Lung cancer risk prediction models based on pulmonary nodules: A systematic review

Zheng Wu | Fei Wang | Wei Cao | Chao Qin | Xuesi Dong | Zhuoyu Yang | Yadi Zheng | Zilin Luo | Liang Zhao | Yiwen Yu | Yongjie Xu | Jiang Li | Wei Tang | Sipeng Shen | Ning Wu | Fengwei Tan | Ni Li | Jie He

1Office of Cancer Screening, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
2Chinese Academy of Medical Sciences Key Laboratory for National Cancer Big Data Analysis and Implement, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
3PET-CT Center, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
4Department of Epidemiology, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China
5Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University, Nanjing, China
6Department of Diagnostic Radiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
7Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Correspondence
Fengwei Tan, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, No.17 Panjiayuannanli, Chaoyang District, Beijing, 100021, China. Email: tanfengwei@cicams.ac.cn
Ni Li, Office of Cancer Screening, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College; Chinese Academy of Medical Sciences Key Laboratory for National Cancer Big Data Analysis and Implement; No.17 Panjiayuannanli, Chaoyang District, Beijing, 100021, China. Email: nli@cicams.ac.cn
Jie He, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, No.17 Panjiayuannanli, Chaoyang District, Beijing, 100021, China. Email: hejie@cicams.ac.cn

Abstract

Background: Screening with low-dose computed tomography (LDCT) is an efficient way to detect lung cancer at an earlier stage, but has a high false-positive rate. Several pulmonary nodules risk prediction models were developed to solve the problem. This systematic review aimed to compare the quality and accuracy of these models.

Methods: The keywords “lung cancer,” “lung neoplasms,” “lung tumor,” “risk,” “lung carcinoma,” “risk,” “predict,” “assessment,” and “nodule” were used to identify relevant articles published before February 2021. All studies with multivariate risk models developed and validated on human LDCT data were included. Informal publications or studies with incomplete procedures were excluded. Information was extracted from each publication and assessed.

Results: A total of 41 articles and 43 models were included. External validation was performed for 23.2% (10/43) models. Deep learning algorithms were applied in 62.8% (27/43) models; 60.0% (15/25) deep learning based researches compared their algorithms with traditional methods, and received better discrimination. Models based on Asian and Chinese populations were usually built on single-center or small sample retrospective studies, and the majority of the Asian models (12/15, 80.0%) were not validated using external datasets.

Conclusion: The existing models showed good discrimination for identifying high-risk pulmonary nodules, but lacked external validation. Deep learning algorithms are increasingly being used with good performance. More researches are required.
INTRODUCTION

Lung cancer causes a significant burden on health care systems. In 2020, lung cancer resulted in the death of 1.8 million people worldwide. In China, lung cancer remains the most commonly diagnosed cancer and the leading cause of cancer death.1

The overall 5-year survival rate of lung cancer ranges from 10% to 20% in most countries.2 However, the prognosis of lung cancer largely depends on the stage of the disease at diagnosis. Although the 5-year survival rate of lung cancer at stage I is above 80%, it is close to 0% for stage IV disease.3 Therefore, early diagnosis and treatment are important to reduce mortality from lung cancer, improve the quality of life and reduce the economic burden from this disease.

Screening with low-dose computed tomography (LDCT) has been shown to be an efficient way to detect lung cancer at an earlier stage and reduce lung cancer mortality.4 Several lung cancer screening trials have been conducted worldwide.4–9 The national lung cancer screening trial (NLST) of the United States has shown that early LDCT screening can detect potentially cancerous lung nodules at an early stage leading to a reduction in lung cancer mortality by 20%. Nevertheless, the false-positive nodule detection rate by LDCT was extremely high at 96.4%,4 eventually leading to unnecessary radiation exposure from further follow-up imaging tests, invasive biopsies, medical expenses, and anxiety among patients.5 Therefore, it is of paramount importance to identify the individuals at higher risk of developing lung cancer based on the pulmonary nodules identified on LDCT scans to recommend appropriate examination and management.

Further examinations in current lung cancer screening programs are recommended solely based on the nodule sizes on the LDCT scans. However, although this method of categorizing pulmonary nodules is easy to implement clinically, it may lead to a high rate of false-positive results. On the contrary, risk prediction models based on pulmonary nodule size, calcification, density, and other relevant imaging information may facilitate the identification of high-risk groups, significantly reduce the false positive rate, and improve the screening program's efficiency.7 Therefore, this method is now recommended by several clinical guidelines to reduce the high false-positive rate of LDCT screening.8,9

As a result, several statistical models have been developed in recent years to predict the risk of developing lung cancer based on the identification of pulmonary nodules on LDCT. However, without a systematic evaluation of the relevant models, it remains unclear which, if any of these models should be used clinically. Therefore, in this study, we reviewed the contemporary published literature to identify current multivariable statistical models used to predict the risk of developing lung cancer from the pulmonary nodules identified on LDCT. In addition, the effectiveness, reliability, bias, and extrapolation of the different models used in these studies were also compared.

METHODS

Search strategy

A literature search was conducted using the PubMed, Cochrane, Embase, and Web of Science electronic databases. The keywords “lung cancer” or “lung neoplasms” or “lung tumor” or “lung carcinoma” and “predict” or “assessment” or “risk” and “nodule” were used to identify all relevant articles published in English from January 1960 to February 2021. We also hand-searched the reference lists of eligible studies to identify additional relevant publications. Further detail about the search strategy used in this study is available in Table S1.

Review methods and selection criteria

Two reviewers independently screened all titles and abstracts and made decisions regarding the potential eligibility of the research articles for full text review. Discrepancies in judgment were resolved by a third reviewer. Studies were eligible if they reported on the development of multivariable risk prediction models for the development of lung cancer based on the pulmonary nodules identified on LDCT and included a detailed description of the procedures used to evaluate and validate the model. Studies with an incomplete description of the procedures used to develop, validate, and evaluate the model were excluded. Informal publications such as conference abstracts were also excluded.

Data extraction

The models used in the studies were divided into two categories; traditional and deep learning models. In the traditional models, raw data (i.e., original image features) were translated into a finite number of feature descriptors (i.e., size, type, or density of nodules) that could be used as predictors for lung cancer. The association between lung cancer risk and each descriptor was tested, quantified, and subsequently developed into an appropriate statistical risk model. In the deep learning algorithm-based models, the use of raw data was allowed and representations needed for detection or classification were automatically discovered,
and the association between lung cancer risk and descriptors is partly unexplainable.\textsuperscript{10,11}

For each of the included studies, basic information about the research methodology, variables used to develop the models, and the methods used to evaluate the models were extracted. The basic information included the first author, publication year, study design, study method, target population, inclusion criteria of participants and nodules, and the number of normal and lung cancer cases used for modeling. The model variables extracted from the studies included: basic information about the clinical and epidemiological characteristics, such as age, sex, smoking, family history, occupational exposure, or history of chronic respiratory diseases; and imaging nodule characteristics, like size, density or shape; other tumor biomarkers like neuron-specific enolase (NSE), or carcinoembryonic antigen (CEA). For the studies based on the deep learning algorithm, it was not possible to extract these variables because of the method used to develop the risk model. The model evaluation criteria included the type of validation (external or internal), the sample size used for verification, the area under the curve (AUC), model calibration slope results, sensitivity, specificity, and the risk threshold. The findings of either the Hosmer-Lemeshow test or the expected to observe ratio (excellent, poor, or uncalibrated) were also recorded. Furthermore, we used the same dataset to compare the performance (AUC, sensitivity, or specificity) of all deep learning models with existing prediction methods or clinically based guidelines published by professional bodies such as the American College of Radiology Lung Imaging Reporting and Data System (ACR Lung-RADS) based on the conclusion in the original text.

Quality assessment

The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) method\textsuperscript{12} was used to evaluate the quality of evidence in traditional models. This method assesses the quality of the publication based on the risk of bias, consistency, accuracy, directness, and publication bias.

Data synthesis

The sample size used in each study was recorded when available and estimated for evaluation purposes when not available. If several models were used to train the algorithm on the same data set, the model with the highest AUC was selected.

Limited statistical power may lead to insufficient power to detect a significant association, resulting in unstable models. To overcome this problem, we calculated the events per variable (EPV) for traditional models. EPV was defined as the number of events divided by the number of predictor variables included in the multivariable model. An EPV value <10 suggests limited statistical power.\textsuperscript{13} Because it was not possible to record and name the variables used in the deep learning models,\textsuperscript{11} the EPV could not be calculated.

RESULTS

Study characteristics and quality assessment

The literature search revealed a total of 3230 publications, of which 630 were found to be duplicated and were, therefore, removed from the evaluation. A total of 2293 articles that did not meet our criteria were excluded from the screening. After evaluating the full texts of the remaining 307 articles, 41 articles met the eligibility criteria and were included for further analysis (Figure 1).

After evaluating the articles, 43 models were identified. Overall the models were based on more than 20 000 Asian, North American, and European participants (Figure 2(a)). After 2018, the number of relevant studies grew rapidly. As a result, over half (67.4%, 29/43) of all models were released after 2018 (Figure 3).

Most models (58.1%, 25/43) were developed based on deep learning algorithms, and the remaining (41.9%, 18/43) were developed using traditional models (Figure 2(b)) such as logistic regression. However, in recent years, the use of deep learning algorithms increased significantly (Table 2).

Only 23% (10/43) of the models were externally validated (Figure 2(c)). Data from multiple sources were used to develop the models in half of the studies (Figure 2(d)). Thirty-three studies used data from cohort studies to develop the models, whereas in eight studies, the models were constructed using the data from screening trials (Tables 3 and 4). Almost all studies (97.6%, 40/41) had
FIGURE 2  Characters of existing models; (a) size and distribution of training sets used for modeling; (b) number and distribution of existing models; (c) number and distribution of models seeking validation in different ways; (d) number and distribution of models from different regions and data sources; and (e) frequency of risk factors used in traditional final models.

FIGURE 3  AUCs and confidence intervals of existing models by regions and time periods.
| First author          | Year | Study design | Study method | Target population | Inclusion criteria of participants | Sample size | Cases of lung cancer | EPV | Data source |
|----------------------|------|--------------|--------------|-------------------|-----------------------------------|-------------|----------------------|-----|-------------|
| Annette McWilliams   | 2013 | Screen trial | Logistic regression | Canadian | 50–74 years old | ≥1 mm | 1871 | 102 | 11.33 | Multicenter |
| Barbara Nemesure      | 2019 | Cohort study | Cox regression | American |                       | 1469 | 85 | 6.54 | Single-center |
| Michael W. Marcus     | 2019 | Screen trial | Logistic regression | English | 50–76 years old | ≥3 mm | 1013 | 52 | 2.60 | Multicenter |
| Martin Tammemagi      | 2018 | Screen trial | Logistic regression | Canadian | 50–74 years old | ≥1 mm | 1871 | 111 | 10.10 | Multicenter |
| Vineet K. Raghu       | 2019 | Cohort study | Logistic regression | American | Smoker | 92 | 50 | 10.00 | Multicenter |
| Joan E. Walter        | 2018 | Screen trial | Logistic regression | Dutch/Belgian | 50–75 years old and smoker | 809 | 50 | 7.14 | Multicenter |
| Xianfeng Li           | 2017 | Cohort study | Fisher discriminant analysis | Chinese | 20–80 years old | 5–30 mm | 39 | 20 | 1.00 | Single-center |
| Michal Reid           | 2019 | Cohort study | Logistic regression | American | ≥18 years old | ≤30 mm | 301 | 200 | 10.00 | Single-center |
| Michael K. Gould      | 2007 | Cohort study | Logistic regression | American | 7–30 mm | 375 | 204 | 13.60 | Multicenter |
| Sungmin Zo            | 2020 | Cohort study | Logistic regression | Korean | | 157 | 90 | 5.29 | Single-center |
| Xiao-Bo Chen          | 2019 | Cohort study | Logistic regression | Chinese | 8–20 mm | 493 | 214 | 11.26 | Single-center |
| Stephen J. Swensen    | 1997 | Cohort study | Logistic regression | American | 4–30 mm | 419 | 145 | 8.06 | Single-center |
| Man Zhang             | 2015 | Cohort study | Logistic regression | Chinese | ≤30 mm | 314 | 248 | 14.59 | Multicenter |
| Bin Zheng             | 2015 | Cohort study | Logistic regression | Chinese | ≤30 mm and GCO^b <50% | 405 | 367 | 11.84 | Single-center |
| Bin Zheng             | 2015 | Cohort study | Logistic regression | Chinese | ≤30 mm and GCO^b ≥50% | 159 | 166 | 5.35 | Single-center |
| Jingsi Dong           | 2014 | Cohort study | Logistic regression | Chinese | | 1679 | 1296 | 58.91 | Single-center |
| Yun Li                | 2012 | Cohort study | Logistic regression | Chinese | | 371 | 229 | 15.27 | Unspecified |
| Li Yang               | 2017 | Cohort study | Logistic regression | Chinese | | 1078 | 721 | 65.55 | Single-center |

^aApproximate number.

^bEPV, events per variable; GCO, ground glass opacity.
| First author                  | Year | Study design   | Targeted population          | Inclusion criteria of participants | Inclusion criteria of nodules | Sample size | Cases of lung cancer | Data source |
|------------------------------|------|----------------|------------------------------|-------------------------------------|-------------------------------|-------------|----------------------|-------------|
| Yoganand Balagurunathan      | 2019 | Screening trial | American                    | 55–74 years old and smoker          | ≥4 mm                         | 244         | 78                   | Multicenter |
| Gerard A. Silvestri         | 2018 | Cohort study   | American and Canadian        | >40 years old                       | 8–30 mm                       | 178         | 29                   | Multicenter |
| Chao Zhang                  | 2019 | Cohort study   | American and Chinese         | Unspecified                         | Unspecified                   | Multicenter |
| Johanna Uthoff              | 2019 | Cohort study   | American                     | Boxer                               | 363                           | 74          | Multicenter |
| Ilaria Bonavita             | 2020 | Cohort study   | American                     | Unspecified                         | Multicenter                   |
| Parnian Ashfar              | 2020 | Cohort study   | American                     | Unspecified                         | Multicenter                   |
| Huafeng Wang                | 2018 | Cohort study   | American                     | Unspecified                         | Multicenter                   |
| Jason L. Causey             | 2018 | Cohort study   | American                     | Unspecified                         | Multicenter                   |
| Samuel Hawkins              | 2016 | Screening trial | American                    | 55–74 years old and smoker          | ≥4 mm                         | 600         | 200                  | Multicenter |
| Samuel Hawkins              | 2016 | Screening trial | American                    | 55–74 years old and smoker          | ≥4 mm                         | 600         | 200                  | Multicenter |
| Andrew V. Kossenkov         | 2019 | Cohort study   | American                     | smoker                              | 583                           | 293         | Multicenter |
| G. A. Soardi                | 2015 | Cohort study   | American                     | smoker                              | 311                           | 199         | Multicenter |
| Zuohong Wu                  | 2021 | Cohort study   | Chinese                      | ≤30 mm                              | 995                           | 772         | Single-center       |
| Stéphane Chauvie            | 2020 | Screening trial | Chinese                      | 45–75 years old and smoker          | 234                           | 32          | Multicenter |
| Shulong Li                  | 2019 | Cohort study   | American                     | Unspecified                         | Multicenter                   |
| Relka Mastoun               | 2021 | Cohort study   | American                     | Unspecified                         | Multicenter                   |
| Yin-Chen Hsu                | 2020 | Cohort study   | Chinese                      | Unspecified                         | Multicenter                   |
| Jiaobao Liu                 | 2020 | Cohort study   | Chinese                      | 6–30 mm                             | 879                           | 601         | Multicenter |
| Rahul Paul                  | 2020 | Cohort study   | American                     | 55–74 years old and smoker          | ≥4 mm                         | 261         | 85                   | Multicenter |
| Muhammed Bilal Zia          | 2020 | Cohort study   | American                     | 1010                                | 926                           | Multicenter |
| Yi-Ming Xu                  | 2020 | Cohort study   | American                     | 1109                                | Multicenter                   |
| Subba R. Digumarthy         | 2019 | Cohort study   | American                     | 36                                  | Unspecified                   | Single-center |
| Yangwei Xiang               | 2019 | Cohort study   | Chinese                      | 588                                 | 462                           | Single-center |
| Litong Mao                  | 2019 | Cohort study   | Chinese                      | 294                                 | 61                            | Single-center |
| Shaun Daly                  | 2013 | Cohort study   | American                     | 136                                 | 69                            | Single-center |
### TABLE 3  Validation of traditional models

| First author                  | Year | Type of validation | Calibration | Sample size | AUC* | Thresholds | Sensitivity | Specificity |
|-------------------------------|------|--------------------|-------------|-------------|------|------------|-------------|-------------|
| Annette McWilliams           | 2013 | External           | Excellent   | 1090        | 0.970| 0.05       | 0.71        | 0.96        |
| Barbara Nemesure              | 2019 | Internal           | Not calibrated | 1455      | 0.860|            | 0.73        | 0.81        |
| Michael W. Marcus            | 2019 | Internal           | Excellent   | 1013        | 0.882|            |             |             |
| Martin T. ammemagi           | 2018 | External           | Excellent   | 3680        | 0.947|            |             |             |
| Vineet K. Raghu              | 2019 | External           | Not calibrated | 126       | 0.882| 0.61       | 0.28        | 1.00        |
| Joan E Walter                | 2018 | Internal           | Excellent   | 809         | 0.850|            |             |             |
| Xianfeng Li                  | 2017 | Internal           | Not calibrated | 39        | 0.921|            |             |             |
| Michal Reid                  | 2019 | External           | Excellent   | 45          | 0.810|            |             |             |
| Michael K. Gould             | 2007 | Internal           | Excellent   | 375         | 0.790|            |             |             |
| Sungmin Zö                  | 2020 | Internal           | Excellent   | 157         | 0.952|            |             |             |
| Xiao-Bo Chen                 | 2019 | External           | Excellent   | 216         | 0.848|            |             |             |
| Stephen J. Swensen           | 1997 | Internal           | Excellent   | 210         | 0.833| 0.10       | 0.93        | 0.47        |
|                               |      |                    |             |             |      | 0.40       | 0.51        | 0.90        |
| Man Zhang                    | 2015 | Internal           | Not calibrated | 120      | 0.910| 0.55       | 0.87        | 0.85        |
| Bin Zheng                    | 2015 | Internal           | Not calibrated | 198     | 0.808|            |             |             |
| Bin Zheng                    | 2015 | Internal           | Not calibrated | 84      | 0.845|            |             |             |
| Jingui Dong                  | 2014 | Internal           | Not calibrated | 1679    | 0.935|            |             |             |
| Yun Li                       | 2012 | External           | Not calibrated | 145     | 0.874| 0.46       | 0.95        | 0.70        |
| Li Yang                      | 2017 | Internal           | Not calibrated | 344     | 0.784| 0.70       | 0.70        | 0.79        |

*AUC, area under curve.

### TABLE 4  Validation of models based on the deep learning algorithm

| First author                  | Year | Sample size | Type of validation | AUC* | Threshold | Sensitivity | Specificity |
|-------------------------------|------|-------------|--------------------|------|-----------|-------------|-------------|
| Yogan and Balagurunathan      | 2019 | 235         | Internal           | 0.850| 0.54      | 0.69        | 0.84        |
| Gerard A. Silvestri           | 2018 | 178         | Internal           | 0.760| 0.05      | 0.97        | 0.44        |
| Chao Zhang                    | 2019 | Unspecified | External           | 0.855| 0.84      | 0.83        |             |
| Johanna Uthoff               | 2019 | 100         | External           | 0.965| 0.38      | 1.00        | 0.96        |
| Ilaria Bonavita               | 2020 | Unspecified | Internal           | Unspecified | 0.964| 0.95       | 0.90        |
| Parnian Afshar               | 2020 | 1010        | Internal           | 0.825| 0.69      | 0.84        |             |
| Huafeng Wang                 | 2018 | 1018        | Internal           | 0.970|           |             |             |
| Jason L. Causey              | 2018 | 1018        | Internal           | 0.993|           |             |             |
| Samuel Hawkins               | 2016 | 600         | Internal           | 0.83  |           |             |             |
| Samuel Hawkins               | 2016 | 600         | Internal           | 0.79  |           |             |             |
| Andrew V. Kossenkov          | 2019 | 158         | External           | 0.825| 0.69      | 0.84        |             |
| G. A. Soardi                 | 2015 | 311         | Internal           | 0.893|           |             |             |
| Zuohong Wu                   | 2021 | 995         | Internal           | 0.851| 0.88      | 0.64        |             |
| Stéphane Chauvie             | 2020 | 234         | Internal           | Unspecified | 0.931| 0.90      | 1.00        |
| Shulong Li                   | 2019 | 1010        | Internal           | 0.931| 0.83      | 0.92        |             |
| Rekka Mastouri               | 2021 | Unspecified | Internal           | 0.92  | 0.92      | 0.92        |             |
| Yin-Chen Hsu                 | 2020 | 836         | Internal           | 0.873| 0.75      | 0.85        |             |
| Jiabao Liu                   | 2020 | 879         | Internal           | 0.938| 0.58      | 0.84        | 0.91        |
| Rahul Paul                   | 2020 | 261         | Internal           | 0.960|           |             |             |
| Muhammed Bilal Zia           | 2020 | 1010        | Internal           | Unspecified | 0.91  | 0.91      | 0.91        |
| Yi-Ming Xu                   | 2020 | 1109        | Internal           | Unspecified | 0.93  | 0.93      | 0.89        |
| Subba R. Digumarthy          | 2019 | 36          | Internal           | 0.708|           |             |             |
| Yangwei Xiang                | 2019 | 588         | Internal           | 0.890| 0.90      | 0.80        |             |
| Liting Mao                   | 2019 | 294         | Internal           | 0.970| 0.81      | 0.81        | 0.92        |
| Shaun Daly                   | 2013 | 81          | External           | 0.676| 0.95      | 0.25        |             |

*AUC, area under curve.*
| Variables | Annette McWilliams | Barbara Nemesure | Michael W. Marcus | Martin Tammmemagi | Vineet K. Raghu | Joan E. Walter | Xianfeng Li | Michal K. Gould | Sungmin Zo | Stephen J. Swensen | Man Zhang | Bin Zheng | Bin Zheng | Jingqiu Dong | Yun Li | Li Yang |
|-----------|--------------------|-----------------|-------------------|-------------------|----------------|----------------|-------------|----------------|-----------|------------------|----------|-----------|-----------|-------------|--------|---------|
| Basic character | | | | | | | | | | | | | | | | |
| Age | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| Sex | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |
| Personal history of other cancer | 1 | 1 | | | | | | | | | | | | | | |
| Family history of lung cancer | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | | | |
| Family history of other cancer | 0 | 0 | | | | | | | | | | | | | | |
| BMI | 0 | 0 | | | | | | | | | | | | | | |
| Exposure of asbestos | 0 | 1 | | | | | | | | | | | | | | |
| FVC | 1 | | | | | | | | | | | | | | | |
| History of respiratory diseases | 1 | 1 | | | | | | | | | | | | | | |
| Smoke | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 |
| Clinical symptoms | | | | | | | | | | | | | | | | |
| Time since previous lung cancer was diagnosed | | | | | | | | | | | | | | | | |
| FEV1 | 0 | 1 | 1 | | | | | | | | | | | | | |
| Biomarkers | | | | | | | | | | | | | | | | |
| Squamous cell carcinoma antigen | | | | | | | | | | | | | | | | |
| NSE | | | | | | | | | | | | | | | | |
| CEA | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | | | | | | | |
| CYFRA21-1 | 1 | 0 | | | | | | | | | | | | | | |
| MiRNA-21-5p | 1 | 0 | | | | | | | | | | | | | | |
| MiR-574-5p | 1 | 0 | | | | | | | | | | | | | | |
| Laboratory indicators | | | | | | | | | | | | | | | | |
| Ferritin | | | | | | | | | | | | | | | | |
| Imaging information | | | | | | | | | | | | | | | | |
| Size | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| Volume | 1 | 1 | | | | | | | | | | | | | | |
| Density | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | | | | | | | |
| Location | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Count | 0 | 0 | 0 | 1 | 0 | | | | | | | | | | | |
| Margin (spiculate) | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | | |
| Satellite lesions | 1 | 1 | | | | | | | | | | | | | | |
| Calcification | | | | | | | | | | | | | | | | |

(Continues)
| Variables\(^a\) | Annette McWilliams\(^38\) | Barbara Nemesure\(^39\) | Michael W. Marcus\(^40\) | Martin Tammemagi\(^31\) | Vinnet K. Raghu\(^42\) | Joan E. Walter\(^43\) | Xianfeng Li\(^44\) | Michal Reid\(^45\) | Michael K. Gould\(^46\) | Sungmin Zo\(^47\) | Stephen J. Swensen\(^49\) | Man Zhang\(^50\) | Bin Zheng\(^1^{31}\) | Bin Zhong\(^2^{31}\) | Jingui Dong\(^32\) | Yun Li\(^53\) | Li Yang\(^34\) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cavitation | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Shape | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Enhancement | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pleural indentation | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Bronchus sign | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Vascular signs | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Emphysema | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Vessels sign | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Vessel number | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Tracheal signs | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Previous CT scan | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Previous X-ray | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Vacuole signs | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Associated pleural effusion | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Enlarged hilar or mediastinal lymph nodes | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Visibility in retrospect | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Carbohydrate antigen | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Neuron-specific enolase | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

\(^{a}\)0 depicts the inclusion of a variable into the model as a candidate variable; 1 depicts retention in the final model.

\(^{b}\)BMI, body mass index; FVC, forced vital capacity; FEV1, forced expiratory volume in one second; NSE, neuron-specific enolase; CEA, carcinoembryonic antigen; CEFRA21-1, cytokeratin fragment antigen 21-1; MiR(NA), MicroRNA.
medium to very low credibility, largely because of publication bias, indirectly, and imprecision (Table S2).

**Development and performance of traditional models**

The model from the Mayo clinic in the United States published in 1997 was the first model used to predict the risk of developing cancer from pulmonary nodules. Since then, 18 traditional models have been developed to predict the pathological characteristics of pulmonary nodules. Seven of these models were based on the North American population; two models were based on the European population, and nine models were based on the Asian population. Of the nine Asian models evaluated in this review, eight models were based on the Chinese population (Table 1).

Traditional models included numerous imaging features such as nodule size, type, location, shape, and margin to determine the pathological characteristics of the pulmonary nodules. In addition, basic information such as age, gender, family history of cancer, and smoking status was also commonly used. However, biomarkers were used in only seven models (Figure 2(e)).

Logistic regression analysis was used to develop most (16/18) traditional models. The models in the other two studies were developed using either Cox regression analysis or Fisher linear discriminant analysis. Most models (14/18) were cohort studies, and the remaining four were constructed using screening test results (Table 1). Based on the regression analysis, the size, margin of the nodules, smoking status, and age of patients were statistically significant in more than half of all models. The addition of biomarkers to tumor markers improved the AUC and statistical significance in three of the seven evaluated models, as shown in Table 5. These findings suggest that although biomarkers were not widely used to develop traditional models, they may have an important role in improving the accuracy of these models.

The AUCs of the models ranged from 0.676 to 0.970. Most models (77.8%, 14/18) performed well on discrimination, with an AUC higher or equal to 0.8. Calibration was assessed in nine models, and the results indicated a good fit. Most studies (61.1%, 11/18) had an EPV higher than

| First author                  | Objects for comparison | Indicators for comparison                      | Superior methods        |
|------------------------------|------------------------|------------------------------------------------|-------------------------|
| Yogan and Balagurunathan14    | None                   |                                                | Deep learning           |
| Gerard A. Silvestri15         | Traditional models     | AUC                                           | Deep learning           |
| Gerard A. Silvestri15         | Clinician              | AUC                                           | Deep learning           |
| Chao Zhang16                  | Clinician              | Accuracy, sensitivity, and specificity         | Deep learning           |
| Johanna Uthoff17              | None                   |                                                | Deep learning           |
| Ilaria Bonavita18             | Clinician              | F1 score                                      | Deep learning           |
| Parnian Afshar19              | None                   |                                                | Deep learning           |
| Huafeng Wang20                | None                   |                                                | Deep learning           |
| Jason L. Causey21             | Clinician              | AUC                                           | Similar                 |
| Samuel Hawkins 1.279          | Lung-RADS              | AUC                                           | Deep learning           |
| Samuel Hawkins 1.279          | Traditional models     | AUC                                           | Similar                 |
| Andrew V. Kossenkov23         | Traditional models     | AUC                                           | Deep learning           |
| G. A. Soardi24                | None                   |                                                | Similar                 |
| Zuohong Wu25                  | Traditional models     | AUC                                           | Deep learning           |
| Stéphane Chauvie26            | Lung-RADS              | PPV, sensitivity, and specificity             | Deep learning           |
| Stéphane Chauvie26            | Traditional models     | PPV, sensitivity, and specificity             | Deep learning           |
| Shulong Li27                  | None                   |                                                | Deep learning           |
| Rekka Mastouri28              | None                   |                                                | Deep learning           |
| Yin-Chen Hsu29                | Lung-RADS              | AUC                                           | Deep learning           |
| Jiabao Liu30                  | Clinician              | AUC                                           | Deep learning           |
| Rahul Paul31                  | None                   |                                                | Deep learning           |
| Muhammad Bilal Zia32          | None                   |                                                | Deep learning           |
| Yi-Ming Xu33                  | Clinician              | Sensitivity                                  | Deep learning           |
| Subba R. Digumarthy34         | None                   |                                                | Deep learning           |
| Yangwei Xiang35               | Traditional models     | AUC                                           | Deep learning           |
| Liting Mao36                  | ACR-Lung RADS8         | Accuracy, sensitivity, and specificity        | Deep learning           |
| Shaun Daly37                  | Traditional models     | AUC                                           | Deep learning           |

* AUC, area under curve; ACR-Lung-RADS, American College of Radiology Lung Imaging Reporting and Data System; PPV, positive predictive value.
of these models was validated using an external dataset. However, five of these models were validated using external data from a similar population from the same countries, and only one model\textsuperscript{38} was verified using data of participants from different origins. The latter model achieved good discrimination with an AUC of 0.970 (Tables 1 and 3).

Compared with the European and American models, the Chinese models lack external validation. Most of the data used to develop the Chinese models were obtained from a single-center or small sample retrospective cohort studies and only two of these studies were validated using an external dataset. However, the discrimination ability of the Chinese models was good, with seven of eight models achieving an AUC higher than 0.8, whereas two models reported excellent calibration. In addition, all Chinese models had an EPV higher than 10. More details can be found in Tables 1, 3, and Figures 2 and 3.

**Development and performance of the deep learning algorithms**

The first study reporting on the development and performance of a deep learning algorithm for the discrimination of pulmonary nodules was published in 2013.\textsuperscript{37} Only biomarkers were included in the development of this model, and the prediction ability was limited, with an AUC of 0.676. The majority of the deep learning models (84%, 21/25) were developed after 2018 and were based on the imaging features of the nodules. This improved the models’ prediction ability, especially when the model was supplemented by epidemiological parameters and biomarkers (Figure 3).

The AUC of the deep learning models was reported in 21 of 25. However, only half of these models (12 of 21) reported the confidence intervals (Table 4). The reported AUCs ranged from 0.676 to 0.970. Most of the deep learning models (68.0%, 17/25) had a good discrimination ability with an AUC higher than 0.8, whereas the other four models (16.0%) had an AUC below 0.8. The majority of the models (84.0%, 21/25 were not validated externally [Table 2]).

Only seven of 18 deep learning models were developed in Asia. Furthermore, all Asian models achieved high discrimination with an AUC above 0.8. However, the sample size of the Asian models was generally small, and only one of these models was validated using an external dataset (Tables 2 and 4).

**Comparison of deep learning models with traditional models**

The discrimination ability of 60.0% (15/25) of the deep learning models was compared with traditional methods. All deep learning models achieved higher or similar discrimination abilities when compared with traditional methods (Table 5).

**DISCUSSION**

LDCT can be used to diagnose lung cancer at an early stage via the identification and classification of pulmonary nodules into different risk categories. However, current pulmonary nodules classification guidelines are based solely on nodule size and density. Other important biomarkers and patient characteristics are mostly ignored, resulting in a very high false-positive rate, over diagnosis, and unnecessary treatment.\textsuperscript{55–57} Various traditional and deep learning models based on clinical, biological, and epidemiological factors have been developed to overcome this problem. To our knowledge, in this manuscript, we present the first systematic review comparing the development, validation, and performance of these models in the characterization of pulmonary nodules identified on LDCT.

In this systematic review, we evaluated the performance of 43 models derived from 41 research articles based on over 20 000 subjects. Our findings indicate that the majority of the traditional and deep learning models achieved an AUC higher than 0.8, suggesting that these models can be used to identify the high-risk population effectively and hence, reduce the false-positive rate and the harms of over diagnosis and treatment.

Since 1997, the development of pulmonary nodule risk prediction models has increased rapidly. Most early models were developed using statistical methods such as regression analysis. Although imaging features such as nodule size, type, location, shape, and margin provide valuable information on the pathological characteristics of the nodules, our findings indicate that the incorporation of clinical characteristics such as age and smoking status can significantly improve the performance of these models. The first study confirming this finding was performed at the Mayo Clinic.\textsuperscript{48}

Since then, various traditional statistic-based models incorporating both imaging and patient characteristics have been developed. Subsequent models also incorporated clinical indicators such as forced vital capacity (FVC) and forced expiratory volume (FEV)\textsubscript{1}, and serum biomarkers such as CEA and NSE, to further improve the prediction efficacy on the models.\textsuperscript{39,40,50–52} Variables including age, size of the nodules, and margin of the nodules should be considered as a priory in machine-learning analyses, as they were consistently considered as predictors of lung cancer in traditional studies.

A limited number of studies incorporated other risk factors such as exposure of asbestos, satellite lesions, bronchus sign, and volume of nodules (Table 5). However, the main limitation of these risk factors is the limited sample size that limits the generalizability of the model. A large number of models were based on single-center and retrospective studies with small sample sizes or data obtained from old studies. Biomarkers were not commonly used in the development of the predictive risk factor model (Table 5, Figure 2(e)). Nodule volume might have been an effective predictor,\textsuperscript{40,42} but was generally not taken into consideration by current models. Because most studies were retrospective, it was not
possible to incorporate time-dependent variables such as variations in biomarkers and nodule size over time into the model. Therefore, time-dependent factors, such as the nodule volume growth rate, were also ignored by most studies.

Deep learning models can learn from various heterogeneous variables to generate homogeneous groups with similar features. These features can be mapped with similar survival models to obtain accurate predictions. Various studies\textsuperscript{15,20,23,29} also suggest that compared with the traditional pulmonary nodule prediction models or expert judgment by clinicians, the use of deep learning algorithms has obvious advantages on discrimination (Table 6). However, although pulmonary nodule risk models based on deep learning algorithms have been used as early as 1993,\textsuperscript{58} they have not been widely used to predict pulmonary nodules until recent years as they still have several limitations. One of the main limitations of deep learning algorithms is that they require large amounts of data, advanced imaging equipment, top-ranked statisticians, and research funds to develop. Despite the high discrimination ability of the deep learning algorithm models evaluated in our systemic review, the GRADE scores of these models were generally low because of their limited sample size, high level of bias, inaccuracy, and indirectness (Table S2). Furthermore, it is difficult to identify the specific variables used to develop the deep learning prediction model, potentially limiting the quality and authenticity of these models.

Few studies were based on the Asian population. The majority of the Asian studies were based on a single center, had a limited sample size, and lacked external validation, which limited the quality of evidence (Tables 3 and 4, Figure 2). It is important to note that the accepted European and United States models may not be suitable for the Asian and Chinese populations because of large population differences, as suggested by Uthoff et al.\textsuperscript{59} and Nair et al.\textsuperscript{60}

Our systemic review has several limitations that have to be acknowledged. First of all, variations between studies, including sample size, research design, data source, and imaging acquisition criteria, made it difficult to quantify, integrate, and extrapolate the results of the different studies. Some of the studies included in our analysis had high publication bias, particularly those that lacked external validity. Additionally, cultural and social risk factors were ignored by most models. Studies evaluating a single risk factor were also excluded from this analysis although these variables were highly predictive of lung cancer and represent the latest trend in the field.

Furthermore, most of the existing models were based on the entire population. Therefore, subgroup analysis based on important risk factors such as smoking status and tumor histology is recommended to improve the prediction performance of current models and adapt these tools according to the specific characteristics of the population being studied. However, this type of research requires large datasets, highlighting the need for further large-scale multicenter prospective studies. Future studies should also focus on developing deep learning based models based on decentralized and deparametric data.\textsuperscript{61} These methods process the raw data directly and therefore, reduce the heterogeneity while improving the models’ performance compared with traditional models.

CONCLUSION

The incidence of lung cancer is increasing, particularly in developing countries. The models evaluated in our study were all developed in Europe, Asia, and the United States. These models showed good discrimination for identifying high-risk pulmonary nodules, particularly when these models combined imaging features with clinical, behavioral characteristics, and other biomarkers. This highlights the need to develop models based on the unique characteristics of different populations, particularly those in developing countries, to reduce the global lung cancer burden. The use of deep learning algorithms increased significantly during the last few years and generally performed better than traditional models. However, more research is required to improve the quality of the deep learning models, particularly for the Asian population, because these models were often based on single-center studies and lacked external validation. Further research should also focus on improving the quality of current screening guidelines by incorporating clinical and epidemiological factors into the evaluation of pulmonary nodules.

ACKNOWLEDGMENTS

This study was funded by grants from the National Key Research and Development Program of China (2018YFC1315000), Non-profit Central Research Institute Fund of Chinese; National Natural Science Foundation of China (8187102812); Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2020PT330001, 2019PT320027, 2019PT320023, 2018RC320010, and 3332019005).

CONFLICT OF INTEREST

The author declares that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

ORCID

Fengwei Tan \(\text{https://orcid.org/0000-0002-8210-684X}\)

Ni Li \(\text{https://orcid.org/0000-0001-5530-7745}\)

Jie He \(\text{https://orcid.org/0000-0002-0285-5403}\)

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49.
2. Santucci C, Carioli G, Bertuccio P, Malvezzi M, Pastorino U, Boffetta P, et al. Progress in cancer mortality, incidence, and survival: a global overview. Eur J Cancer Prev. 2020;29:367–81.
3. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. J Thorac Oncol. 2016;11:39–51.

4. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365:395–409.

5. Henschke CI, Naidich DP, Yankelevitz DF, McGuinness G, McCauley DI, Smith JP, et al. Early lung cancer action project: initial findings on repeat screenings. Cancer. 2001;92:153–9.

6. Church TR, Black WC, Aberle DR, Berg CD, Clingan KL, Duan F, et al. Results of initial low-dose computed tomographic screening for lung cancer. N Engl J Med. 2013;368:1980–91.

7. Swensen SJ, Jett JR, Sloan JA, Midhun DE, Hartman TE, Sykes AM, et al. Screening for lung cancer with low-dose spiral computed tomography. Am J Respir Crit Care Med. 2002;165:508–13.

8. Baldwin DR, Callister ME. The British Thoracic Society guidelines on the investigation and management of pulmonary nodules. Thorax. 2015;70:794–8.

9. Oudkerk M, Devaraj A, Vliegenthart R, Henzler T, Prosch H, et al. Probability of cancer in lung nodules using sequential management of uncertain findings on repeat screenings. Cancer. 2001;92:153–9.

10. Chan HP, Samala RK, Hadijski LM, Zhou C. Deep learning in medical image analysis. Adv Exp Med Biol. 2020;1213:3–21.

11. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436.

12. Iorio A, Spencer FA, Falavigna M, Chayen E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating expert level of lung cancer detection and classification using a deep convolutional neural network. Oncologist. 2019;24:1159–67.

13. Hawkins S, Wang H, Liu Y, Garcia A, Stringfield O, Kremer H, et al. Predicting malignant nodules from screening CT scans. J Thorac Oncol. 2020;15:216–21.

14. Causey JL, Zhang J, Ma S, Jiang B, Qualls JA, Politte DG, et al. Highly accurate model for prediction of lung nodule malignancy with CT scans. Sci Rep. 2018;8:3286.

15. Hawkins S, Wang H, Liu Y, Garcia A, Stringfield O, Kremer H, et al. Predicting malignant nodules from screening CT scans. J Thorac Oncol. 2016;11:2120–8.

16. Kossenkov AV, Qureshi R, Dawany NB, Wickramasinghe J, Liu Q, Majumdar RS, et al. A gene expression classifier from whole blood distinguishes benign from malignant lung nodules detected by low-dose CT. Cancer Res. 2019;79:262–73.

17. Soardi GA, Perandini S, Motton M, Montemazzi S. Assessing probability of malignancy in solid solitary pulmonary nodules with a new Bayesian calculator: improving diagnostic accuracy by means of expanded and updated features. Eur Radiol. 2015;25:155–62.

18. Wu Z, Huang T, Zhang S, Cheng D, Li W, Chen B. A prediction model to evaluate the pretest risk of malignancy in solitary pulmonary nodules: evidence from a large Chinese southwestern population. J Cancer Res Clin Oncol. 2021;147:275–85.

19. Chauvie S, De Maggi A, Baralis I, Dalmasso F, Berchialla P, Priotto R, et al. Artificial intelligence and radiomics enhance the positive predictive value of digital chest tomosynthesis for lung cancer detection within SOS clinical trial. Eur Radiol. 2020;30:4134–40.

20. Li S, Xu P, Li B, Chen L, Zhou Z, Hao H, et al. Predicting lung nodule malignancies by combining deep convolutional neural network and handicraft features. Phys Med Biol. 2019;64:175012.

21. Mastouri R, Khilfa N, Neji H, Hantous-Zannad S. A bilinear convolutional neural network for lung nodules classification on CT images. Int J Comput Assist Radiol Surg. 2021;16:91–101.

22. Hsu YC, Tsai YH, Yen HH, Hsu LS, Tsai YH, Lin YC, et al. Artificial neural networks improve detection of small lung cancers using low-dose CT images. Eur J Radiol. 2020;120:103882.

23. Zia MB, Juan ZJ, Zhou XJ, Xiao N, Wang JW, Khan A. Classification of malignant and benign lung nodules and prediction of image label class using multi-deep model. Int J Adv Comp Sci Appl. 2020;11:35–41.

24. Xu YM, Zhang T, Xu H, Qi L, Zhang W, Zhang YD, et al. Deep learning in CT images: automated pulmonary nodule detection for subsequent management using convolutional neural network. Cancer Manag Res. 2020;12:2979–92.

25. Paul R, Schabath M, Gillies R, Hall L, Goldgof D. Convolutional neural network ensembles for accurate lung nodule malignancy prediction 2 years in the future. Comput Biol Med. 2020;122:103882.

26. Zia MB, Juan ZJ, Zhou XJ, Xiao N, Wang JW, Khan A. Classification of malignant and benign lung nodules and prediction of image label class using multi-deep model. Int J Adv Comp Sci Appl. 2020;11:35–41.

27. Paul R, Schabath M, Gillies R, Hall L, Goldgof D. Convolutional neural network ensembles for accurate lung nodule malignancy prediction 2 years in the future. Comput Biol Med. 2020;122:103882.

28. Zia MB, Juan ZJ, Zhou XJ, Xiao N, Wang JW, Khan A. Classification of malignant and benign lung nodules and prediction of image label class using multi-deep model. Int J Adv Comp Sci Appl. 2020;11:35–41.

29. Xu YM, Zhang T, Xu H, Qi L, Zhang W, Zhang YD, et al. Deep learning in CT images: automated pulmonary nodule detection for subsequent management using convolutional neural network. Cancer Manag Res. 2020;12:2979–92.

30. Paul R, Schabath M, Gillies R, Hall L, Goldgof D. Convolutional neural network ensembles for accurate lung nodule malignancy prediction 2 years in the future. Comput Biol Med. 2020;122:103882.

31. Paul R, Schabath M, Gillies R, Hall L, Goldgof D. Convolutional neural network ensembles for accurate lung nodule malignancy prediction 2 years in the future. Comput Biol Med. 2020;122:103882.

32. Zia MB, Juan ZJ, Zhou XJ, Xiao N, Wang JW, Khan A. Classification of malignant and benign lung nodules and prediction of image label class using multi-deep model. Int J Adv Comp Sci Appl. 2020;11:35–41.

33. Xu YM, Zhang T, Xu H, Qi L, Zhang W, Zhang YD, et al. Deep learning in CT images: automated pulmonary nodule detection for subsequent management using convolutional neural network. Cancer Manag Res. 2020;12:2979–92.

34. Paul R, Schabath M, Gillies R, Hall L, Goldgof D. Convolutional neural network ensembles for accurate lung nodule malignancy prediction 2 years in the future. Comput Biol Med. 2020;122:103882.

35. Paul R, Schabath M, Gillies R, Hall L, Goldgof D. Convolutional neural network ensembles for accurate lung nodule malignancy prediction 2 years in the future. Comput Biol Med. 2020;122:103882.

36. Zia MB, Juan ZJ, Zhou XJ, Xiao N, Wang JW, Khan A. Classification of malignant and benign lung nodules and prediction of image label class using multi-deep model. Int J Adv Comp Sci Appl. 2020;11:35–41.

37. Paul R, Schabath M, Gillies R, Hall L, Goldgof D. Convolutional neural network ensembles for accurate lung nodule malignancy prediction 2 years in the future. Comput Biol Med. 2020;122:103882.

38. Zia MB, Juan ZJ, Zhou XJ, Xiao N, Wang JW, Khan A. Classification of malignant and benign lung nodules and prediction of image label class using multi-deep model. Int J Adv Comp Sci Appl. 2020;11:35–41.

39. Paul R, Schabath M, Gillies R, Hall L, Goldgof D. Convolutional neural network ensembles for accurate lung nodule malignancy prediction 2 years in the future. Comput Biol Med. 2020;122:103882.

40. Paul R, Schabath M, Gillies R, Hall L, Goldgof D. Convolutional neural network ensembles for accurate lung nodule malignancy prediction 2 years in the future. Comput Biol Med. 2020;122:103882.

41. Paul R, Schabath M, Gillies R, Hall L, Goldgof D. Convolutional neural network ensembles for accurate lung nodule malignancy prediction 2 years in the future. Comput Biol Med. 2020;122:103882.

42. Paul R, Schabath M, Gillies R, Hall L, Goldgof D. Convolutional neural network ensembles for accurate lung nodule malignancy prediction 2 years in the future. Comput Biol Med. 2020;122:103882.

43. Paul R, Schabath M, Gillies R, Hall L, Goldgof D. Convolutional neural network ensembles for accurate lung nodule malignancy prediction 2 years in the future. Comput Biol Med. 2020;122:103882.

44. Paul R, Schabath M, Gillies R, Hall L, Goldgof D. Convolutional neural network ensembles for accurate lung nodule malignancy prediction 2 years in the future. Comput Biol Med. 2020;122:103882.
43. Walter JE, Heuvelmans MA, Bock GH, Yousaf-Khan U, Groen HJM, Aalst CMV, et al. Characteristics of new solid nodules detected in incidence screening rounds of low-dose CT lung cancer screening: the NELSON study. Thorax. 2018;73:741–7.
44. Li X, Zhang Q, Jin X, Cao L. Combining serum miRNAs, CEA, and CYFRA21-1 with imaging and clinical features to distinguish benign and malignant pulmonary nodules: a pilot study: Xianfeng Li et al.: combining biomarker, imaging, and clinical features to distinguish pulmonary nodules. World J Surg Oncol. 2017;15:107.
45. Reid M, Choi HK, Han X, Wang X, Mukhopadhyay S, Kou L, et al. Development of a risk prediction model to estimate the probability of malignancy in pulmonary nodules being considered for biopsy. Chest. 2019;156:367–75.
46. Gould MK, Ananth L, Barnett PG. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. Chest. 2007;131:383–8.
47. Zo S, Woo SY, Kim S, Lee JE, Jeong BH, Um SW, et al. Predicting the risk of malignancy of lung nodules diagnosed as indeterminate on radial endobronchial ultrasound-guided biopsy. J Clin Med. 2020;9:3652.
48. Chen XB, Yan RY, Zhao K, Zhang F, Li YJ, Wu L, et al. Nomogram for the prediction of malignancy in small (8-20 mm) indeterminate solid solitary pulmonary nodules in Chinese populations. Cancer Manag Res. 2019;11:9439–48.
49. Donnelly EF, Kazerooni EA, Lee E, Henry TS, Boiselle PM, Crabtree TD, et al. ACR appropriateness criteria lung cancer screening. J Am Coll Radiol. 2018;15:S341–6.
50. Mazzone PJ, Silverstri GA, Souter LH, Caverly TJ, Kanne JP, Katki HA, et al. Screening for lung cancer: CHEST guideline and expert panel report. Chest. 2021;160:e427–94.
51. Dong JS, Sun N, Li JG, Liu ZY, Zhang BH, Chen ZL, et al. Development and validation of clinical diagnostic models for the probability of malignancy in solitary pulmonary nodules. Thorac Cancer. 2014;5:162–8.
52. Li Y, Wang J. A mathematical model for predicting malignancy of solitary pulmonary nodules. World J Surg. 2012;36:830–5.
53. Yang L, Zhang Q, Bai L, Li TY, He C, Ma QL, et al. Assessment of the cancer risk factors of solitary pulmonary nodules. Oncotarget. 2017;8:29318–27.
54. Wood DE, Kazerooni EA, Baum SL, Eapen GA, Ettinger DS, Hou L, et al. Lung cancer screening, version 3.2018, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2018;16:412–41.
55. Wood DE, Kazerooni EA, Lee E, Henry TS, Boiselle PM, Crabtree TD, et al. ACR appropriateness criteria® lung cancer screening. J Am Coll Radiol. 2018;15:5341–6.