



Lung cancer disease prediction with CT scan and histopathological images feature analysis using deep learning techniques

Vani Rajasekar^a, M.P. Vaishnnavi^b, S. Premkumar^c, Velliangiri Sarveshwaran^{d,*}, V. Rangaraaj^a

^a Department of Computer Science and Engineering, Kongu Engineering College, Erode, India

^b Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India

^c Department of Computer Science and Engineering, Galgotias University, Uttar Pradesh, India

^d Department of Computational Intelligence, SRM Institute of Science and Technology, Kattankulathur Campus, Chennai, India

ARTICLE INFO

Keywords:

Lung cancer
Histopathological images
CT Scan images
Deep learning
Diagnosis

ABSTRACT

Lung cancer is characterized by the uncontrollable growth of cells in the lung tissues. Early diagnosis of malignant cells in the lungs, which provide oxygen to the human body and excrete carbon dioxide because of important processes, is critical. Because of its potential importance in patient diagnosis and treatment, the use of deep learning for the identification of lymph node involvement on histopathological slides has attracted widespread attention. The existing algorithm performs considerably less in recognition accuracy, precision, sensitivity, F-Score, Specificity, etc. The proposed methodology shows enhanced performance in the metrics with six different deep learning algorithms like Convolution Neural Network (CNN), CNN Gradient Descent (CNN GD), VGG-16, VGG-19, Inception V3 and Resnet-50. The proposed algorithm is analyzed based on CT scan images and histopathological images. The result analysis shows that the detection accuracy is better when histopathological tissues are considered for analysis.

1. Introduction

Cancer has long been recognised as a hazardous disease that can lead to death. Lung cancer is found to be one among frequent malignancies worldwide, and it is the important reason of mortality from cancer in both industrialised and developing countries. The majority of these individuals have non-small cell lung cancer (NSCLC), which has a 5-year mortality rate of only 18%. Despite current medical advances resulting in a significant raise in overall cancer survival rates, this progress is lower significant in lung cancer because the majority of symptoms and recognised patients have advanced illness [1]. The cancer cells migrate from the lungs to the lymph glands and then to the bloodstream. Due to the obvious natural movement of lymph, cancer cells usually spread towards the middle of the chest. If this cancer had spread to other organs, metastasis will occur, and early identification can help avoid this. Nonsurgical methods, including as radiation, chemotherapy, surgical intervention, or monoclonal antibodies, are frequently used to treat late-stage lesions. This emphasises the critical importance of employing follow-up radiography to monitor therapeutic response and track

tumour radiographic changes with time [2] (see Figs. 1–4).

Researchers recently established predictive algorithms based on differentially expressed genes to categorise lung cancer patients with different health outcomes, such as relapse and total survival4. Biomarkers have been proven to be important in the treatment of NSCLC in previous research. Radiographic tumour characteristics may now be assessed quantitatively rather than qualitatively, thanks to artificial intelligence (AI), a technique known as "radiomics" [3]. Several studies have shown that noninvasively describing tumour characteristics has greater predictive power than standard clinical assessments. Pathologists have been using traditional microscopes and glass slides to make an accurate diagnosis since the mid-nineteenth century. The traditional procedure has the pathologist manually reviewing several glass slides, which takes a great amount of time and energy. The introduction of slide scanning equipment that can increase production digital slides has ushered classical pathology into the digital world, bringing various benefits to the histopathology process [4]. The major advantage is the capability to utilize computer simulations, such as automated image analysis, to assist health care professionals in the investigation and

* Corresponding author.

E-mail addresses: vanikecit@gmail.com (V. Rajasekar), vaishnnavi.cse@sathyabama.ac.in (M.P. Vaishnnavi), premambal@gmail.com (S. Premkumar), velliangiris@gmail.com (V. Sarveshwaran), rangaraaj1@gmail.com (V. Rangaraaj).

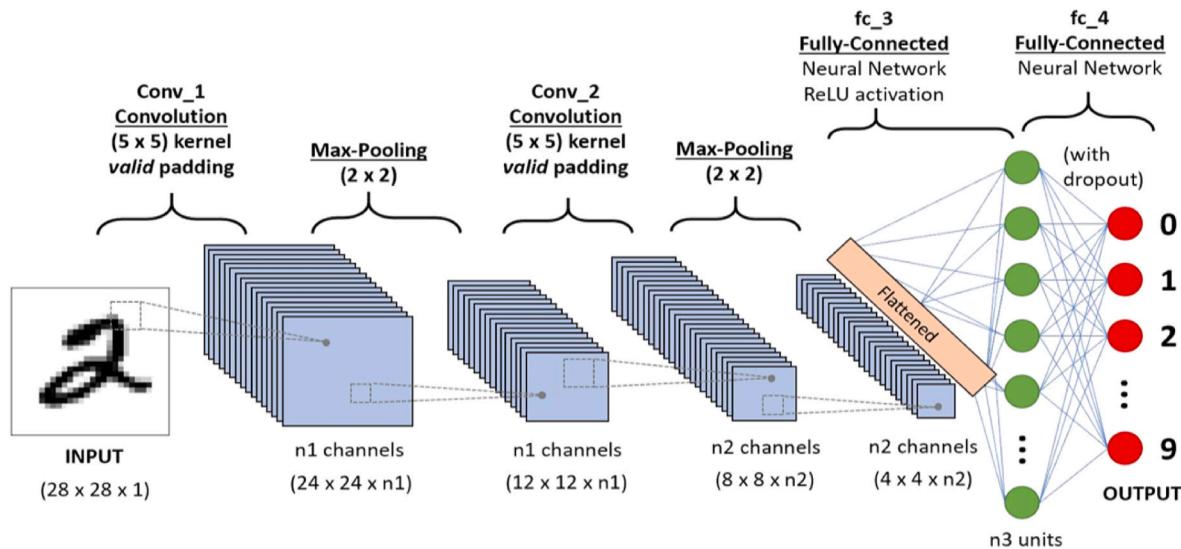


Fig. 1. Proposed CNN architecture.

quantitative determination of slides, reducing the amount of time needed for manual testing and trying to improve pathologist precision, repeatability, and workflow efficiency. The use of deep learning approaches to aid diagnostics has recently sparked a great deal of interest in histopathology.

Convolutional neural networks (CNNs) in general have shown tremendous promise in medical disease identification. CNNs have been employed in pathology with the promising outcomes from tumour cell identification in metastatic breast cancer to grade glioma and colon cancer. When it comes to classifying cancer and making the necessary therapeutic options, a thorough examination of lymph glands is essential. Multiple lymph node levels are considered when determining prognosis and appropriate staging necessitates meticulous examination of lymph node health [5]. Histopathological pictures have recently been discovered to be an accurate predictor of various therapeutic biomarkers. Manually screening several slides, on the other hand, can be tiresome and time-consuming for pathologists, and people are capable of making mistakes since they must keep track of which sections they already studied. Several kind of available solutions in these competitions were capable of demonstrating greater efficiency in terms of accuracy micro metastases than a pathologist operating in a busy practise. In lung cancer, primary tumour metastases have an identical role in predicting cancer stage, therapy options, and outcomes as they do in breast cancer [6].

In this proposed methodology, a deep learning-based lung cancer

disease prediction with CT scan and histopathological images was employed. The prediction analysis is made with different algorithms like CNN with quasi-convex gradient descent, VGG16, InceptionV3, VGG 19 and Resnet50. A novel strategy for CNN optimization with quasi-gradient descent is employed which results in higher accuracy of lung cancer detection and reduced false positives. The main contributions of the proposed approaches are listed as follows:

- Designing a deep learning methodology for lung cancer detection based on CT scan images and Histopathological images
- Models are designed with CNN, CNN GD, Inception V3, Resner-50, VGG-16, and VGG-19
- Analyze the performance with various metrics like recognition accuracy, precision, Specificity, Sensitivity, F-Score.

2. Related works

The development of malignant cells in the lungs is known as lung cancer. Overall men and women's mortality rates have increased as a result of growing cancer incidence. Lung cancer is a disease wherein the cells in the lungs quickly multiply. Lung cancer cannot be eradicated, but it can be reduced [7]. The number of people affected with lung cancer is precisely equal to the number of people who smoke continuously. Lung cancer treatment was evaluated using classification approaches such as Naive Bayes, SVM, Decision Tree, and Logistic

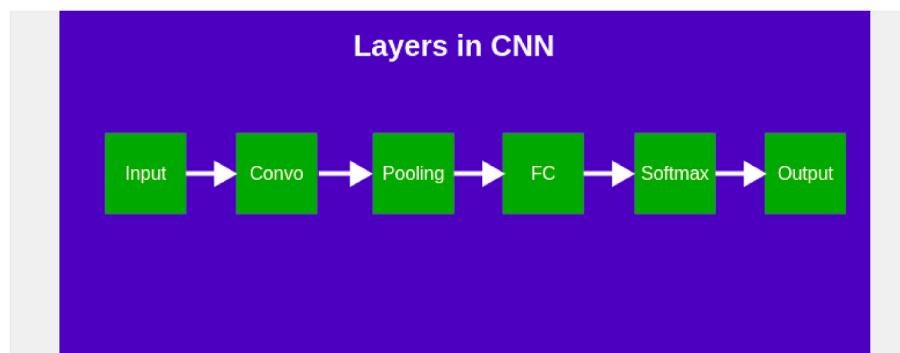


Fig. 2. Different layers in CNN.

- Input Layer:

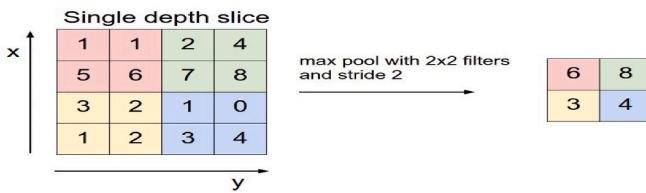


Fig. 3. Pooling layer.

Regression. Pradhan et al. [8] conduct an empirical evaluation of multiple machine learning methods that can be used to identify lung cancer using IoT devices. A survey of roughly 65 papers employing machine learning techniques to forecast various diseases was conducted in this study. The study focuses on a variety of machine learning methods for detecting a variety of diseases in order to identify a gap in prospective lung cancer detection in medical IoT. Deep residual learning is used by Bhatia et al. [9] to identify lung cancer from CT scans. With the UNet and ResNet algorithms, we propose a series of pre-processing approaches for emphasising cancer-prone lung regions and retrieving characteristics. The extracted features are fed through several classifiers, namely Adaboost and Random Forest, and the individual predictions are ensembled to calculate the likelihood of a CT scan becoming cancerous.

Without learning inadequate human information, Shin et al. [10,11] use deep learning to investigate the characteristics of cell exosomes and determine the similarities in human plasma extracellular vesicles. The deep learning classifier was tested with exosome SERS data from regular and lung cancer cell lines and was able to categorise them with 95% efficiency. The deep learning algorithm projected that 90.7% of patients' plasma exosomes were more similar to lung cancer cell extracellular vesicles than the mean of healthy controls in 43 patients, encompassing stage I and II cancer patients. In the ability to forecast lung ADC subtypes, researchers looked at four clinical factors: age, sex, tumour size, and smoking status, as well as 40 radiomic markers. The LIFEx software was used to extract radiomic characteristics from PET scans of segmented cancers. The clinical and radio mic variables were

ranked, and a subset of meaningful features was chosen based on Gini coefficient scores for histopathological class relationships [12]. In the estimation of survival, a deep learning network with a tumour cell and metastatic staging system was used to examine the dependability of individual therapy suggestions supplied by the deep learning preservation neural network. The C statistics were employed to evaluate the performance of the model. The computational intelligence survival neural network model's longevity forecasts and treatment strategies were made easier with the use of a customer interface [13].

A lung cancer detection model that utilizes image analysis and machine intelligence to identify the occurrence of lung cancer in CT scans and blood tests has been developed. Despite the fact that CT scan findings are more efficient than mammograms, patient CT scan pictures are divided into normal and abnormal categories [14,15]. Even in the same tumour stage, non-small-cell cancer patients have a wide range of clinical performance and results. This research investigates deep learning applications such as medical imaging, which could help with patient stratification by automating the measurement of radiographic properties. Ardila et al. [16] suggest a novel system for predicting lung cancer risk based on a patient's historical and current computerized tomography dimensions. Our model obtains a 94.4% state-of-the-art efficiency. When prior computed tomography imagery was unavailable, their model excelled all six radiologists, reducing false positives by 11% and false negatives by 5%. For CT scans of the lungs, Lakshmanaprabhu et al. [17] introduce a new automatic diagnosis categorization system. The CT scan of lung pictures was processed using Optimal Deep Neural Network (ODNN) and Linear Discriminate Analysis in their research (LDA). The remainder of the paper is organized as section 3 describes the deep learning techniques used in the proposed methodology. Section 4 describes the result comparison of various proposed techniques. Section 5 describes the conclusion and future extension of the proposed work.

3. Proposed methodologies

3.1. CNN

Multilayer perceptrons are regularised variations of CNNs. Multilayer perceptrons are a dynamic network that is linked together where each neuron is interconnected to all neurons in the next layer. Because of their total connection, these systems are susceptible to an information explosion. CNNs need very little pre-processing compared to certain other image categorization algorithms. This means that the system uses automated retraining to enhance the filters whereas traditional algorithms require hand-engineering. The lack of prior information and human interaction in extracting features is a significant benefit. Normalization or preventing overfitting can be accomplished in a variety of methods, including compensating parameters throughout training or reducing interconnectivity. CNNs consider a new method to normalization: they reap the benefits of another patient's patterns in data and generate models of increasing the complexity using smaller and simpler patterns imprinted in their filters. As a result, CNNs are at the bottom end of the connectedness and complexity range.

3.1.1. Layers in CNN

The various five layers in CNN are a) Input layer b) Convolutional layer c) Pooling layer d) Fully connected layer e) Softmax/logistic layer and f) Output layer.

This layer in CNN should have to include the imaging information. A three-dimensional matrix is used to represent image data. This layer requires reshaping into a single column. Before feeding an image over the dimensions of $28 \times 28 = 784$, it will be converted to 784×1 .

• Convolutional layer (Convolution + ReLU)

This layer is also called the features extraction layer since it retrieves components of the image. A piece of the information is originally

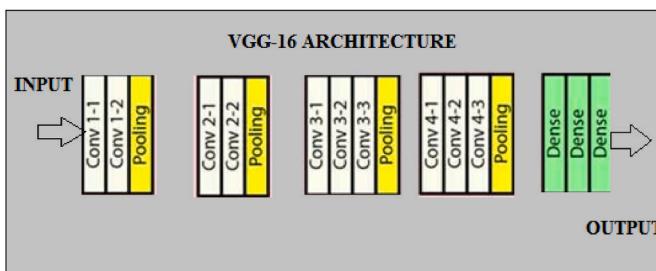


Fig. 4. VGG-16 architecture.

- The first two layers are 3×3 filter convolution layers with a total of 64 filters. Because the same convolution is utilised, the size will be $224 \times 224 \times 64$. The incremental is always 1 and the filter has always been 3×3 .
- Subsequently, the volume height and breadth were reduced from $224 \times 224 \times 64$ to $112 \times 112 \times 64$ using a pooling layer with such a maxpool size of 2×2 and an incremental of 2.
- Two convolution layers containing 128 filters are then applied. This will result in the creation of a different dimension of $112 \times 112 \times 128$.
- The volume decreases to $56 \times 56 \times 128$ after utilising the pooling layer.
- Two 256-filter fully connected layers were added, accompanied by a down-sampling layer that reduced the size to $28 \times 28 \times 256$.
- The max pool layer separates two further stacks, each one with three convolution layers.
- The $7 \times 7 \times 512$ level is reduced to a fully connected (FC) layer with 4096 inputs and classes 1000 softmax output after the last max pooling.

attached to the Convolutional layer to perform convolution and filtration [18–20]. The result size is represented by a single integer returned by the procedure. Then slide the filter in each of the next regions of interest with the same input images using a Stride and continue the operation. The same technique is repeated till the whole process is controlled. The result will be used as the feed for the following layer. The Convolutional layer also has ReLU activation, which lowers all negative numbers to zero.

To reduce the temporal volume of the source images, max pooling is used after convolution. It is utilised between two convolution layers. Applying FC after the Convolutional layer without employing pooling or maximum pooling will be operationally demanding. As a result, using maximal pooling seems to be the only technique to minimise the input image's geographic volume. In the above instance, max pooling was applied to a single spatial dimension with a Span of 2.

The four-by-four-dimensional input is reduced to two-by-two dimensions. The pooling layer has no characteristics, but it does have two hyperparameters, Filter(F) and Stride (S).

In general, if the input dimension is $W_1 \times H_1 \times D_1$, then

$$W_2 = (W_1 - F) / S + 1 \quad (1)$$

$$H_2 = (H_1 - F) / S + 1 \quad (2)$$

$$D_2 = D_1 \quad (3)$$

where W_2 , H_2 and D_2 are the width, height and depth of output.

• Fully Connected Layer (FC)

The completely linked layer includes weights, biases, and neurons. It links neurons in one layer to those in another. It is used to teach individuals how to categorise images into different groups.

• Softmax Layer (FC)

The Softmax or Stochastic layer is the final layer of CNN. It was at the very bottom of the Convolution layers. Softmax is suitable for multi-classification, whereas stochastic is suitable for binary classification.

• Output Layer (FC)

The output label in the form of one hot encoding is available in the output layer.

3.1.2. CNN with quasi gradient descent

Gradient descent (GD) is a recurrent first-order optimization process

for locating a function's local minimum or maximum. To reduce a cost or loss function, these techniques are frequently used in machine learning (ML) and deep learning (DL). The two specific requirements of gradient descent are.

- Differentiable
- Convex

Spatially separated functions have a derivative at each point in their domain, although not all functions meet those standards. Convex implies that the line segment joins different function points lies on or above the curve of the univariate variable. For two points x_1 , x_2 laying on the function's curve this condition is expressed in equation 4.

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \quad (4)$$

where λ signifies a point's position on a section line and its value should be between 0 which is left most point and 1 which is right most point. Similarly, 0.5 denotes an intermediate location. Compute the second derivative to see if its value has always been greater than zero to see if a univariate function is convex.

$$\frac{d^2f(x)}{dx^2} \quad (5)$$

With a gradient descent approach, quasi-convex functions can also be used. However, they frequently feature "saddle points," also known as "minimax points," where the method might become stuck.

3.2. VGG16

VGG16 is a CNN architecture that took first place in the 2014 ILSVR ImageNet contest. One of the greatest Vision model architectures remains. VGG16 is distinct in that, rather than many hyperparameters, authors centred on the 3×3 filter's convolution layer in step 1 and always used the same padding and max pool layer in step 2. All through the architecture, the sequence of convolutional levels and maximal pool tiers is maintained [21]. The VGG architecture serves as the foundation for cutting-edge object classification models. The VGGNet, which was created as a deep neural network, outperforms baselines on a variety of tasks and datasets other than ImageNet. Furthermore, it remains one of the best most widely used image recognition architectures today.

3.3. Inception V3

Convolutional Neural Networks are used in the Inception V3 learning

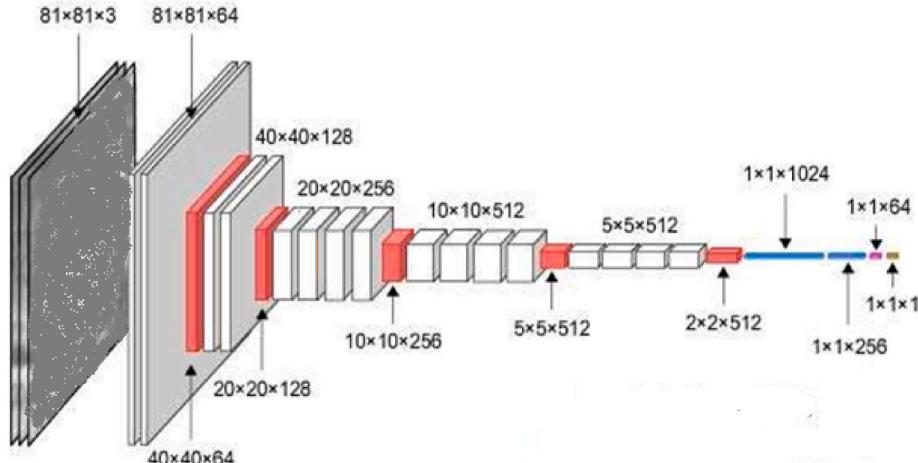
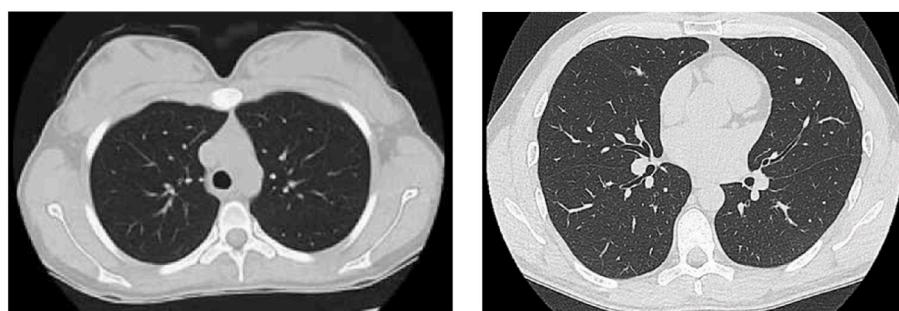
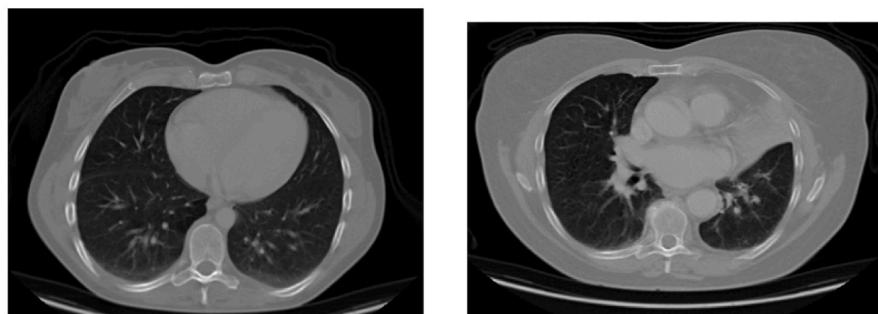


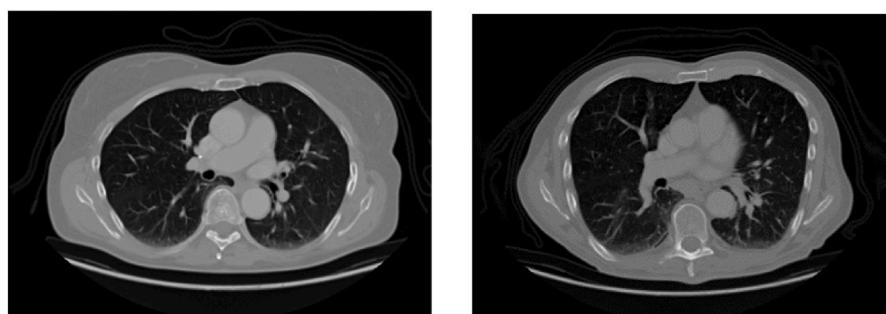
Fig. 5. VGG-19 architecture.



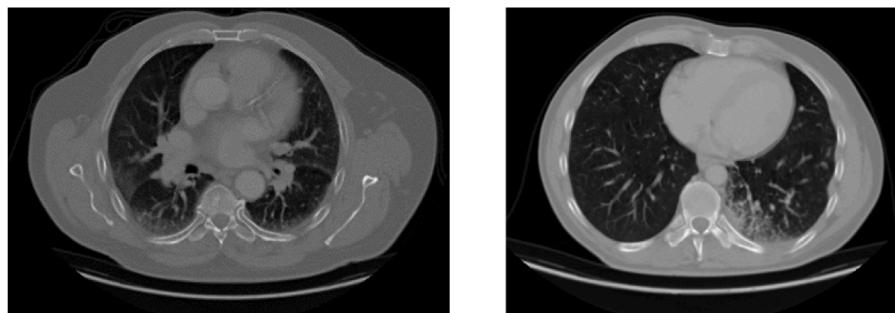
a) Normal Lung



b) Adenocarcinoma



c) Large cell carcinoma



d) Squamous cell carcinoma

Fig. 6. CT scan images of Lung dataset.

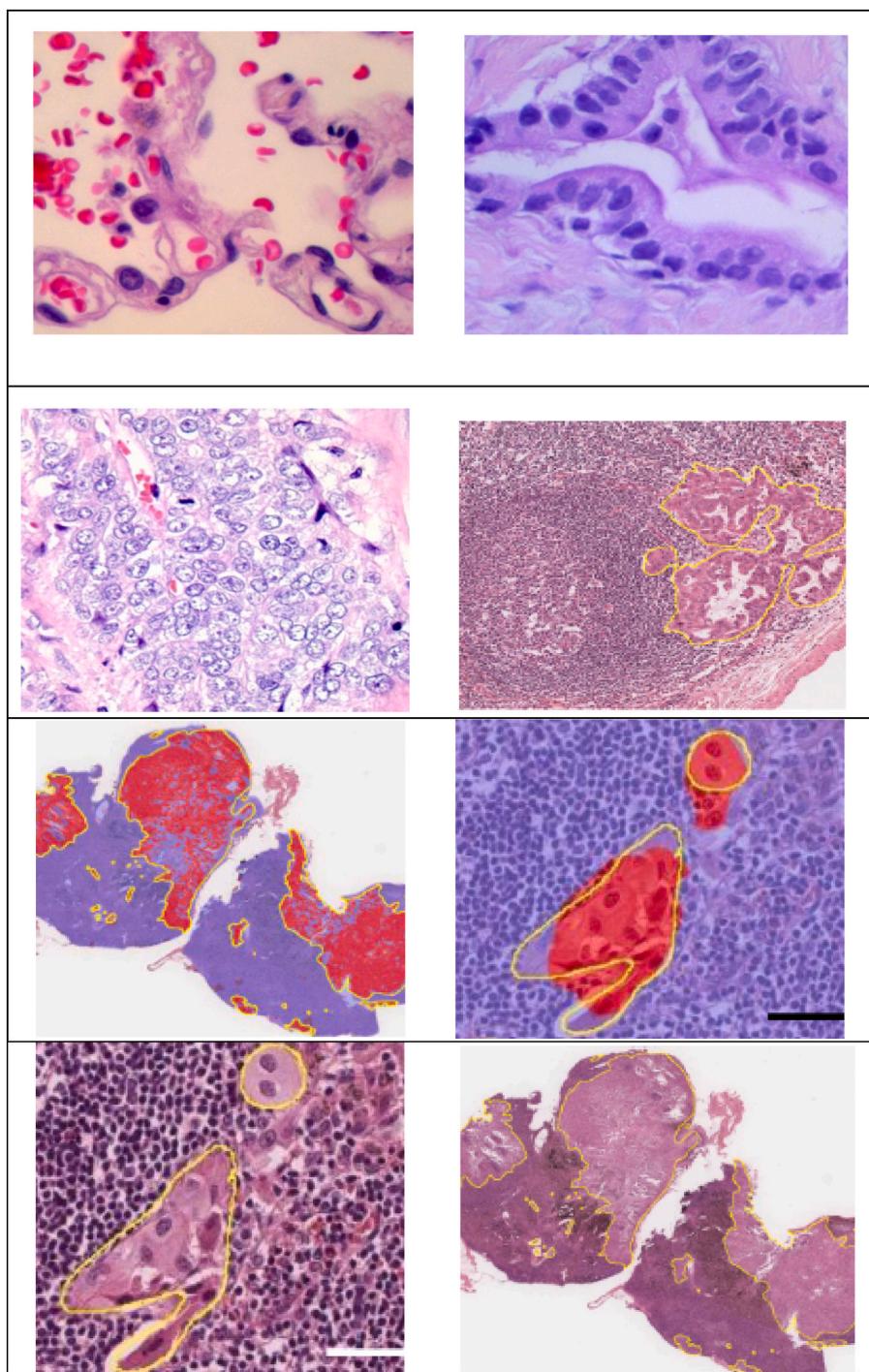


Fig. 7. CT scan images of Lung dataset.

algorithm for picture categorization. The Inception V3 is a much more enhanced form of the basic model Inception V1, which was first launched as Google Net in 2014. When several deep levels of convolutions were used in a model, overfitting of the data happened. To get around this, the V1 model uses the concept of several filters of different diameters on the same level. Therefore, instead of having in-depth layers in our inception models, we now have parallel layers, which makes the system larger instead of deeper. One of the great benefits of the Inception V1 model was the considerable dimension reduction. To improve the model, the larger Convolutions were factored into smaller

Convolution layers. Consider the basic module of the conceptualization V1 module [22–24]. It has 55 convolutional layers, which are computationally very expensive, as previously stated. To reduce the computational cost, the 55% convolutional layer was replaced by 33% convolutional layers. Max pooling and average pooling were typically lower than the grid size of feature maps. In order to efficiently reduce the grid size, the activation dimensions of the network filters are increased in the inception V3 design.

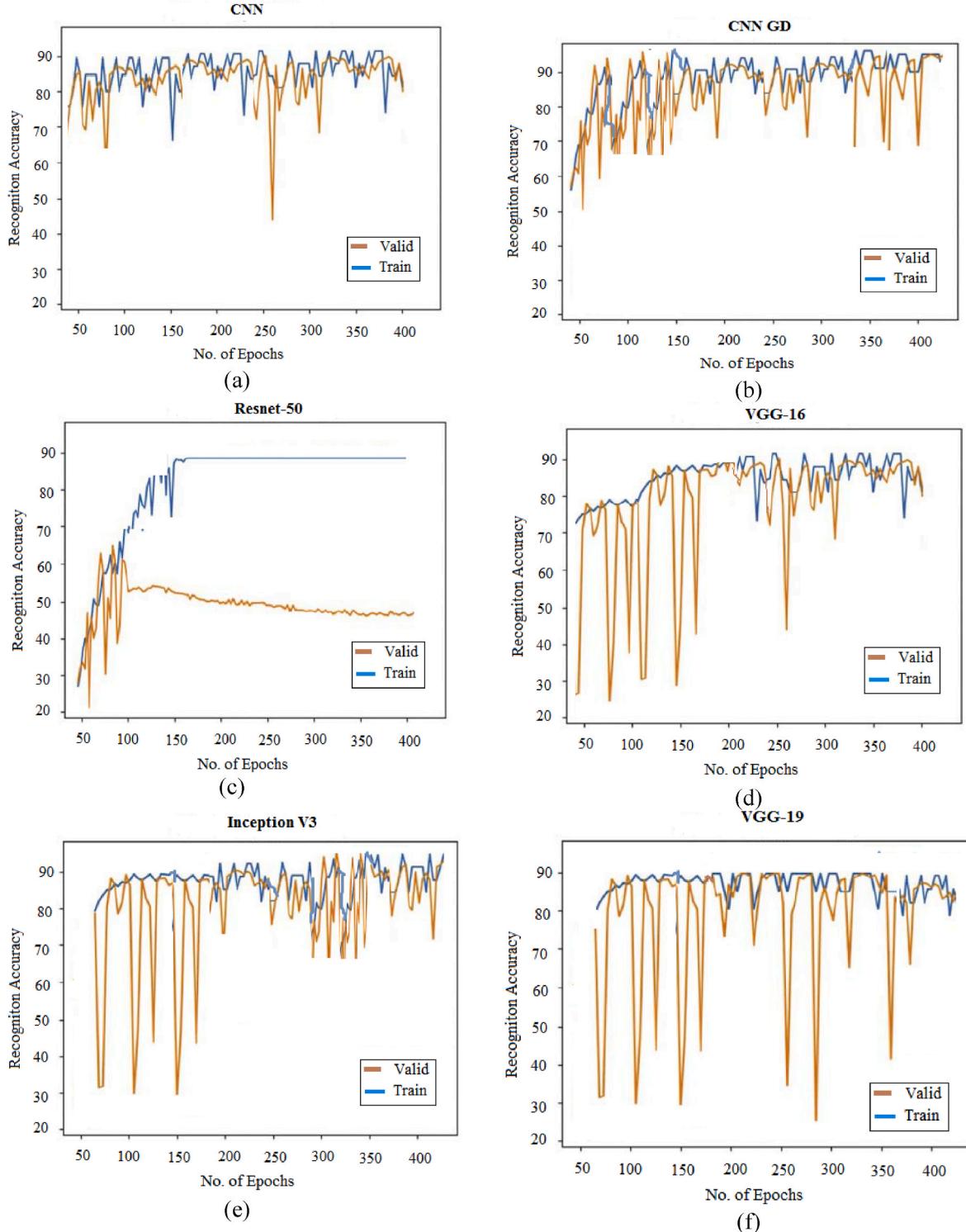


Fig. 8. Recognition accuracy comparison on a) CNN b) CNN GD c) Resnet-50 d) VGG-16 e) Inception V3 f) VGG-19.

3.4. VGG-19

The VGG19 model is slight variation of the VGG model with 19 layers (16 convolution layers, 3 Fully connected layers, 5 MaxPool layers and 1 SoftMax layer). There are 19.6 billion FLOPs in VGG19. This structure is assigned with a fixed size ($224 * 224$) RGB picture as input, indicating that the matrix was of size ($224, 224, 3$). The required pre-processing was to subtract the average RGB value from each pixel, which was

calculated for the whole training set. Kernels are used with a size of $(3 * 3)$ and a stride size of 1 pixel to span the entire image. To maintain the resolution of the image, spatial padding was applied. With stride 2, maximum pooling was achieved over a $2 * 2$ pixel window. This has been accompanied by the Rectified linear unit (ReLu) to bring non-linearity into the method to enhance classification and reduce processing time, while earlier models employed tanh or sigmoid components. Three fully connected layers were constructed, the first two layers have

Table 1
Classification results of various models.

Parameters	CNN	CNN GD	VGG 16	Inception V3	VGG 19	Resnet 50
Accuracy (%)	93.64	97.86	96.52	93.54	92.17	93.47
Precision (%)	89.76	96.39	92.14	90.57	91.46	93.31
Sensitivity (%)	91.70	96.79	93.71	91.73	93.20	91.14
Specificity (%)	88.64	97.40	92.91	92.29	92.00	92.73
FPR (%)	91.42	98.63	93.48	93.75	93.39	95.01
F-Score (%)	92.58	97.96	94.24	95.68	94.78	91.63

been of size 4096, followed by a layer with 1000 channels for 1000-way ILSVRC categorization, and finally a softmax function.

3.5. Resnet-50

One of the most major benefits of the ResNet is that it avoids bad consequences while growing network depth. The fundamental goal of the recurrent neural network is to build a deeper neural network. It is essential for keeping precision and margin of error as one proceeds deeper into the execution of a large number of hidden layers. This is where identification mappings can assist. Continue to learn the residuals to align the anticipated with the actual. Deep learning researchers desire to add so many layers to extract significant information from complex images. As a result, the initial layers may detect edges, while the latter layers may identify recognised objects, such as automobile tyres [25]. However, if the network has more than 30 layers, the efficiency falls and the accuracy drops. This contradicts the belief that adding layers to a neural network will improve its performance. This is not due to over-fitting, because in that instance, dropout and regularisation techniques can be used to eliminate the problem entirely. It's mostly there because of the well-known vanishing gradient issue (see Fig. 5).

4. Results and discussions

The Cancer Imaging Archive provided the dataset for this investigation, which included normal and tumorous CT images of 15,500 patients. The size of the dataset for training includes the real-time data collected from 180 distinct patients. In terms of scanning modes and manufacturers, the dataset is likewise quite diverse. The photos were most likely taken as part of regular treatment rather than as part of a supervised literature review or clinical trial. Experienced professionals analyzed and labelled each item in the dataset. The dataset is freely available. The various images of the lung and its disease categories are shown in Fig. 6. From the total dataset size 93,000 images are used for training the model and 1080 images are used for testing the model. A confusion matrix was built to examine the tumour identification outcomes at the slide stage in detail, providing reliability in terms of percentage for tumour prediction in various-sized metastases. The confusion matrix is computed with four values TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative).

The performance of the proposed approach is analyzed in terms of accuracy, precision, sensitivity, specificity, False Positive Rate (FPR) and F-Score. The values of these parameters are computed as follows. The metrics are computed as follows.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (9)$$

$$\text{FPR} = \frac{FP}{TN + FP} \quad (10)$$

$$AUC = \int_0^1 \frac{TP}{TP + FN} d \frac{FP}{FP + TN} \quad (11)$$

$$F\text{-Score} = \frac{2 \times TP}{(2 \times TP) + FP + FN} \quad (12)$$

In addition to the CT scan images, the histopathological images are also considered for the identification of lung cancer and are shown in Fig. 7. The training and testing dataset are split from the main dataset in the ratio of 3:1 respectively. The various deep learning models are executed against the two datasets and the values of the result are shown in Table 1 (see Fig. 8).

The result analysis of recognition accuracy is denoted with six different models CNN, CNN GD, Resnet-50, VGG-16, Inception V3 and VGG-19. The CNN GD shows a higher accuracy of 97.86%. The higher precision is obtained in the CNN GD model with 96.39% whereas the lower precision is 89.76% in CNN. The higher sensitivity rate is 96.79% whereas the lower sensitivity rate is 91.14% in Resnet-50. The specificity is obtained maximum in CNN GD with 97.40% whereas the lower rate is obtained at 88.64%. The higher False Positive Rate is obtained maximum at 98.63% in CNN GD and the lower FPR is 91.42% in CNN. Similarly, the CNN-GD shows the maximum F-Score of 97.96% whereas 91.63% is obtained for Resnet-50 as the lowest F-Score.

The methods used for the detection of lung cancer in the proposed approach are CNN, CNN GD, Inception V3, Resnet-50, VGG-16 and VGG-19. CNN are effective models for image classification and recognition. CNN GD is the basic algorithm for CNN model optimization. The advantages of using CNN GD are the training data helps these models learn over time, and the cost function within gradient descent specifically acts as a barometer, gauging its accuracy with each iteration of parameter updates. Resnet uses a larger number of layers to train the model and that can be trained easily without increasing the training error percentage. VGG 16 and VGG 19 have faster training speeds, fewer training samples per time, and higher accuracy compared with existing methods. The inception family makes several improvements including using Label smoothing, factorized 7×7 convolutions, and the use of an auxiliary classifier to propagate label information lower down the network. Thus, the usage of all the models in the proposed architecture provides enhanced accuracy with a lower recognition rate.

5. Conclusion

In the proposed approach six deep learning models were built for efficient diagnosis of lung cancer. The methods used for the detection of lung cancer are CNN, CNN GD, Inception V3, Resnet-50, VGG-16 and VGG-19. The experimental analyses were done based on the CT scan images and Histopathological images. The performance of the proposed approach was done based on the recognition accuracy, F-Score, Precision, Sensitivity, Specificity etc. Because of the inherent advantage of the proposed methodology, this will be an efficient method for lung cancer detection and will be beneficial to needed people. The enhanced performance is seen in the CNN GD model with an accuracy of 97.86%. The higher precision is 96.39% and the sensitivity rate is 96.79%. The specificity is obtained maximum in CNN GD with 97.40%. Similarly, the CNN-GD shows the maximum F-Score of 97.96%. In future, the proposed methodology can be enhanced by including fuzzy genetic optimization techniques with the deep learning approaches.

Credit author statement

Vani Rajasekar: Conceptualization, Methodology, Software. **Vaishnnave M:** Data curation, Writing- Original draft preparation. **Prem Kumar S:** Visualization, Investigation. **Rangaraaj V:** Software, Validation.: **Velliangiri Sarveshwaran:** Writing- Reviewing and Editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Velliangiri Sarveshwaran reports administrative support was provided by SRM Institute of Science and Technology. Velliangiri Sarveshwaran reports a relationship with SRM Institute of Science and Technology that includes: employment. Velliangiri Sarveshwaran has patent pending to Assignee. nil.

Data availability

Data will be made available on request.

References

- [1] Y. Xu, A. Hosny, R. Zeleznik, C. Parmar, T. Coroller, I. Franco, H.J. Aerts, Deep learning predicts lung cancer treatment response from serial medical imaging, *Clin. Cancer Res.* 25 (11) (2019) 3266–3275.
- [2] M.I. Faisal, S. Bashir, Z.S. Khan, F.H. Khan, An evaluation of machine learning classifiers and ensembles for early stage prediction of lung cancer, December, in: 2018 3rd International Conference on Emerging Trends in Engineering, Sciences and Technology (ICEEST), IEEE, 2018, pp. 1–4.
- [3] N. Coudray, P.S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. Tsirigos, Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning, *Nat. Med.* 24 (10) (2018) 1559–1567.
- [4] D.M. Ibrahim, N.M. Elshennawy, A.M. Sarhan, Deep-chest: multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases, *Comput. Biol. Med.* 132 (2021), 104348.
- [5] M. Avanzo, J. Stancanello, G. Pirrone, G. Sartor, Radiomics and deep learning in lung cancer, *Strahlenther. Onkol.* 196 (10) (2020) 879–887.
- [6] S.H. Hyun, M.S. Ahn, Y.W. Koh, S.J. Lee, A machine-learning approach using PET-based radiomics to predict the histological subtypes of lung cancer, *Clin. Nucl. Med.* 44 (12) (2019) 956–960.
- [7] P.R. Radhika, R.A. Nair, G. Veena, A comparative study of lung cancer detection using machine learning algorithms, 2019, February, in: IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), IEEE, 2019, pp. 1–4.
- [8] K. Pradhan, P. Chawla, Medical Internet of things using machine learning algorithms for lung cancer detection, *J. Manag. Anal.* 7 (4) (2020) 591–623.
- [9] S. Bhatia, Y. Sinha, L. Goel, Lung cancer detection: a deep learning approach, in: *Soft Computing for Problem Solving*, Springer, Singapore, 2019, pp. 699–705.
- [10] H. Shin, S. Oh, S. Hong, M. Kang, D. Kang, Y.G. Ji, Y. Choi, Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes, *ACS Nano* 14 (5) (2020) 5435–5444.
- [11] V. Rajasekar, B. Predić, M. Saracevic, M. Elhoseny, D. Karabasevic, D. Stanujkic, P. Jayapaul, Enhanced multimodal biometric recognition approach for smart cities based on an optimized fuzzy genetic algorithm, *Sci. Rep.* 12 (1) (2022) 1–11.
- [12] S.H. Hyun, M.S. Ahn, Y.W. Koh, S.J. Lee, A machine-learning approach using PET-based radiomics to predict the histological subtypes of lung cancer, *Clin. Nucl. Med.* 44 (12) (2019) 956–960.
- [13] Y. She, Z. Jin, J. Wu, J. Deng, L. Zhang, H. Su, C. Chen, Development and validation of a deep learning model for non-small cell lung cancer survival, *JAMA Netw. Open* 3 (6) (2020) e205842, e205842.
- [14] W. Rahane, H. Dalvi, Y. Magar, A. Kalane, S. Jondhale, Lung cancer detection using image processing and machine learning healthcare, 2018, March, in: International Conference on Current Trends towards Converging Technologies (ICCTCT), IEEE, 2018, pp. 1–5.
- [15] V. Rajasekar, S. Krishnamoorthi, M. Saracović, D. Pepic, M. Zajmovic, H. Zogic, Ensemble machine learning methods to predict the balancing of ayurvedic constituents in the human body : ensemble machine learning methods to predict, *Comput. Sci.* 23 (1) (2022), <https://doi.org/10.7494/csci.2022.23.1.4315>.
- [16] A. Hosny, C. Parmar, T.P. Coroller, P. Grossmann, R. Zeleznik, A. Kumar, H. J. Aerts, Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study, *PLoS Med.* 15 (11) (2018), e1002711.
- [17] D. Ardila, A.P. Kiraly, S. Bharadwaj, B. Choi, J.J. Reicher, L. Peng, S. Shetty, End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography, *Nat. Med.* 25 (6) (2019) 954–961.
- [18] S.K. Lakshmanaprabu, S.N. Mohanty, K. Shankar, N. Arunkumar, G. Ramirez, Optimal deep learning model for classification of lung cancer on CT images, *Future Generat. Comput. Syst.* 92 (2019) 374–382.
- [19] H. Chao, H. Shan, F. Homayounieh, R. Singh, R.D. Khera, H. Guo, P. Yan, Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography, *Nat. Commun.* 12 (1) (2021) 1–10.
- [20] S. Krishnamoorthi, P. Jayapaul, V. Rajasekar, R.K. Dhanaraj, C. Iwendi, A futuristic approach to generate random bit sequence using dynamic perturbed chaotic system, *Turk. J. Electr. Eng. Comput. Sci.* 30 (1) (2022) 35–49.
- [21] B. Ricciuti, G. Jones, M. Severgnini, J.V. Alessi, G. Recondo, M. Lawrence, M. Awad, Early plasma circulating tumor DNA (ctDNA) changes predict response to first-line pembrolizumab-based therapy in non-small cell lung cancer (NSCLC), *J. Immunother. Cancer* 9 (3) (2021).
- [22] Y. Chen, E. Zitello, R. Guo, Y. Deng, The function of lncRNAs and their role in the prediction, diagnosis, and prognosis of lung cancer, *Clin. Transl. Med.* 11 (4) (2021) e367.
- [23] J. Hanaoka, M. Yoden, K. Hayashi, T. Shiratori, K. Okamoto, R. Kaku, A. Sonoda, Dynamic perfusion digital radiography for predicting pulmonary function after lung cancer resection, *World J. Surg. Oncol.* 19 (1) (2021) 1–10.
- [24] V. Rajasekar, J. Premalatha, K. Sathy, Cancelable Iris template for secure authentication based on random projection and double random phase encoding, *Peer-to-Peer Network. Appl.* 14 (2) (2021) 747–762.
- [25] M. Gaga, J. Chorostowska-Wynimko, I. Horváth, M.C. Tammemagi, D. Shitrit, V. H. Eisenberg, Q. Zhou, Validation of Lung EpiCheck, a novel methylation-based blood assay, for the detection of lung cancer in European and Chinese high-risk individuals, *Eur. Respir. J.* 57 (1) (2021).