CADTH

CADTH RAPID RESPONSE REPORT: SUMMARY OF ABSTRACTS

Artificial Intelligence for Classification of Lung Nodules: Clinical Utility, Diagnostic Accuracy, Cost-Effectiveness, and Guidelines

Service Line:	Rapid Response Service
Version:	1.0
Publication Date:	November 14, 2019
Report Length:	12 Pages

CADTH

Authors: Shannon Hill, Melissa Walter

Cite As: *Artificial Intelligence for Classification of Lung Nodules: Clinical Utility, Diagnostic Accuracy, Cost-Effectiveness, and Guidelines*. Ottawa: CADTH; 2019 Nov. (CADTH rapid response report: summary of abstracts).

Funding: CADTH receives funding from Canada's federal, provincial, and territorial governments, with the exception of Quebec.

Questions or requests for information about this report can be directed to **requests@cadth.ca**

Research Questions

1. What is the clinical utility of artificial intelligence for nodule classification in screening, incidental identification, or known or suspected malignancies for lung cancer?
2. What is the diagnostic accuracy of artificial intelligence for nodule classification in screening, incidental identification, or known or suspected malignancies for lung cancer?
3. What is the cost-effectiveness of artificial intelligence for nodule classification in screening, incidental identification, or known or suspected malignancies for lung cancer?
4. What are the evidence-based guidelines regarding artificial intelligence for nodule classification in screening, incidental identification, or known or suspected malignancies for lung cancer?

Key Findings

Two randomized controlled trials and twelve non-randomized studies were identified regarding the clinical utility and diagnostic accuracy of artificial intelligence for nodule classification in screening, incidental identification, or known or suspected malignancies for lung cancer.

Methods

A limited literature search was conducted by an information specialist on key resources including MEDLINE, the Cochrane Library, the University of York Centre for Reviews and Dissemination (CRD) databases, the websites of Canadian and major international health technology agencies, as well as a focused Internet search. The search strategy was comprised of both controlled vocabulary, such as the National Library of Medicine's MeSH (Medical Subject Headings), and keywords. The main search concepts were artificial intelligence and lung nodules. No filters were applied to limit the retrieval by study type. Where possible, retrieval was limited to the human population. The search was also limited to English language documents published between January 1, 2014 and October 31, 2019. Internet links were provided, where available.

Selection Criteria

One reviewer screened citations and selected studies based on the inclusion criteria presented in Table 1.

Table 1: Selection Criteria

Population	Patients with lung nodules (< 3 cm) suspected of having lung cancer - Patients identified during a screening program - Patients identified incidentally when having a scan for an unrelated reason - Patients with a known or suspected malignancy
Intervention	Artificial intelligence algorithms for reading computed tomography (or computerized axial tomography) scans to classify nodules, excluding detection alone (e.g., machine learning [supervised and unsupervised learning, support vector machines, random forests, black box learning], deep learning, artificial neural network [convolutional neural networks])
Comparator	Radiologist-read computed tomography (or computerized axial tomography) scan; clinical decision support tools or criteria (e.g., Vancouver Lung Cancer Risk Prediction Model)
Outcomes	Q1: Clinical utility (e.g., avoidance of biopsy, psychological distress, early or appropriate treatment, detection of cancer, survival, additional testing) Q2: Diagnostic accuracy (i.e., clinical validity: sensitivity, specificity, negative predictive value, positive predictive value, true positives, precision, accuracy, recall) Q3: Cost effectiveness (e.g., incremental cost effectiveness ratio, quality adjusted life years) Q4: Evidence-based guidelines
Study Designs	Health technology assessments, systematic reviews, meta-analyses, randomized control trials, non-randomized studies, economic evaluations, evidence-based guidelines.

Results

Rapid Response reports are organized so that the higher quality evidence is presented first. Therefore, health technology assessment reports, systematic reviews, and meta-analyses are presented first. These are followed by randomized controlled trials, non-randomized studies, economic evaluations, and evidence-based guidelines.

Two randomized controlled trials[1,2] and twelve non-randomized studies[3-14] were identified regarding the clinical utility and diagnostic accuracy of artificial intelligence for nodule classification in screening, incidental identification, or known or suspected malignancies for lung cancer.

Overall Summary of Findings

Two randomized controlled trials[1,2] and twelve non-randomized studies[3-14] were identified regarding the clinical utility and diagnostic accuracy of artificial intelligence (AI) for nodule classification in screening, incidental identification, or known or suspected malignancies for lung cancer.

The authors from the first randomized controlled trial[1] found that radiologist's diagnostic predictions improved with the use of computer aided diagnostic prediction disclosure. The authors of the second randomized controlled trial[2] found equivalent performances for benign and malignant nodule classification between the computer model and human observers, but human observers performed better when nodules were matched in size.

The authors of two non-randomized studies[3,4] that focused on the clinical utility of AI for nodule classification concluded that the use of AI can offer classification accuracy improvement over human observation alone. The authors of the non-randomized studies[3,4] stated that classification accuracy with the use of AI can reduce the time required for results interpretation,[3] improve results interpretation,[3,4] and offer aid in treatment for patients.[3]

The authors of nine non-randomized studies[5-12,14] that focused on the diagnostic accuracy of AI for nodule classification found that AI generally achieved a higher accuracy measurement for nodule classification compared to human observation or previously reported data (the ACR lung-RADS). The exception of one non-randomized study[13] reported that the computerized classification methods used in the study achieved a lower accuracy measurement compared to human observation.

Further details of the included studies can be found in Table 2.

Table 2: Summary of Included Studies

Author, Year	Purpose and Population	Artificial Intelligence Intervention	Comparator	Results
		Randomized Controlled Trials		
Perandini,[1] 2019	The purpose of this study was to assess the accuracy gain of a CAD versus radiologist judgement for characterizing solitary pulmonary nodules using CT. This study included randomly selected solitary pulmonary nodules with a definitive diagnosis. **N=100**	Bayesian analysis-based computer aided diagnosis (Bayesian Inference Malignancy Calculator).	Seven radiologists using a one to five risk chart to evaluate CT scans.	Radiologist predictions were evaluated using ROC curve before and after the disclosure of CAD predictions. The overall ROC area under curve was 0.758 before disclosure of CAD predictions, and 0.803 after disclosure of CAD predictions.
van Riel,[2] 2017	The purpose of this study was to compare human observers to a computer model for differentiating malignant and benign pulmonary nodules from baseline CT scans. The population group included chest CT scans from the Danish Lung Cancer Screening Trial with randomly selected benign nodule scans. **N=300**	A mathematically derived computer model called PanCan that assigned a malignancy probability score to each nodule.	11 human observers to assign a malignancy probability score to each nodule. Seven observers assessed morphological nodule characteristics.	Performance of assigned malignancy probability scores was assessed using AUC. For risk-assessment of nodules of all sizes, performances between computer model and human observers were equivalent (AUC 0.932 versus 0.910, p=0.184). For differentiating malignant nodules from size-matched benign nodules,

Author, Year	Purpose and Population	Artificial Intelligence Intervention	Comparator	Results
				human observers performed better (AUC 0.819 versus 0.706, p < 0.001).
Non-Randomized Studies (Clinical Utility Studies)				
Li,[3] 2019	The purpose of this study was to provide a reference for the application of AI for diagnosis and follow-up of multiple pulmonary nodules. The population included patients with synchronous and metachronous multiple pulmonary nodules. **N=53**	Use of AI for diagnosis (AI properties were not specified in the abstract).	Postoperative pathological tests for diagnosis.	Agreement between artificial intelligence diagnosis and postoperative pathological tests was 88.8% for identifying benign and malignant nodules. The authors concluded that AI as a diagnostic aid shows more accurate and objective results for diagnosing multiple pulmonary nodules which reduces time required for results interpretation and can aid in follow-up and treatment for patients.
Chen,[4] 2017	The purpose of this study was to determine if various texture features can differentiate between malignant and benign solid pulmonary nodules using a machine model. The study was based on patients from granuloma-endemic regions.	Machine learning models were trained on texture features on DTPI images for discriminating between benign and malignant solid pulmonary nodules.	Standard clinical metrics and visual interpretation of DTPI images.	The authors concluded that machine learning methods found optimal discrimination between malignant and benign nodules and achieved significant improvements over standard clinical metrics and visual interpretation methods.
Non-Randomized Studies (Diagnostic Accuracy Studies)				
Gong,[5] 2019	The purpose of this study was to develop a CT based radiomic feature analysis approach for the diagnosis of ground-	A LOOCV method was used to build models. The models were embedded with a Relied feature selection, SMOTE and	Two radiologists.	Performance of the machine learning classifiers was assessed using the average AUC. When compared to the two

Author, Year	Purpose and Population	Artificial Intelligence Intervention	Comparator	Results
	glass opacity pulmonary nodules, and to assess the performance of CADx in classifying benign and malignant nodules associated with histopathological subtypes. This study included histopathology-confirmed ground-glass opacity nodules collected from two cancer centers. **N=182**	three machine learning classifiers (Support vector machine, logistic regression, and Gaussian Naïve Bayes).		radiologists, the scheme yielded a higher accuracy (61.3% versus radiologist 1 and 53.1% versus radiologist 2). The authors concluded that using CT based radiomic feature analysis is a feasible approach to distinguish between benign and malignant nodules, and CADx had a higher performance in diagnosing nodules compared to radiologists.
Mao,[6] 2019	The purpose of this study was to assess the usefulness of a quantitative radiomic model for predicting malignancy in small solid pulmonary nodules. This study included malignant and benign small solid pulmonary nodules detected in baseline low-dose CT screening.	A radiomic predictive model.	The ACR lung-RADS.	The radiomic prediction model yielded a SE of 81.0% and SP of 92.2%. The accuracy of the radiomic prediction model was higher than the ACR-lung RADS (89.9% versus 76.5%). The authors concluded that a radiomic model can improve the accuracy in predicting malignancy of small solid pulmonary nodules.
Wu,[7] 2019	The purpose of this study was to propose a classification scheme to help radiologists differentiate benign and malignant pulmonary nodules on CT scans. The data set included nodules collected from the LIDC.	Malignant-benign classification using RF machine learning model with clustering analysis.	A variant composite rank of malignancy from four radiologists.	The classification scheme was verified by three experiments and resulted in an AUC of 0.9702, 0.9190, and 0.8662, respectively. The authors concluded that the method of this classification scheme

Author, Year	Purpose and Population	Artificial Intelligence Intervention	Comparator	Results
	N=952			can improve the accuracy of classifying benign and malignant nodules based on CT images.
Zhang,[8] 2019	The purpose of this study is to integrate a deep learning algorithm to detect and classify pulmonary nodules from CT images. Clinical CT data was obtained through open-source data sets and multi-center data sets.	A three-dimensional CNN.	Manual assessments done by different ranks of doctors.	The CNN found to have a SE and SP of 84.4% and 83.0%, respectively. The authors stated that the CNN showed to perform better that the manual assessment. The authors concluded that CNN with a deep learning algorithm may help radiologist by providing accurate information for diagnosing pulmonary nodules.
Causey,[9] 2018	This purpose of this study is to present NoduleX which is a systematic approach to predict lung nodule malignancy from CT data. Over 1000 lung nodules were analyzed from the LIDC/IDRI cohort.	A deep learning CNN (NoduleX).	Nodule identification and classification of four thoracic radiologists.	NoduleX showed to achieve a higher accuracy for nodule malignancy classification. This was measured using the AUC. The authors reported an AUC of ~0.99. The authors concluded that this AUC corresponds with the analysis of the data by radiologists.
Choi,[10] 2018	The purpose of this study was to develop a radiomics prediction model to improve pulmonary nodule malignancy classification. This study included pulmonary nodules	A prediction model was constructed by using an SVM-LASSO model.	The ACR lung-RADS.	The authors reported that the best SVM-LASSO model consisted of two features and achieved an accuracy of 84.6% and 0.89 AUC compared to the Lung-RADS which

Author, Year	Purpose and Population	Artificial Intelligence Intervention	Comparator	Results
	from the LIDC image collection. **N=72**			achieved an accuracy of 72.2% and 0.77 AUC. The authors concluded that the SVM-LASSO model achieved a higher accuracy than Lung-RADS.
Abbas,[11] 2017	The purpose of this study was to implement a CADe system, known as Nodular-Deep, to classify pulmonary lung nodules into benign and malignant classes. This system was tested on CT scans from the LIDC. **N=1200**	Nodular-Deep which used CNN and RNN algorithms combined with softmax linear classifier.	Segmentation done by a radiologist.	The performance of Nodular-Deep was validated using SE, SP, and AUC measurements. Nodular-Deep achieved 94% SE, 96% SP and 0.85 AUC in determining benign and malignant nodules. The authors concluded that Nodular-Deep outperforms manual segmentation by a radiologist.
Alilou,[12] 2017	The purpose of this study was to evaluate 3D shape features for discriminating benign from malignant nodules on lung CT images. This study included 82 and 67 patients from two different institutions.	An SVM classifier was combined with a feature selection scheme.	Two expert radiologists and one pulmonologist.	The authors validated the SVM classifier and comparators using the AUC. The SVM classifier yielded an AUC of 0.72 and 0.64 for manually and automatically segmented nodules, respectively. The comparators yielded an AUC of 0.82, 0.68, and 0.58 for the two radiologists and one pulmonologist, respectively.
Armato,[13] 2016	The purpose of this study was to describe and report the performance of the LUNGx Challenge for classifying benign and	Ten groups that applied their own computerized classification system methods.	Six radiologists.	Classification was validated using the AUC. The computerized classification methods achieved a

Author, Year	Purpose and Population	Artificial Intelligence Intervention	Comparator	Results
	malignant lung nodules based on CT scans. This study included nodules selected on approximate size matching from two cohorts. **N=73**			range from 0.50 to 0.68 AUC. The radiologists achieved a range from 0.70 to 0.85 AUC.
Dhara,[14] 2016	The purpose of this study was to classify benign and malignant pulmonary nodules using SVM. This study included a dataset of nodules from the LIDC/IDRI public database. **N=891**	An SVM using a semi-automated technique for classification.	Segmentation of nodules by trained radiologists.	The classification of the SVM was validated by the A z. The proposed classifying method achieved an A z of 0.9505, 0.8822 and 0.8488 based on three different configurations. The authors concluded that the proposed SVM method outperforms segmentation of pulmonary nodules by trained radiologists.

A z = area under the receiver operating characteristic curve; ACR lung-RADS = American College of Radiologists lung CT screening reporting and data system; AI = artificial intelligence; AUC = area under the receiver operating characteristic curve; CAD = computer aided diagnosis; CADe = computer aided diagnostic; CADx = computer aided diagnosis; CNN = convolutional neural network; CT = computer tomography; DTPI = dual time DFG/PET CT; LIDC = lung image database consortium; LIDC/IDRI = lung image database consortium and image database resource initiative; LOOCV = leave-one-out cross-validation; RF = random forest; RNN = recurrent neural network; ROC = receiver operating characteristics; SE = sensitivity; SMOTE = synthetic minority oversampling technique; SP = specificity; SVM = support vector machine; SVM-LASSO = support vector machine with a least absolute shrinkage and selection operator.

References Summarized

Health Technology Assessments

No literature identified.

Systematic Reviews and Meta-analyses

No literature identified.

Randomized Controlled Trials

1. Perandini S, Soardi GA, Motton M, et al. Enhanced characterization of solid solitary pulmonary nodules with Bayesian analysis-based computer-aided diagnosis. *World J Radiol*. 2016 Aug;8(8):729-734.
PubMed: PM27648166

2. van Riel SJ, Ciompi F, Winkler Wille MM, et al. Malignancy risk estimation of pulmonary nodules in screening CTs: Comparison between a computer model and human observers. *PLoS One*. 2017 Nov;12(11):e0185032.
[PubMed: PM29121063]

Non-Randomized Studies

Clinical Utility Studies

3. Li X, Hu B, Li H, You B. Application of artificial intelligence in the diagnosis of multiple primary lung cancer. *Thorac Cancer*. 2019 Nov;10(11):2168-2174.
[PubMed: PM31529684]

4. Chen S, Harmon S, Perk T, et al. Diagnostic classification of solitary pulmonary nodules using dual time ^{18}F-FDG PET/CT image texture features in granuloma-endemic regions. *Sci Rep*. 2017 Aug;7(1):9370.
[PubMed: PM28839156]

Diagnostic Accuracy Studies

5. Gong J, Liu J, Hao W, et al. Computer-aided diagnosis of ground-glass opacity pulmonary nodules using radiomic features analysis. *Phys Med Biol*. 2019 Jul;64(13):135015.
[PubMed: PM31167172]

6. Mao L, Chen H, Liang M, et al. Quantitative radiomic model for predicting malignancy of small solid pulmonary nodules detected by low-dose CT screening. *Quant Imaging Med Surg*. 2019 Feb;9(2):263-272.
[PubMed: PM30976550]

7. Wu W, Hu H, Gong J, et al. Malignant-benign classification of pulmonary nodules based on random forest aided by clustering analysis. *Phys Med Biol*. 2019 Jan;64(3):035017.
[PubMed: PM30702087]

8. Zhang C, Sun X, Dang K, et al. Toward an expert level of lung cancer detection and classification using a deep convolutional neural network. *Oncologist*. 2019 Sep;24(9):1159-1165.
[PubMed: PM30996009]

9. Causey JL, Zhang J, Ma S, et al. Highly accurate model for prediction of lung nodule malignancy with CT scans. *Sci Rep*. 2018 Jun;8(1):9286.
[PubMed: PM29915334]

10. Choi W, Oh JH, Riyahi S, et al. Radiomics analysis of pulmonary nodules in low-dose CT for early detection of lung cancer. *Med Phys*. 2018 Apr;45(4):1537-1549.
[PubMed: PM29457229]

11. Abbas Q. Nodular-Deep: classification of pulmonary nodules using deep neural network. *Int J Med Res Health Sci*. 2017;6(8):111-118.
https://www.ijmrhs.com/abstract/nodulardeep-classification-of-pulmonary-nodules-using-deep-neural-network-12959.html

12. Alilou M, Beig N, Orooji M, et al. An integrated segmentation and shape-based classification scheme for distinguishing adenocarcinomas from granulomas on lung CT. *Med Phys*. 2017 Jul;44(7):3556-3569.
PubMed: PM28295386

13. Armato SG 3rd, Drukker K, Li F, et al. LUNGx Challenge for computerized lung nodule classification. *J Med Imaging (Bellingham)*. 2016 Oct;3(4):044506.
PubMed: PM28018939

14. Dhara AK, Mukhopadhyay S, Dutta A, Garg M, Khandelwal N. A combination of shape and texture features for classification of pulmonary nodules in lung CT images. *J Digit Imaging*. 2016 Aug;29(4):466-475.
PubMed: PM26738871

Economic Evaluations

No literature identified.

Guidelines and Recommendations

No literature identified.