagram that outlines the steps used to implement the proposed model. The process starts with loading the lung cancer dataset, which contains images corresponding to the three above-mentioned classes. As medical datasets, like other datasets, could suffer from class imbalance that can lead to overfitting, an oversampling technique was integrated. After the dataset was balanced, the next step was to pass each image to the AlexNet pre-trained neural network for extracting features. Then, the GA parameters were defined to configure the algorithm and optimize feature selection. Adjusting GA variables such as population size, mutation rate, crossover probability, and the number of generations helps GA find the most relevant features in the images. This ensures reducing the dimensionality of the data by selecting only valuable features while obtaining high classification accuracy. GA is used to select the most relevant features and reduce feature space to improve the performance of ML in detecting lung cancer. Once optimal features are selected, the next step involves using an ML classifier to classify patient diagnosis into three categories. In summary, this method implements image oversampling, Alexnet, GA, and ML.

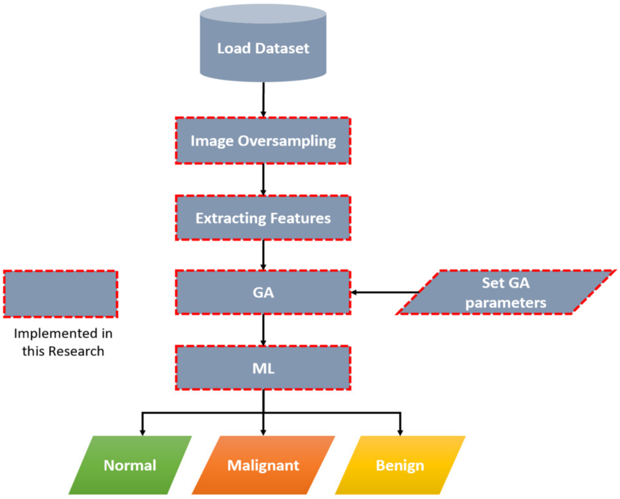

Fig. 2. Flow diagram of the proposed approach.

B. *Dataset*

This study used the IQ-OTH/NCCD dataset, which contains a collection of CT scan slices [38-40] collected from Iraqi hospitals in three months in 2019. Slices represent images of the human chest, captured using a Siemens SOMATOM CT scanner, with a protocol of 120 kV, 1 mm slice thickness, and varying window widths and centers for reading, ranging from 350 to 1200 Hounsfield Unit. Each scan consists of 80 to 200 slices, showing the lungs from different angles and sides. The dataset has three classes, Benign, Malignant, and Normal, and a total of 1,097 images, as listed in Table I. However, the dataset has imbalanced classes, especially for the Benign class, which could contribute to overfitting in classification tasks. To address this issue, the proposed model employs oversampling to increase the number of images. This is achieved by identifying the class with the highest number of images, in this case, the Malignant class with 561 images, and then increasing the other classes to map the class with the highest number of images.

TABLE I. CLASSES OF THE IQ-OTH/NCCD DATASET

Category	Number of original images	Number of images after oversampling
Benign	120	561
Malignant	561	561
Normal	416	561
Total	1,097	1,683

C. Theory of Genetic Algorithm (GA)

GA is an optimization method derived from the natural selection process [41]. GA starts with a sample of individuals in a population and then generates a new one using genetic operators on the sample population to find the best individual fit [42]. GA is used to solve complex problems employing the concepts of population, chromosome, gene, fitness function, selection, crossover, mutation, and termination. Population represents a group of solutions to the problem, as shown in:

$$P(n) = S_1, S_2, \dots, S_i \qquad (1)$$

where P refers to population, n to the generation number, S to the individual solution or chromosome, and i to i^{th} solution. Chromosome represents an individual solution to the problem, usually in binary string, as shown in (2):

$$S_i = G_1, G_2, \dots, G_m \qquad (2)$$

where G refers to a gene and m is the length of the chromosome. Each single chromosome consists of elements called genes and can take a binary value (0 or 1).

A fitness function is used to evaluate the individual solution (chromosome), as shown in:

$$F(S_i) \rightarrow R \qquad (3)$$

where F is a fitness function that can be formulated depending on the problem and R is a real number. Selection is used to select the best individuals, which is also known as survival. Crossover is a method used to combine parts from two parents to produce offspring. Mutation is used to alter some parts of the offspring.

To understand the operation of GA, assume that we want to find the minimum solution of the following function:

$$f(x) = (\mathrm{x} - 2) \qquad (4)$$

Assume also that the value of x is an integer between 0 and 15. The aim is to find a value between 0 and 15 that gives the smallest solution of $f(x)$. At the beginning, a random population is generated, and each population is represented as a 4-bit binary string. Therefore, the first population could be:

- Value of $x = 4$ (binary: 0100)
- Value of $x = 15$ (binary: 1111)
- Value of $x = 2$ (binary: 0010)

As the objective is to find the minimum solution for $f(x)$ in (4), the smallest value is the best solution.

- $f(4) = (4 - 2) = 2$
- $f(15) = (15 - 2) = 13$
- $f(2) = (2 - 2) = 0$

$x = 2$ is the best solution since it produces the minimum value. Now, the fittest population is chosen to produce the next generation, which can be $x = 4$ (binary 0100) and $x = 2$ (binary 0010), since they generated the smallest fitness values. The crossover method can be applied to the selected individuals, Parent 1: 0100 and Parent 2: 0010. This is done to produce Offspring 1 and Offspring 2. Assume that the first offspring is generated from the first two bits of Parent 1 and the second two bits of Parent 2. Then Offspring 1 is 0110. Offspring 2 is generated from the last two bits of Parent 1 and the first two bits of Parent 2, which gives 0000. Therefore, in the case of Offspring 1, $x = 6$, and for Offspring 2, $x = 0$. The mutation method can be used to ensure diversity in the population. This can be done by flipping the last two bits of Offspring 2, for example. Offspring 2 now becomes 0011 ($x = 3$). Therefore, the new population is $x = 6$ and $x = 3$. The evaluation of the fitness function using the new population gives:

$$f(6) = (6 - 2) = 4$$

$$f(3) = (3 - 2) = 1$$

The smallest value of the fitness function using the new population is $f(3)$. However, it is not smaller than $f(2)$, thus the algorithm is terminated as the optimal solution has already been selected.

D. Feature Selection Using GA

Algorithm 1 describes the GA feature selection approach. In the input, the dataset that contains lung cancer images is loaded and each image is read. Then AlexNet is used to extract all features from each image. After that, the following GA parameters are set: population size, max generation, and number of features. The value of each parameter was selected based on experiments. The expected output of the algorithm is the most important features organized in columns, and each row is labeled with the class to which it belongs (Normal, Malignant, Benign).

Oversampling is performed to increase the number of images since the dataset is not balanced. The oversampling technique used is random duplication of images, so the number of images in each class matches the class that has the maximum number of images. Subsequently, AlexNet is used to extract all features from the images, and the results are fed to GA to select the most important features. Finally, the GA output is saved in a CSV file to be used later by the ML model.

```
Algorithm 1: Feature Selection Using GA
Input:
  Load dataset
  Read images
  Load AlexNet
  Set GA parameters:
    Set populationSize
    Set maxGenerations
    Set numFeatures
Output:
  importantFeatures.csv
```

```
Procedure:
  Perform oversampling
  Extract features using AlexNet
  Run GA
  Save relevant features to
  importantFeatures.csv
End Algorithm
```

IV. RESULTS AND DISCUSSION

This section presents the experimental results of this research. It begins by showing a defined set of parameters used for each classifier. Then, it provides a comparative study of various ML classifiers that integrate the GA approach to evaluate their effectiveness. Finally, this section describes a comparative analysis of the proposed model with recent studies using the IQ-OTH/NCCD dataset.

Table II provides information about each classifier and the corresponding parameters used to classify lung cancer images. Default values were used with each classifier to achieve fairness. The Random Forest (RF) classifier uses Gini impurity as the splitting criterion, and the splitter is set to the best. The Decision Tree (DT) classifier also uses the Gini criterion and 100 decision trees. The SVM classifier uses a linear kernel, and the regularization is set to 1.0 to avoid overfitting. The K-Nearest Neighbor (KNN) classifier relies on Singular Value Decomposition (SVD). The hidden layer of the Feedforward Neural Network (FFNN) classifier consists of 100 neurons, uses the Rectified Linear Unit (ReLU) as the activation function, and Adam as an optimizer. The Linear Discriminant Analysis (LDA) classifier employs five neighbors.

TABLE II. CLASSIFIER PARAMETERS

Classifier	Parameters
RF	Criterion: Gini, splitter: best
SVM	Kernel: linear, regularization: 1.0
KNN	Solver: SVD
FFNN	Hidden layer sizes: 100, activation: ReLU, solver: Adam
LDA	Number of neighbors: 5
DT	Number of decision trees :100, criterion: Gini

The performance comparison of ML models used along with the GA algorithm is displayed in Table III. These models were evaluated using metrics including precision, recall, F1-score, and accuracy. The RF classifier achieves an impressive performance whereas the SVM classifier performs a perfect balanced performance. In addition, the SVM classifier exhibits identical accuracy with the RF classifier, which shows its proficiency in the context of lung cancer classification. The KNN classifier shows an accuracy of 97.92%, which is lower than that of RF and SVM. The FFNN classifier outperforms all the ML methods considered in this study, with perfect precision reaching 100%. In addition, the classifier shows reliable and strong performance with recall, F1-score, and accuracy of 99.72%, 99.86%, and 99.70%, respectively. The LAD classifier yields classification performance close to the SVM classifier, and it produces an identical accuracy to RF and SVM. LDA results do not outperform the FFNN, but it shows a solid performance compared to the other ML methods. The DT is a simpler classifier, and according to the results, it is not one of the top performers in context of lung cancer classification. The DT classifier shows the lowest scores across all metrics compared with RF, SVM, KNN, FFNN, and LDA

TABLE III. ML CLASSIFICATION PERFORMANCE USING GA

Model	Precision	Recall	F1-score	Accuracy
RF	98.88%	99.16%	99.02%	99.11%
SVM	98.77%	98.77%	98.77%	99.11%
KNN	97.88%	97.69%	97.40%	97.92%
FFNN	100%	99.72%	99.86%	99.70%
LDA	98.72%	98.87%	98.77%	99.11%
DT	94.93%	95.12%	94.95%	94.36%

Table IV presents a comparative analysis of research studies conducted in 2023-24. These studies used different classifiers on the IQ-OTH/NCCD dataset, which is the same dataset used in this study. According to their results, most studies achieved classification accuracy greater than 99.00% except [36, 43] which obtained 98.32% and 98.54%, respectively. In [44], a CNN with BiLSTM achieved 99.20%. In [34], ResNet101 with SqueezeNet reached 99.09%. In [35, 37], comparable results were obtained, with 99.64% and 99.54% accuracy, respectively. This study, using GA with FFNN, outperformed these models by achieving 99.70% accuracy.

TABLE IV. COMPARISON OF DIFFERENT CLASSIFIERS EVALUATED ON IQ OTH/NCCD

Study	Model	Accuracy
[34]	ResNet101 + SqueezeNet (Hybrid model)	99.09%
[35]	CNN + SMOTE	99.64%
[36]	LCSCDM	98.54%
[37]	MENet (Xception, InceptionResNetV2, MobileNetV2)	99.54%
[43]	binary count ratio	98.32%
[44]	CNN-Bi LSTM	99.20%
This study	**GA-FFNN**	**99.70%**

V. CONCLUSION

Medical images are often analyzed using DL models that require a lot of computational resources. This study proposed an approach that leverages the power of GA and ML to detect lung cancer. GA was applied to select the most important features from lung cancer CT images, helping to reduce feature space size and enhance classification performance when using conventional ML models. This approach used GA to select the most relevant features and ML to perform the classification task. A comparative analysis was performed to evaluate the effectiveness of the proposed model with various ML models. The proposed model was compared with state-of-the-art classifiers on the same dataset. The experimental results confirm that the proposed model achieved high classification performance by integrating GA and FFNN. In the future, the proposed model should be tested with more diverse datasets to validate its robustness and generalizability.

REFERENCES

[1] R. L. Siegel, A. N. Giaquinto, and A. Jemal, "Cancer statistics, 2024," *CA: a cancer journal for clinicians*, vol. 74, no. 1, pp. 12–49, 2024, https://doi.org/10.3322/caac.21820.

[2] J. Kuon *et al.*, "Impact of molecular alterations on quality of life and prognostic understanding over time in patients with incurable lung cancer: a multicenter, longitudinal, prospective cohort study," *Supportive Care in Cancer*, vol. 30, no. 4, pp. 3131–3140, Apr. 2022, https://doi.org/10.1007/s00520-021-06736-2.

[3] M. A. Balcı, L. M. Batrancea, Ö. Akgüller, and A. Nichita, "A Series-Based Deep Learning Approach to Lung Nodule Image Classification," *Cancers*, vol. 15, no. 3, Jan. 2023, Art. no. 843, https://doi.org/10.3390/cancers15030843.

[4] H. Zhang, Y. Peng, and Y. Guo, "Pulmonary nodules detection based on multi-scale attention networks," *Scientific Reports*, vol. 12, no. 1, Jan. 2022, Art. no. 1466, https://doi.org/10.1038/s41598-022-05372-y.

[5] Y. Li, J. Chang, and Y. Tian, "Improved cost-sensitive multikernel learning support vector machine algorithm based on particle swarm optimization in pulmonary nodule recognition," *Soft Computing*, vol. 26, no. 7, pp. 3369–3383, Apr. 2022, https://doi.org/10.1007/s00500-021-06718-w.

[6] A. Lahiri *et al.*, "Lung cancer immunotherapy: progress, pitfalls, and promises," *Molecular Cancer*, vol. 22, no. 1, Feb. 2023, Art. no. 40, https://doi.org/10.1186/s12943-023-01740-y.

[7] J. S. Deutsch *et al.*, "Association between pathologic response and survival after neoadjuvant therapy in lung cancer," *Nature Medicine*, vol. 30, no. 1, pp. 218–228, Jan. 2024, https://doi.org/10.1038/s41591-023-02660-6.

[8] S. Tang, T. Fan, X. Wang, C. Yu, C. Zhang, and Y. Zhou, "Cancer Immunotherapy and Medical Imaging Research Trends from 2003 to 2023: A Bibliometric Analysis," *Journal of Multidisciplinary Healthcare*, vol. 17, pp. 2105–2120, Dec. 2024, https://doi.org/10.2147/JMDH.S457367.

[9] W. He, B. Li, R. Liao, H. Mo, and L. Tian, "An ISHAP-based interpretation-model-guided classification method for malignant pulmonary nodule," *Knowledge-Based Systems*, vol. 237, Feb. 2022, Art. no. 107778, https://doi.org/10.1016/j.knosys.2021.107778.

[10] M. A. Heuvelmans *et al.*, "Lung cancer prediction by Deep Learning to identify benign lung nodules," *Lung Cancer*, vol. 154, pp. 1–4, Apr. 2021, https://doi.org/10.1016/j.lungcan.2021.01.027.

[11] P. Dutande, U. Baid, and S. Talbar, "LNCDS: A 2D-3D cascaded CNN approach for lung nodule classification, detection and segmentation," *Biomedical Signal Processing and Control*, vol. 67, May 2021, Art. no. 102527, https://doi.org/10.1016/j.bspc.2021.102527.

[12] B. Yin, M. Sun, J. Zhang, W. Liu, C. Liu, and Z. Wang, "AFA: adversarial frequency alignment for domain generalized lung nodule detection," *Neural Computing and Applications*, vol. 34, no. 10, pp. 8039–8050, May 2022, https://doi.org/10.1007/s00521-022-06928-9.

[13] F. Shariaty, M. Orooji, E. N. Velichko, and S. V. Zavjalov, "Texture appearance model, a new model-based segmentation paradigm, application on the segmentation of lung nodule in the CT scan of the chest," *Computers in Biology and Medicine*, vol. 140, Jan. 2022, Art. no. 105086, https://doi.org/10.1016/j.compbiomed.2021.105086.

[14] X. Zhang, S. Li, B. Zhang, J. Dong, S. Zhao, and X. Liu, "Automatic detection and segmentation of lung nodules in different locations from CT images based on adaptive -hull algorithm and DenseNet convolutional network," *International Journal of Imaging Systems and Technology*, vol. 31, no. 4, pp. 1882–1893, 2021, https://doi.org/10.1002/ima.22580.

[15] H. Zhang and H. Zhang, "LungSeek: 3D Selective Kernel residual network for pulmonary nodule diagnosis," *The Visual Computer*, vol. 39, no. 2, pp. 679–692, Feb. 2023, https://doi.org/10.1007/s00371-021-02366-1.

[16] L. Sun *et al.*, "Attention-embedded complementary-stream CNN for false positive reduction in pulmonary nodule detection," *Computers in Biology and Medicine*, vol. 133, Jun. 2021, Art. no. 104357, https://doi.org/10.1016/j.compbiomed.2021.104357.

[17] C.-F. J. Kuo, J. Barman, C. W. Hsieh, and H.-H. Hsu, "Fast fully automatic detection, classification and 3D reconstruction of pulmonary nodules in CT images by local image feature analysis," *Biomedical Signal Processing and Control*, vol. 68, Jul. 2021, Art. no. 102790, https://doi.org/10.1016/j.bspc.2021.102790.

[18] K. Chen *et al.*, "Clinical impact of a deep learning system for automated detection of missed pulmonary nodules on routine body computed tomography including the chest region," *European Radiology*, vol. 32, no. 5, pp. 2891–2900, May 2022, https://doi.org/10.1007/s00330-021-08412-9.

[19] E. J. Ostrin *et al.*, "Contribution of a Blood-Based Protein Biomarker Panel to the Classification of Indeterminate Pulmonary Nodules," *Journal of Thoracic Oncology*, vol. 16, no. 2, pp. 228–236, Feb. 2021, https://doi.org/10.1016/j.jtho.2020.09.024.

[20] H. Farhat, G. E. Sakr, and R. Kilany, "Deep learning applications in pulmonary medical imaging: recent updates and insights on COVID-19," *Machine Vision and Applications*, vol. 31, no. 6, Jul. 2020, Art. no. 53, https://doi.org/10.1007/s00138-020-01101-5.

[21] A. Holzinger, K. Keiblinger, P. Holub, K. Zatloukal, and H. Müller, "AI for life: Trends in artificial intelligence for biotechnology," *New Biotechnology*, vol. 74, pp. 16–24, May 2023, https://doi.org/10.1016/j.nbt.2023.02.001.

[22] A. Aldoseri, K. N. Al-Khalifa, and A. M. Hamouda, "Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges," *Applied Sciences*, vol. 13, no. 12, Jan. 2023, Art. no. 7082, https://doi.org/10.3390/app13127082.

[23] S. L. Goldenberg, G. Nir, and S. E. Salcudean, "A new era: artificial intelligence and machine learning in prostate cancer," *Nature Reviews Urology*, vol. 16, no. 7, pp. 391–403, Jul. 2019, https://doi.org/10.1038/s41585-019-0193-3.

[24] V. Kaul, S. Enslin, and S. A. Gross, "History of artificial intelligence in medicine," *Gastrointestinal Endoscopy*, vol. 92, no. 4, pp. 807–812, Oct. 2020, https://doi.org/10.1016/j.gie.2020.06.040.

[25] A. Prelaj *et al.*, "Machine Learning Using Real-World and Translational Data to Improve Treatment Selection for NSCLC Patients Treated with Immunotherapy," *Cancers*, vol. 14, no. 2, Jan. 2022, Art. no. 435, https://doi.org/10.3390/cancers14020435.

[26] S. T. Vemula, M. Sreevani, P. Rajarajeswari, K. Bhargavi, J. M. R. S. Tavares, and S. Alankritha, "Deep Learning Techniques for Lung Cancer Recognition," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 14916–14922, Aug. 2024, https://doi.org/10.48084/etasr.7510.

[27] B. He *et al.*, "Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker," *Journal for Immunotherapy of Cancer*, vol. 8, no. 2, Jul. 2020, Art. no. e000550, https://doi.org/10.1136/jitc-2020-000550.

[28] W. Ali and F. Saeed, "Hybrid Filter and Genetic Algorithm-Based Feature Selection for Improving Cancer Classification in High-Dimensional Microarray Data," *Processes*, vol. 11, no. 2, Feb. 2023, Art. no. 562, https://doi.org/10.3390/pr11020562.

[29] P. G. Mikhael *et al.*, "Sybil: A Validated Deep Learning Model to Predict Future Lung Cancer Risk From a Single Low-Dose Chest Computed Tomography," *Journal of Clinical Oncology*, vol. 41, no. 12, pp. 2191–2200, Apr. 2023, https://doi.org/10.1200/JCO.22.01345.

[30] I. Naseer, T. Masood, S. Akram, A. Jaffar, M. Rashid, and M. Amjad Iqbal, "Lung Cancer Detection Using Modified AlexNet Architecture and Support Vector Machine," *Computers, Materials & Continua*, vol. 74, no. 1, pp. 2039–2054, 2023, https://doi.org/10.32604/cmc.2023.032927.

[31] S. Wankhade and V. S., "A novel hybrid deep learning method for early detection of lung cancer using neural networks," *Healthcare Analytics*, vol. 3, Nov. 2023, Art. no. 100195, https://doi.org/10.1016/j.health.2023.100195.

[32] A. Syed Musthafa, K. Sankar, T. Benil, and Y. N. Rao, "A hybrid machine learning technique for early prediction of lung nodules from medical images using a learning-based neural network classifier," *Concurrency and Computation: Practice and Experience*, vol. 35, no. 3, 2023, Art. no. e7488, https://doi.org/10.1002/cpe.7488.

[33] V. K. Gugulothu and S. Balaji, "An early prediction and classification of lung nodule diagnosis on CT images based on hybrid deep learning techniques," *Multimedia Tools and Applications*, May 2023, https://doi.org/10.1007/s11042-023-15802-2.

[34] S. Zafar, J. Ahmad, Z. Mubeen, and G. Mumtaz, "Enhanced Lung Cancer Detection and Classification with mRMR-Based Hybrid Deep Learning Model," *Journal of Computing & Biomedical Informatics*, vol. 7, no. 02, Sep. 2024.

[35] M. M. Musthafa, I. Manimozhi, T. R. Mahesh, and S. Guluwadi, "Optimizing double-layered convolutional neural networks for efficient lung cancer classification through hyperparameter optimization and advanced image pre-processing techniques," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, May 2024, Art. no. 142, https://doi.org/10.1186/s12911-024-02553-9.

[36] A. Mohammed Qadir, P. Ahmed Abdalla, and D. Faiq Abd, "A Hybrid Lung Cancer Model for Diagnosis and Stage Classification from Computed Tomography Images," *Iraqi Journal for Electrical and Electronic Engineering*, vol. 20, no. 2, pp. 266–274, Dec. 2024, https://doi.org/10.37917/ijeee.20.2.23.

[37] S. Majumder, N. Gautam, A. Basu, A. Sau, Z. W. Geem, and R. Sarkar, "MENet: A Mitscherlich function based ensemble of CNN models to classify lung cancer using CT scans," *PLOS ONE*, vol. 19, no. 3, 2024, Art. no. e0298527, https://doi.org/10.1371/journal.pone.0298527.

[38] H. F. Al-Yasriy, M. S. Al-Husieny, F. Y. Mohsen, E. A. Khalil, and Z. S. Hassan, "Diagnosis of Lung Cancer Based on CT Scans Using CNN," *IOP Conference Series: Materials Science and Engineering*, vol. 928, no. 2, Aug. 2020, Art. no. 022035, https://doi.org/10.1088/1757-899X/928/2/022035.

[39] H. F. Kareem, M. S. AL-Huseiny, F. Y. Mohsen, E. A. Khalil, and Z. S. Hassan, "Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 3, Mar. 2021, Art. no. 1731, https://doi.org/10.11591/ijeecs.v21.i3.pp1731-1738.

[40] H. F. Al-Yasriy, "The IQ-OTH/NCCD lung cancer dataset." Kaggle, https://doi.org/10.34740/KAGGLE/DS/672399.

[41] A. Rajab, "Genetic Algorithm-Based Multi-Hop Routing to Improve the Lifetime of Wireless Sensor Networks," *Engineering, Technology & Applied Science Research*, vol. 11, no. 6, pp. 7770–7775, Dec. 2021, https://doi.org/10.48084/etasr.4484.

[42] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 8091–8126, Feb. 2021, https://doi.org/10.1007/s11042-020-10139-6.

[43] S. Saechueng and U. Suttapakti, "Binary Count Ratio for Lung Cancer Classification in Computerized Tomography Scan Images," in *2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Osaka, Japan, Feb. 2024, pp. 070–074, https://doi.org/10.1109/ICAIIC60209.2024.10463210.

[44] B. Mostafa, M. Sakr, and A. Keshk, "Employing the Capabilities of LSTM and Bi-LSTM for Lung Cancer Detection and Classification.," *International Journal of Intelligent Engineering & Systems*, vol. 17, no. 5, 2024.