

Artificial Neural Networks Improve LDCT Lung Cancer Screening: A Comparative Validation Study

Yin-Chen Hsu

Chiayi Chang Gung Memorial Hospital

Yuan-Hsiung Tsai

Chiayi Chang Gung Memorial Hospital

Hsu-Huei Weng

Chiayi Chang Gung Memorial Hospital

Li-Sheng Hsu

Chiayi Chang Gung Memorial Hospital

Ying-Huang Tsai

Chang Gung Memorial Hospital

Yu-Ching Lin

Chiayi Chang Gung Memorial Hospital

Ming-Szu Hung

Chiayi Chang Gung Memorial Hospital

Yu-Hung Fang

Chiayi Chang Gung Memorial Hospital

Chien-Wei Chen (✉ chienwei33@gmail.com)

<https://orcid.org/0000-0002-5085-9919>

Research article

Keywords: Early detection of cancer, receiver operating characteristic (ROC) curves, sensitivity and specificity, machine learning, data visualization

Posted Date: April 28th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-24642/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on October 22nd, 2020. See the published version at <https://doi.org/10.1186/s12885-020-07465-1>.

Abstract

Background: This study proposes a prediction model for the automatic assessment of lung cancer risk based on an artificial neural network (ANN) with a data-driven approach to the low-dose computed tomography (LDCT) standardized structure report.

Methods: This comparative validation study analysed a prospective cohort from Chiayi Chang Gung Memorial Hospital, Taiwan. In total, 836 asymptomatic patients who had undergone LDCT scans between February 2017 and August 2018 were included, comprising 27 lung cancer cases and 809 controls. A derivation cohort of 602 participants (19 lung cancer cases and 583 controls) was collected to construct the ANN prediction model. A comparative validation of the ANN and Lung-RADS was conducted with a prospective cohort of 234 participants (8 lung cancer cases and 226 controls). The areas under the curves (AUCs) of the receiver operating characteristic (ROC) curves were used to compare the prediction models.

Results: At the cut-off of category 3, the Lung-RADS had a sensitivity of 12.5%, specificity of 96.0%, positive predictive value of 10.0%, and negative predictive value of 96.9%. At its optimal cut-off value, the ANN had a sensitivity of 75.0%, specificity of 85.0%, positive predictive value of 15.0%, and negative predictive value of 99.0%. The area under the ROC curve was 0.764 for the Lung-RADS and 0.873 for the ANN ($P=0.01$). The heatmap plot demonstrates the leading items, i.e., solid nodules, partially solid nodules, and ground-glass nodules, as the significant predictors of malignant outcomes.

Conclusions: Compared to the Lung-RADS, the ANN provided better sensitivity for the detection of lung cancer in an Asian population. In addition, the ANN provided a more refined discriminative ability than the Lung-RADS for lung cancer risk stratification with population-specific demographic characteristics. When lung nodules are detected and documented in a standardized structured report, ANNs may better provide important insights for lung cancer prediction than conventional rule-based criteria.

Trial registration: Not applicable.

Background

Lung cancer is the leading cause of cancer mortality worldwide [1]. The National Lung Screening Trial (NLST) showed that low-dose computed tomography (LDCT) screening could reduce lung cancer mortality by 20% compared to CXR [2]. With the increasing use of LDCT for lung cancer screening, the American College of Radiology (ACR) introduced the Lung Computed Tomography (CT) Screening Reporting and Data System (Lung-RADS), which assigns groups for screening populations [3]. Aimed at high-risk smokers in the USA, the validity of the Lung-RADS remains unclear in areas with a high prevalence of non-smoking-related lung cancer, such as China, Taiwan, and Japan [4]. In Taiwan, more than 95% of lung cancer patients are non-smokers, most of whom have adenocarcinoma [5, 6]. Given the wide range of lung cancer demographics in Asia, the implementation of the Lung-RADS is not yet universal [7]. To address ambiguity, medical institutions have developed various structured reporting

systems [8]. However, there is no current evidence showing explicit superiority for any reporting system in assessing lung cancer risks.

The artificial neural network (ANN) is a field of artificial intelligence technology characterized by simulating biological neural systems based on mathematical theories [9]. ANNs modify their behaviour by adjusting the weights between hidden units until the output correctly converges to the ground truth, and they are particularly adept at classification problems with different input data [10]. With the ability to analyse complex nonlinear relationships between predictors and diseases, well-trained ANNs make predictions with greater accuracy than conventional rule-based criteria [11].

This study aims to propose a reporting system based on an ANN with a data-driven approach to the LDCT standardized structured report. We further explore determinants for predicting lung cancer in this study population.

Methods

Study design and participants

The Institutional Review Board of Chang Gung Medical Foundation approved this study. From February 2017 through August 2018, a total of 836 consecutive asymptomatic participants who underwent both CXR and LDCT at Chiayi Chang Gung Memorial Hospital, Taiwan, for lung cancer screening were prospectively enrolled. The inclusion criteria were age between 40 and 80 years old and willingness to participate in follow-up imaging or diagnostic workup. Subjects were excluded if they had positive CXR findings or a known medical history of any malignant disease. Serial imaging reports, basic patient information, and demographic data were obtained. The diagnosis of lung cancer was confirmed based on surgical resection or lung biopsy and was validated until the index date of July 30, 2019. Patients who did not have confirmed lung cancer based on a hospital-based cancer registry prior to the index date were classified as benign. Figure 1 shows the flowchart of the study.

LDCT image acquisition and interpretation

All LDCT scans were performed with a 64-slice multidetector CT (Somatom Sensation 64; Siemens Healthcare, Erlangen, Germany) in a low-dose setting without contrast enhancement (volumetric CT dose index ≤ 2.0 mGy for a standard patient). The scan parameters were 120 kVp, 25 effective mAs, soft-tissue kernel (B30f), and 3 mm slice thickness. All equipment specifications and acquisition parameters followed the recommendations of the ACR Society of Thoracic Radiology Practice Parameters for the Performance and Reporting of Lung Cancer Screening Thoracic CT [12]. Each LDCT baseline scan was reported by one thoracic radiologist with 7 years of experience. The standardized structured reports described the size, shape, location, and texture of the lung nodules, as well as other incidental findings. The density of each lung nodule was reported according to the definition from the Fleischner Society

guidelines [13, 14]. The size of each lung nodule was measured on lung windows and recorded as recommended by the Lung-RADS.

Development of the ANN

Using data scraping techniques, 22 input features were automatically extracted from the standardized LDCT structured reports to develop an ANN. Four of the inputs were clinical information and LDCT parameters. Another seven inputs consisted of nodule patterns and nodule sizes, which were routinely assessed based on the Lung-RADS standardized lexicon. The remaining inputs were additional interpretations, which consisted of 11 descriptive features transformed by assigning 0 or 1 for binary categories. Table 1 lists all 22 input features of the ANN.

Table 1
Clinical descriptors of the derivation and validation cohorts

Characteristics		Derivation cohort	Validation cohort	<i>P</i>
		(N = 602)	(N = 234)	
Sex ^a	Male, n (%)	243 (40.37%)	113 (48.29%)	0.038 ^b
	Female, n (%)	359 (59.63%)	121 (51.71%)	
Age (y) ^a		61.96 ± 6.47	60.93 ± 7.92	0.053
LDCT parameters	Dose (mSv) ^a	1.47 ± 0.28	1.50 ± 0.27	0.161
	DLP (mGy·cm) ^a	50.00 ± 13.11	51.24 ± 11.62	0.206
Pattern of nodules	Nodules of interest, n (%) ^a	1.15 (0–32)	1.31 (0–8)	0.330
	Number of involved lobes, n (%) ^a	0.78 (0–5)	1.00 (0–5)	0.007 ^b
Size of nodules	Solid nodule (mm) ^a	1.76 (0-136.00)	1.76 (0-36.75)	1.000
	PS nodule (mm) ^a	0.49 (0-20.40)	0.58 (0-7.30)	0.498
	GGN (mm) ^a	0.84 (0–31.00)	0.39 (0-10.30)	0.038 ^b
	Calcified nodule (mm) ^a	0.38 (0-19.25)	0.56 (0-7.05)	0.100
	Fat-containing nodule (mm) ^a	0.05 (0-28.15)	0.00 (0)	0.506

^a The 22 input items for developing the ANN.

^b The *P*-values indicated a significant difference between the training and internal validation groups and were obtained by t-test and chi-square test.

BMI, body mass index; DLP, dose length product; GGN, ground-glass nodule; PS nodule, part-solid nodule.

The values are given as the mean ± SD, range or n (%).

Characteristics		Derivation cohort	Validation cohort	<i>P</i>
Intra-pulmonary findings	Linear atelectasis, n (%) ^a	441 (73.26%)	110 (47.00%)	< 0.001 ^b
	Plate-like atelectasis, n (%) ^a	78 (12.96%)	19 (8.12%)	0.050
	Plate-like GGN, n (%) ^a	145 (24.09%)	40 (17.09%)	0.029 ^b
	Bronchiectasis, n (%) ^a	39 (6.48%)	9 (3.85%)	0.143
	Emphysema, n (%) ^a	52 (8.64%)	30 (12.82%)	0.068
	Fibrotic change, n (%) ^a	156 (25.91%)	42 (17.95%)	0.015 ^b
Extra-pulmonary findings	Mediastinal tumour, n (%) ^a	34 (5.65%)	9 (3.85%)	0.290
	Thyroid nodule, n (%) ^a	20 (3.32%)	2 (0.85%)	0.045 ^b
	Adrenal nodule, n (%) ^a	6 (1.00%)	0 (0.00%)	0.125
	Hepatic nodule, n (%) ^a	68 (11.30%)	20 (8.55%)	0.245
	Renal nodule, n (%) ^a	16 (2.66%)	10 (4.27%)	0.229
Lung-RADS	Category 1	323 (53.66%)	115 (49.14%)	0.240
	Category 2	228 (37.87%)	109 (46.58%)	0.021 ^b
	Category 3	36 (5.98%)	7 (3.00%)	0.080
	Category 4	15 (2.49%)	3 (1.28%)	0.279
^a The 22 input items for developing the ANN.				
^b The <i>P</i> -values indicated a significant difference between the training and internal validation groups and were obtained by t-test and chi-square test.				
BMI, body mass index; DLP, dose length product; GGN, ground-glass nodule; PS nodule, part-solid nodule.				
The values are given as the mean ± SD, range or n (%).				

Feed-forward neural networks based on the back-propagation algorithm were constructed using Keras version 2.2.4 [15], a high-level neural network application programming interface that can simplify the ANN construction process. The ANN consisted of the first two hidden layers, followed by a dropout layer to prevent over-fitting and a dense layer as the output layer [16]. The performances of the prediction models were monitored during training to achieve optimal tuning of the hyperparameters. Figure 2 shows the structure of an ANN.

Validation and risk group identification

In the training process, the ANN was internally validated via “three-fold cross-validation” [17]. The dataset was divided into three equal parts. At each cycle, one of the three parts was selected as the test set and removed from the dataset, while the remaining cases were used as the training set of the ANN. This process was repeated until the entire dataset had been used once as the test set. Finally, the ANN was validated with the prospective validation cohort.

To investigate the determining factors for predicting lung cancer, the feature importance was evaluated by visualizing the weights connecting each input unit to each hidden unit in the first layer [18]. By transforming these weight values to a colour scale, the weight values for each input feature were presented as light (positive value) or dark (negative value) spots. The significant predictors for predicting lung cancer were highlighted based on their weights with large absolute values.

Statistical analyses

Statistical analyses were performed using MedCalc 18.9.1 (MedCalc Software, Ostend, Belgium). Observed distributions were tested against the hypothesized normal distribution (Kolmogorov–Smirnov test). Data are reported as the mean \pm standard deviation or number (%) unless otherwise indicated. To determine and compare the performance of the Lung-RADS and ANN, the sensitivity and specificity of the lung cancer classification at different thresholds were analysed based on area under the receiver operating characteristic (AUC-ROC) curve analyses. The optimal diagnostic thresholds of the ROC curves were determined using maximized Youden’s [19] index. ROC curves were compared using the method described by DeLong et al. [20]. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+), and negative likelihood ratio (LR–) of each model for lung cancer diagnosis were calculated [21]. In all analyses, $P < 0.05$ was considered to indicate statistical significance.

Results

Demographic and clinical characteristics

The study cohort included a total of 836 consecutive asymptomatic participants who had undergone LDCT for lung cancer screening (27 lung cancer cases and 809 controls) at our institution. Between February 2017 and February 2018, 602 participants were included in the derivation cohort. Among the participants in the derivation cohort, 29 subjects underwent surgical resection or biopsy for tissue sampling. Nineteen of those subjects were diagnosed with lung cancer (adenocarcinoma, $n = 19$), and the remaining ten had benign lesions (pneumonia, $n = 5$; pulmonary fibrosis, $n = 4$; and pulmonary hamartoma, $n = 1$). Between March and August 2018, 234 participants were included in the validation cohort. Nine of these subjects underwent tissue sampling, eight of whom were diagnosed with lung cancer (adenocarcinoma, $n = 8$); the remaining subjects had benign lesions (pulmonary fibrosis, $n = 1$). Despite the adoption of identical inclusion criteria, there were several significant differences in

demographic features between the training and validation cohorts. The full demographic and clinical descriptions of each cohort are presented in Table 1.

For the derivation cohort (n = 602), the distribution of baseline Lung-RADS categories was as follows: category 1 (53.66%), category 2 (37.87%), category 3 (5.98%), and category 4 (2.49%). Among the subjects in this cohort, the 19 lung cancer participants (3.16%) included 6 with category 2, 5 with category 3, and 8 with category 4; none had category 1. For the validation cohort (n = 234), the distribution of baseline Lung-RADS categories was as follows: category 1 (49.14%), category 2 (46.58%), category 3 (3.00%), and category 4 (1.28%). Among the subjects in this cohort, the 8 lung cancer participants (3.42%) included 7 with category 2 and 1 with category 3; none had category 1 or category 4.

Performance of prediction models

Table 2 presents the contingency results of both lung cancer assessment models. Most of the non-cancer cases were correctly identified by both the Lung-RADS and ANN (specificity: 96.0% and 85.0%, respectively), but more lung cancer cases were correctly identified by the ANN (sensitivity: 12.5% and 75.0%, respectively). Figure 3 presents the ROC curves and AUCs for assessing the overall validity of both tools. There was a significant difference between the AUCs of the Lung-RADS and ANN (AUC 0.764 vs. 0.873, respectively, $P = 0.013$). Table 3 presents the sensitivity, specificity, PPV, NPV, LR+, and LR- of the two risk assessment tools. For Lung-RADS, a positive predictive value of 10.0% (95% CI: 1.6 to 43.7%) and negative predictive value of 96.9% (95% CI: 96.0 to 97.6%) were calculated at the cut-off point of category 3, which adhered to the original definition of a “positive scan”. For the ANN, a positive predictive value of 15.0% (95% CI: 9.6 to 22.6%) and negative predictive value of 99.0% (95% CI: 96.7 to 99.7%) were calculated at the optimal cut-off value. The likelihood ratios confirm that the results according to both lung cancer risk classification tools differ from those according to chance.

Table 2
Contingency table for the Lung-RADS and ANN models (n = 234)

Scale/model	Lung-RADS			ANN		
	No	Yes	Sum	No	Yes	Sum
Disease (-)	217	9	226	192	34	226
Disease (+)	7	1	8	2	6	8
Sum	224	10	234	194	40	234

Table 3
Performance analysis for the Lung-RADS and ANN models (n = 234)

Scale/model	Lung-RADS	ANN
Cut-off	Category 3	> 0.012
AUC (95% CI)	0.764 (0.705, 0.817)	0.873 (0.823, 0.913)
Classification accuracy (%)	93.16	84.62
Sensitivity (95% CI)	12.50 (0.3, 52.7)	75.00 (34.9, 96.8)
Specificity (95% CI)	96.02 (92.6, 98.2)	84.96 (79.6, 89.4)
PPV (95% CI)	10.0 (1.6, 43.7)	15.0 (9.6, 22.6)
NPV (95% CI)	96.9 (96.0, 97.6)	99.0 (96.7, 99.7)
LR+ (95% CI)	3.14 (0.5, 21.9)	4.99 (3.0, 8.3)
LR- (95% CI)	0.91 (0.7, 1.2)	0.29 (0.1, 1.0)
AUC, area under the curve; CI, confidence interval; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.		

Feature importance and risk group identification

Figure 4 shows a heatmap visualizing the feature importance of the ANN. In this plot, the rows correspond to the 22 input items, while the columns correspond to the weights connecting the inputs to the 10 hidden units in the first layer of the ANN. The significant predictors are highlighted by outliers in the weight values and can be recognized by the abruptly strong contrast to the other features. Three of the features, i.e., solid nodules, partially solid nodules, and ground-glass nodules (GGNs), indicated with dark bars in the plot, were potential predictors of malignant outcomes. By contrast, the presence of calcified nodules, indicated with a light bar in the plot, was recognized as a potential benign predictor.

Discussion

In lung cancer screening, LDCT is used to detect pulmonary nodules and evaluate their size and morphology. Most pulmonary nodules are small (< 5 mm in diameter) and benign, and their morphology is variable [22]. Across the lung cancer screening literature, the major challenge faced by this diagnostic imaging modality is the difficulty of defining a “positive scan [23, 24].” The false-positive rate of the Lung-RADS has increased due to the large degree of variation in lung cancer demographics between populations, thus limiting the reliability of this tool [25]. In addition, application of the unitary criteria without appropriate validation may result in false-positive results, overdiagnosis, and unnecessary costs [26]. In this study, the Lung-RADS predicted lung cancer risks for the validation cohort with an AUC of 0.76, which indicated suboptimal decisive power to assess lung cancer risks in the population. The principles of the Lung-RADS are uniformity of radiology interpretation, risk assessment, and nodule management in LDCT lung cancer screening programmes, and although the clinical presentations of lung

cancer are likely to vary greatly between populations, some of these imaging findings are not assessed. One possible remedy for this obstacle is the development of a validated prediction model for lung cancer risk using artificial intelligence algorithms, such as ANNs. In this study, the ANN took many risk factors into account, and it predicted lung cancer risks for the validation cohort with an AUC of 0.87. Among high-risk groups, overdiagnosis and unnecessary procedures might be avoided when patients are identified correctly by ANNs. Compared to the Lung-RADS, ANNs may be more robust in the prediction of lung cancer. Additionally, the standardized structured reports in this study involved the use of lung nodule descriptions from the Lung-RADS lexicon suggested by the ACR. As these input features can be easily identified and are generally assessed by radiologists, the ANN-based LDCT reporting system is both cost-effective and user-friendly.

Along with risk group identification, we investigated the predictors of lung cancer that are potentially useful for identifying patients at high risk for lung cancer. According to the heatmap derived from the ANN, three features, i.e., solid nodules, partially solid nodules, and GGNs, were identified as significant predictors of malignant outcomes. In conformity by the NLST and Lung-RADS criteria, there is a strong implication that the ANN predicts lung cancer mainly based on the documented nodule size in each category. Furthermore, this study addressed the diversity of lung cancer risk assessments in populations with a high percentage of non-smoking-related lung cancer. Among the subjects in this study, more than one-third of the confirmed lung cancer lesions presented with GGNs < 20 mm (5 of 19 lung cancer cases in the derivation cohort and 5 of 8 lung cancer cases in the validation cohort). When the Lung-RADS was applied, these patients were classified as category 2 and may have been falsely reassured by the “negative” screening results and thus did not return for follow-up scans. Among the 5 of 8 lung cancer cases in the validation cohort, the ANN could identify all (100%) of these patients who had pulmonary lesions and initially presented with GGNs < 20 mm, which were finally confirmed as adenocarcinoma. In several studies performed in Asian cohorts, the majority of lung cancer patients were non-smokers with pulmonary adenocarcinoma spectrum lesions, which typically presented as pure GGNs or partially solid nodules [27, 28]. The current literature shows that larger GGNs (variable cut-off, range 10.5 ~ 15.0 mm) tend to be more aggressive or appear as invasive pulmonary adenocarcinoma [29, 30]. This is a particular concern in Asian populations, where it would be important to report these GGNs and develop corresponding algorithms with follow-up strategies. Therefore, the ANN potentially assimilates population-specific demographic characteristics and provides important insights that improve the efficacy of lung cancer screening programmes.

There were several limitations to this study. First, classification models based on machine learning tend to be unstable in small datasets. Therefore, both models in this study were externally validated using a prospective cohort. Second, the positive and negative predictive values were influenced by the prevalence of disease in the study population. The prevalence of inpatient falls being estimated as 3% is a rough estimate as mentioned above and is therefore arbitrary to some extent. Finally, the short follow-up period may have caused partial verification. A large-scale prospective study with long-term follow-up is required to explore the benefits of using an ANN as part of an LDCT lung cancer screening programme.

Conclusions

Compared to the Lung-RADS, the ANN may have substantially improved the sensitivity for the detection of lung cancer in an Asian population. Furthermore, ANNs have a more refined discriminative ability than the Lung-RADS for lung cancer risk stratification with population-specific demographic characteristics. When lung nodules are detected and documented in a standardized structured report, ANNs may better provide important insights for lung cancer prediction than conventional rule-based criteria. The effects of using an ANN in clinical practice must be examined carefully in further prospective large cohort studies.

Abbreviations

ACR = American College of Radiology; ANN = artificial neural network; AUC = area under the curve; GGN = ground-glass nodule; LDCT = low-dose computed tomography; Lung-RADS = Lung CT Screening Reporting and Data System; NLST = National Lung Screening Trial; NPV = negative predictive value; PPV = positive predictive value; ROC = receiver operating characteristic

Declarations

Ethics approval and consent to participate

The study was approved by the Institutional Review Board (IRB) of Chang Gung Medical Foundation, in accordance with the ethical standards of the responsible committee on human experimentation (IRB Nos. 201801905B0). Written consents were obtained from study participants.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Chang Gung Memorial Hospital. (Contract Nos. PMRPG6F0021 and CMRPG6H0651). The funding body did not play any role in design, in the collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.

Authors' contributions

YCH conceptualized the study, performed analysis and wrote the main manuscript text. CWC assisted with the writing of the manuscript and prepared all figures. YHT oversaw the project. LSH and HHW assisted with study design and statistical analysis. YHT, YCL, MSH and YHF assisted with collection of patients' meta data and interpretation. All authors read and approved the final manuscript.

Acknowledgements

We acknowledge Springer Nature Author Service for editing this manuscript.

References

1. American Cancer Society. Cancer Facts & Figs. 2018. Atlanta: American Cancer Society; 2018.
2. National Lung Screening Trial Research Team. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365:395–409.
3. Lung CT screening reporting and data system (Lung-RADS). American College of Radiology. 2014. <https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads>. Accessed 1 Dec 2018.
4. Detterbeck FC, Marom EM, Arenberg DA, Franklin WA, Nicholson AG, Travis WD, et al. The IASLC lung cancer staging project: background data and proposals for the application of TNM staging rules to lung cancer presenting as multiple nodules with ground glass or lepidic features or a pneumonic type of involvement in the forthcoming eighth edition of the TNM classification. *J Thorac Oncol*. 2016;11:666–80.
5. Chen KY, Chang CH, Yu CJ, Kuo SH, Yang PC. Distribution according to histologic type and outcome by gender and age group in Taiwanese patients with lung carcinoma. *Cancer*. 2005;103:2566–74.
6. Ha SY, Choi SJ, Cho JH, Choi HJ, Lee J, Jung K, et al. Lung cancer in never-smoker Asian females is driven by oncogenic mutations, most often involving EGFR. *Oncotarget*. 2015;6:5465–74.
7. Carter BW, Lichtenberger JP 3rd, Wu CC, Munden RF. Screening for lung cancer: lexicon for communicating with health care providers. *AJR Am J Roentgenol*. 2018;210:473–9.
8. Hsu HT, Tang EK, Wu MT, Wu CC, Liang CH, Chen CS, et al. Modified lung-RADS improves performance of screening LDCT in a population with high prevalence of non-smoking-related lung

- cancer. *Acad Radiol*. 2018;25:1240–51.
9. Bishop CM. *Neural networks for pattern recognition*. New York: Oxford University Press; 1995. 482 p.
 10. Baker JA, Kornguth PJ, Lo JY, Williford ME, Floyd CE Jr. Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon. *Radiology*. 1995;196:817–22.
 11. Weng SF, Reys J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*. 2017;12:e0174944.
 12. Kazerooni EA, Austin JH, Black WC, Dyer DS, Hazelton TR, Leung AN, et al. ACR–STR practice parameter for the performance and reporting of lung cancer screening thoracic computed tomography (CT): 2014 (Resolution 4). *J Thorac Imaging*. 2014;29:310–6.
 13. MacMahon H, Austin JH, Gamsu G, Herold CJ, Jett JR, Naidich DP, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner society. *Radiology*. 2005;237:395–400.
 14. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: From the fleischner society 2017. *Radiology*. 2017;284:228–43.
 15. Chollet F. Keras. GitHub; <https://github.com/fchollet/keras%7D%7D>; 2015.
 16. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15:1929–58.
 17. James G, Witten D, Hastie T, Tibshirani R. *An introduction to statistical learning*. New York: Springer; 2013.
 18. Müller AC, Guido S. *Introduction to machine learning with python: a guide for data scientists*. California: O'Reilly Media, Inc; 2016.
 19. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32–5.
 20. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–45.
 21. Sheskin DJ. *Handbook of parametric and nonparametric statistical procedures*. New York: CRC Press; 2003.
 22. van Riel SJ, Sanchez CI, Bankier AA, Naidich DP, Verschakelen J, Scholten ET, et al. Observer variability for classification of pulmonary nodules on low-dose CT images and its effect on nodule management. *Radiology*. 2015;277:863–71.
 23. Gierada DS, Pilgram TK, Ford M, Fagerstrom RM, Church TR, Nath H, et al. Lung cancer: interobserver agreement on interpretation of pulmonary findings at low-dose CT screening. *Radiology*. 2008;246:265–72.
 24. Balata H, Evison M, Sharman A, Crosbie P, Booton R. CT screening for lung cancer: are we ready to implement in Europe? *Lung Cancer*. 2019;134:25–33.
 25. Haiman CA, Stram DO, Wilkens LR, Pike MC, Kolonel LN, Henderson BE, et al. Ethnic and racial differences in the smoking-related risk of lung cancer. *N Engl J Med*. 2006;354:333–42.

26. Patz EF Jr, Pinsky P, Gatsonis C, Sicks JD, Kramer BS, Tammemagi MC, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med.* 2014;174:269–74.
27. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers—a different disease. *Nat Rev Cancer.* 2007;7:778–90.
28. Saito S, Espinoza-Mercado F, Liu H, Sata N, Cui X, Soukiasian HJ. Current status of research and treatment for non-small cell lung cancer in never-smoking females. *Cancer Biol Ther.* 2017;18:359–68.
29. Jin X, Zhao SH, Gao J, Wang DJ, Wu J, Wu CC, et al. CT characteristics and pathological implications of early stage (T1N0M0) lung adenocarcinoma with pure ground-glass opacity. *Eur Radiol.* 2015;25:2532–40.
30. Lee HY, Choi YL, Lee KS, Han J, Zo JI, Shim YM, et al. Pure ground-glass opacity neoplastic lung nodules: histopathology, imaging, and management. *AJR Am J Roentgenol.* 2014;202:W224-33.

Figures

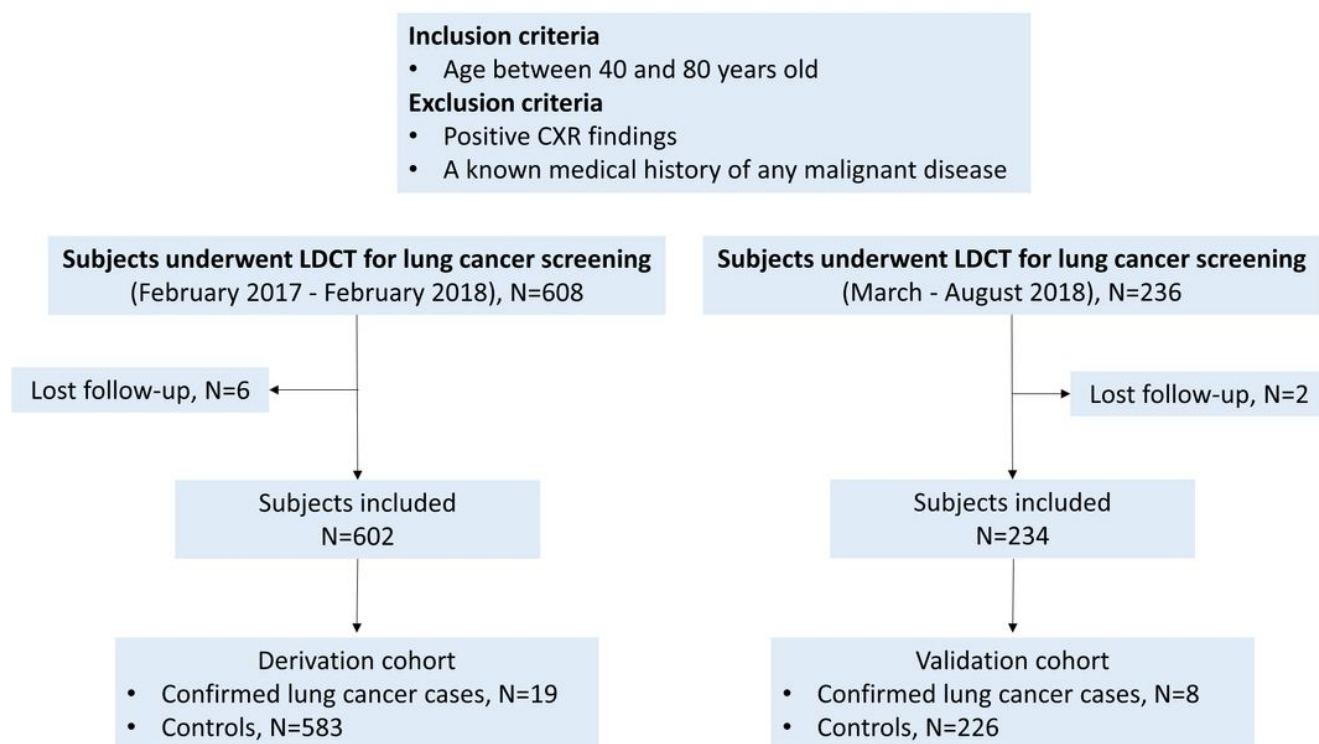


Figure 1

Flow diagram.

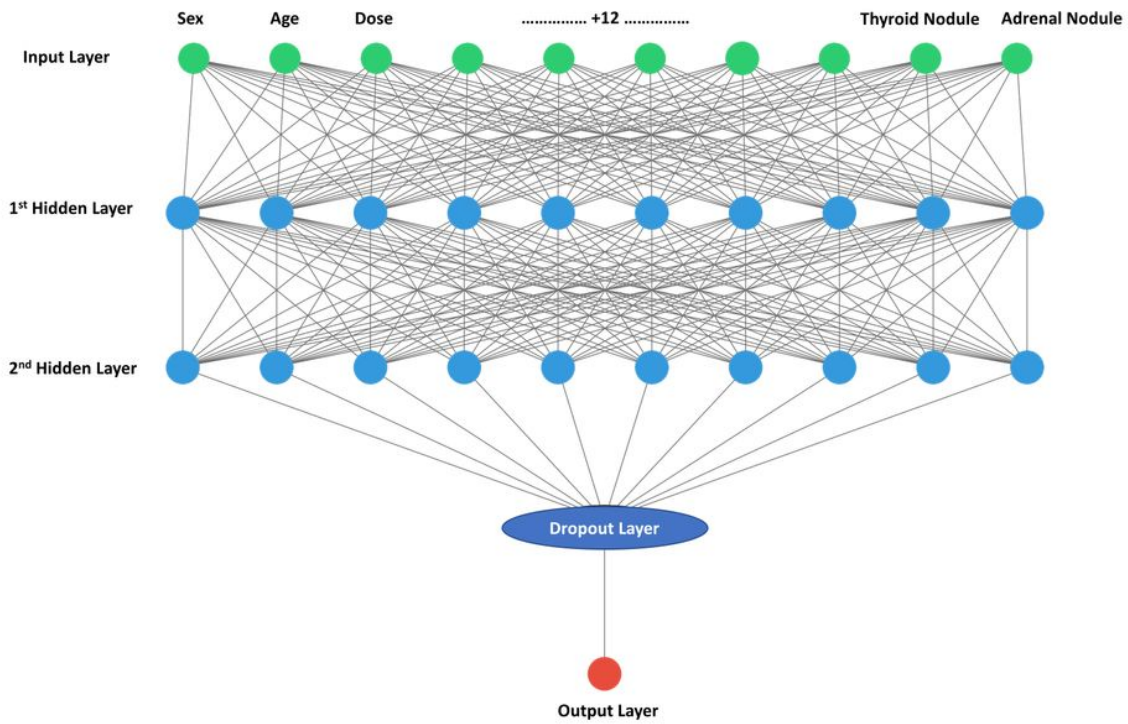


Figure 2

Structure of an ANN.

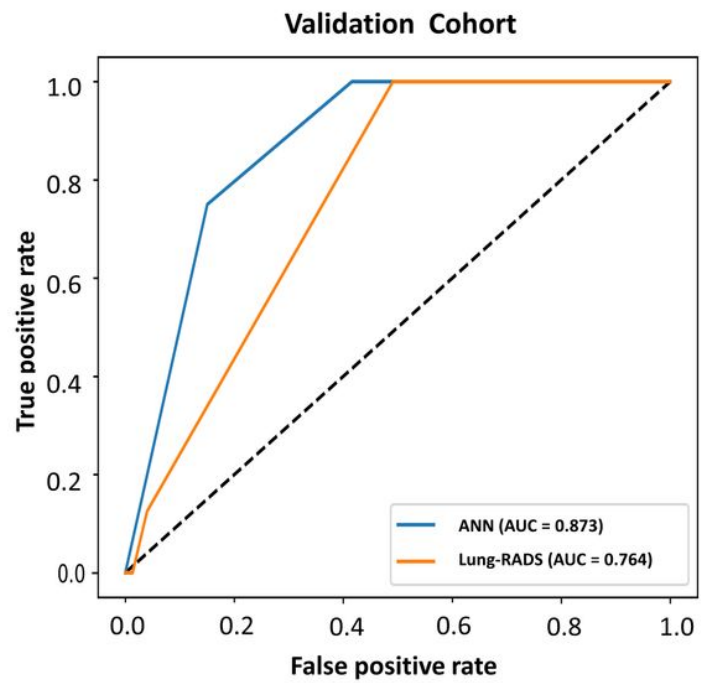


Figure 3

ROC curves for the Lung-RADS and ANN model.

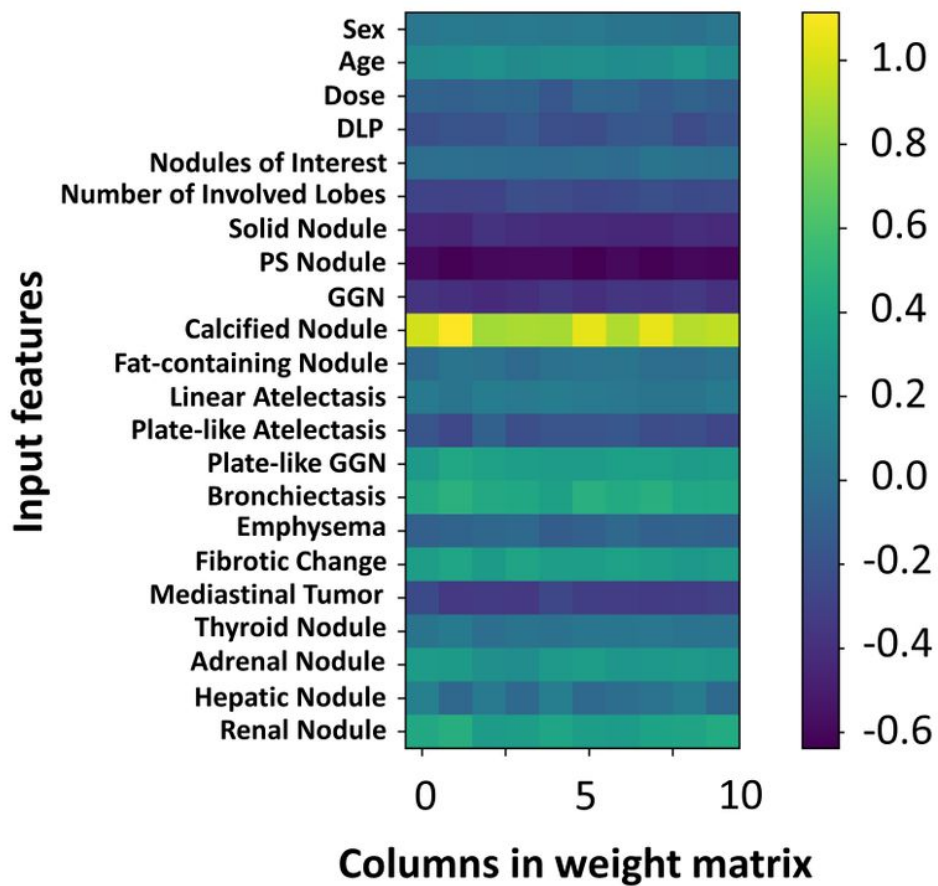


Figure 4

The heatmap plot visualizing feature importance of the ANN model.