ARTICLE INFO

Article History:
Received 7 July 2020
Revised 14 November 2020
Accepted 17 November 2020
Available online 25 December 2020

Keywords:
Tumor metastasis
Medical imaging
Artificial intelligence
Deep learning
Diagnostic meta-analysis

ABSTRACT

Background: Early diagnosis of tumor metastasis is crucial for clinical treatment. Artificial intelligence (AI) has shown great promise in the field of medicine. We therefore aimed to evaluate the diagnostic accuracy of AI algorithms in detecting tumor metastasis using medical radiology imaging.

Methods: We searched PubMed and Web of Science for studies published from January 1, 1997, to January 30, 2020. Studies evaluating an AI model for the diagnosis of tumor metastasis from medical images were included. We excluded studies that used histopathology images or medical wave-form data and those focused on the region segmentation of interest. Studies providing enough information to construct contingency tables were included in a meta-analysis.

Findings: We identified 2620 studies, of which 69 were included. Among them, 34 studies were included in a meta-analysis with a pooled sensitivity of 82% (95% CI 79–84%), specificity of 84% (82–87%) and AUC of 0.90 (0.87–0.92). Analysis for different AI algorithms showed a pooled sensitivity of 87% (83–90%) for machine learning and 86% (82–89%) for deep learning, and a pooled specificity of 89% (82–93%) for machine learning, and 87% (82–91%) for deep learning.

Interpretation: AI algorithms may be used for the diagnosis of tumor metastasis using medical radiology imaging with equivalent or even better performance to health-care professionals, in terms of sensitivity and specificity. At the same time, rigorous reporting standards with external validation and comparison to health-care professionals are urgently needed for AI application in the medical field.

Funding: College students’ innovative entrepreneurial training plan program.

© 2020 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

Tumor metastasis, including lymph node metastasis (LNM) and distant metastasis (DM), contributes to cancer-related death. Regarding tumor classification, N and M staging are essential for both the treatment strategy, like the plan for surgery and chemoradiotherapy, and prognosis prediction [1,2]. Thus, it is crucial to conduct a complete and accurate pre-operative clinical evaluation of tumor metastasis.

Medical imaging is commonly used to visualize tumor dissemination and quantify the severity, providing valuable information for diagnosis, staging and treatment decision [3] with satisfactory diagnostic accuracy. For example, the sensitivity and specificity of contrast-enhanced ultrasound (CEUS), multidetector computed tomography (MDCT), magnetic resonance imaging (MRI), and fluorodeoxyglucose (FDG) positron emission tomography (PET)/CT in the detection of colorectal cancer liver metastasis was 80–97% [4], which is similar in other diseases [5,6]. However, owing to the uncoordinated ratio of doctors to patients and the difficulty of radiological diagnosis, making a correct and timely diagnosis from medical imaging is challenging [7].

Artificial intelligence (AI) has already shown great promise to address this problem through automated diagnosis from medical imaging [8,9]. In the 1980s, artificial neural networks (ANNs) were developed [10], resulting in a surge of machine learning (ML) based on statistical models. In the 1990s, various ML models were successively proposed, such as support vector machines (SVM) [11] and random forests (RF) [12]. It is not until 2006 that deep learning (DL), a new branch of ML, gained great attention [13,14]. Since then, DL, such as convolutional neural networks (CNN) and deep neural networks (DNN), has been applied in many fields, including photo
2. Methods

2.1. Search strategy and selection criteria

In this systematic review and meta-analysis, we searched for studies that developed or validated an AI model for the diagnosis of tumor metastasis (LNM and DM) from medical radiology imaging. We searched PubMed and Web of Science for studies published from January 1, 1997, to January 30, 2020, with no restrictions on regions, languages, or publication types. Studies were included if they evaluated an AI model for the diagnosis of tumor metastasis from medical images. We found one systematic review comparing performance of AI algorithms with health-care professionals for all diseases, but we did not find systematic reviews focusing on tumor metastasis.

2.2. Data collection

Three reviewers (QZ, LY and JL) extracted data independently using a predefined data extraction sheet, and uncertainties were resolved by another reviewer (BZ). We extracted binary diagnosis accuracy data and constructed contingency tables, which included true-positive (TP), false-positive (FP), true-negative (TN), and false-negative (FN) results if the study provided enough information. Sensitivity and specificity results were calculated from contingency tables.

We evaluated the performance of the AI model, we conducted a meta-analysis from studies providing enough information to construct contingency tables. If a study provided several contingency tables for different algorithms or primary tumors, we treated them as independent items.

The quality of the included studies was evaluated by the reviewers (QZ and KG) and conformed to the revised version of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) [29].

2.3. Statistical analysis

Receiver operating characteristic (ROC) curves were constructed to evaluate the accuracy of the AI model. The ROC figures provide average sensitivity and specificity across included studies with a 95% confidence interval (CI) of the summary operating point. The ROC figures also provide the 95% prediction region representing the confidence intervals for forecasts of sensitivity and specificity in a future study. Areas under the ROC curve (AUCs) with 95% CI were also calculated. Odds ratio (OR) and 95% CI for each study was calculated to estimate the performance of the AI algorithms.
We calculated heterogeneity between studies using the \( \chi^2 \) test (threshold \( P = 0.1 \)), which was quantified using the \( I^2 \) statistic. We also conducted the subgroup analysis and regression analysis to identify the sources of heterogeneity. Random effects models were used during the process. \( P \) value of 0.05 or less was considered to indicate a statistically significant difference.

Two separate analyses were performed according to different algorithms and whether studies were externally validated. Following its development, we divided AI algorithms into ML algorithms (ANN, KNN, SVM, RF, logistic regression and decision tree) and DL algorithms (CNN, DNN and DCNN). External validation means studies were validated by out-of-sample dataset.

To compare diagnostic performance between AI algorithms and health-care professionals, we did another separate analysis for studies providing contingency tables for both health-care professionals and AI algorithm using the same sample. We evaluated the quality of included studies according to QUADAS-2 by RevMan (Version 5.3). Stata (Version 15.0) was used in the ROC curves, the calculation of AUC, subgroup analysis, Deeks’ Funnel Plot Asymmetry Test and forest plots. Data analysis was performed by BZ. This study is registered with PROSPERO, CRD42020172924.

### 3. Results

Fig. 1 summarized our literature search for eligible studies. Our search identified 2620 records, of which 1991 were screened after removing 629 duplicates. 1898 articles were excluded as they did not meet the inclusion criteria. 93 full-text articles were assessed for eligibility and 24 articles were excluded when scanning the full text. As a result, 69 studies were included in the systematic review. Among the 69 studies, 34 studies provided enough information to construct contingency tables and calculate test performance parameters, and were therefore included in the meta-analysis.

These 69 studies described 72 patient cohorts. In these studies, target conditions were divided into LNM (45 studies) and DM (26 studies) (2 studies involved both LNM and DM), which included bone metastasis (13 studies), brain metastasis (3 studies), liver metastasis (4 studies), lung metastasis (2 studies) and others (4 studies). Primary tumors comprised breast cancer (10 studies), head and neck cancer (9 studies), gastric cancer (7 studies), lung cancer (6 studies), colorectal cancer (5 studies), prostate cancer (3 studies) and other primary tumors (6 studies). Thirteen studies did not report this. In addition, 10 studies contained several different primary tumors. Study characteristics are shown in Tables 1, 2 and 3. All included studies used retrospective data and were not open-access. Seven studies excluded low-quality images which meant that the location and size of the lesion on the images did not match that seen at pathologic examination or one or two of the most representative images were selected for each patient, while 62 studies did not report this. Comparison between AI models and health-care professionals by the same test set was only provided in 8 studies. As for the verification of the model, 7 studies collected out-of-sample dataset to do an external validation, and the others were internally validated. Furthermore, different algorithms including DL (23 studies) and ML (34 studies) were included in the systematic review. Four studies used both DL and ML algorithms and 5 studies did not report the detailed types of algorithms.

Fig. 1. Study selection.

We accepted all forms of the reference standard for the diagnosis of metastasis. Forty-three studies used histopathology; 21 studies used varying models of expert evaluation; 10 studies used other imaging types to confirm the diagnosis; 7 studies used existing clinical notes; 4 studies used clinical follow-up, and 1 study did not report this. A part of studies applied several different references.

A total of 34 studies and 123 contingency tables were included in the meta-analysis. In these studies, primary tumors included breast cancer (7 studies), head and neck cancer (7 studies), gastrointestinal cancer (4 studies), lung cancer (5 studies) and others (3 studies). 4 studies had several different primary tumors; 4 studies did not report this. There were 25 studies targeting LNM and 10 studies targeting DM (1 study related to both LNM and DM). None of the 8 studies included in the systematic review with comparison between AI models and health-care professionals were excluded in the meta-analysis. After removing 3 from the 7 studies included in the systematic review with external validation because of the lack of contingency tables, only 4 studies were used for the meta-analysis.

In addition, we investigated the international research situation of this subject, finding that the studies mostly concentrated on China, America and Japan, with 31, 11 and 11 studies respectively. Included studies were also widely distributed in South Korea and Europe, South America, Australia and the Middle East had some sporadic distribution as well (Fig. 2).

The quality of studies included in the meta-analysis was assessed by the QUADAS-2 score [29] (Supplementary figure 1). Three and 5 studies showed a high risk respectively for patient selections and reference standards because these studies did not clarify whether enrolled patients were consecutive or use non-histopathology methods as reference standard, which we think were acceptable. So, these studies were not excluded.
Table 1
Participant demographics for the 69 included studies.

| First author and year | Participants Inclusion criteria | Exclusion criteria | Patient/Sample | Positive Patients(samples)/Negative Patients(samples) | Mean age (SD; range), year | Percentage of male participants |
|------------------------|---------------------------------|-------------------|----------------|------------------------------------------------------|---------------------------|---------------------------------|
| Mitsu Koizumi et al. (2020) [40] | NR | Skeletal metastasis did not meet the criteria of the term 'disseminated'; no skeletal metastasis | 54/54 | 54(NR)/0(NR) | NR | NR |
| Jing Li et al. (2020) [41] | Patients underwent gastrectomy plus lymph node dissection and were diagnosed gastric adenocarcinomas; patients were scanned with GSI mode; without any local or systematic treatment before CT scans and surgery; with definite postoperative pathologic data. | Invisible lesion on CT images; with a minimum diameter of tumor less than 5 mm insufficient to outline a valid ROI; insufficient stomach distention; poor image quality for post-processing. | 204/NR | 122(NR)/82(NR) | Training set:59(12;28–81) Test set:59(11;28–74) | Training set:72% Test set:72% |
| L. Zhang et al. (2020) [42] | NR | NR | 51/NR | 32(NR)/19(NR) | NR | 47% |
| Li-Qiang Zhou et al. (2020) [43] | Patients with histologically confirmed primary breast cancer who underwent surgery; T1 or T2 primary breast cancer with clinically negative LNs and no preoperative therapy; standard preoperative breast US | T3 or T4 stage; physically positive LNs; imaging positive LNs; preoperative therapy; low quality US images | Cohort1: 756/974 Cohort 2: 78/81 | Training set:343(441)/337(436) Testing set: internal validation:37(49)/39 (48) external validation:40(43)/38 (38) | Training set:48(NR;24–81) Test set: internal validation:50 (NR;25–82) external validation:46 (NR;30–74) | NR |
| Endre Grøvik et al. (2020) [44] | The presence of known or possible metastatic disease; no prior surgical or radiation therapy; the availability of all required MRI sequences; patients with $\geq$1 metastatic lesion | NR | 156/156 | 156(156)/0(0) | 63(12;29–92) | 33% |
| Yu Zhao et al. (2019) [45] | Patients with metastatic castration-resistant prostate cancer | Unqualified imaging quality of 3D T1 MPRAGE; data missing; skull metastases and meningeal metastases | Dataset 1: 1201/1201 Dataset 2: 231/231 Dataset 3: 220/220 | Dataset 1:1201(1201)/0(0) Dataset 2:231(231)/0(0) Dataset 3:220(220)/0(0) | Dataset 1:58(18;NR) Dataset 2:60(18;NR) Dataset 3:59(15;NR) | Dataset 1:57% Dataset 2:53% Dataset 3:52% |
| Jie Xue et al. (2019) [46] | Definitely pathohistological results of the primary tumor lesion; patients with only metatatic lesions in brain; with an age over 18 years old; 3D T1 MPRAGE sequence was acquired. | Absence of contrast-enhanced CT imaging data after chemotherapy and prior to post-chemotherapy LN dissection; insufficient image quality; insufficient matching of histopathology to the individual LNs | Dataset 1: 1201/1201 Dataset 2: 231/231 Dataset 3: 220/220 | Dataset 1:1201(1201)/0(0) Dataset 2:231(231)/0(0) Dataset 3:220(220)/0(0) | Dataset 1:58(18;NR) Dataset 2:60(18;NR) Dataset 3:59(15;NR) | Dataset 1:57% Dataset 2:53% Dataset 3:52% |
| Bettina Baessler et al. (2019) [47] | Patients with retroperitoneally metastasized testicular germ cell tumors prior to post-chemotherapy LN dissection | NR | 193/NR | 193(NR)/0(NR) | 69.6(7.9;NR) | NR |
(continued on next page)
| First author and year                  | Participants   | Inclusion criteria                                                                 | Exclusion criteria                                                                 | Patient/Sample | Positive Patients(samples)/Negative Patients(samples) | Mean age (SD; range), year | Percentage of male participants |
|---------------------------------------|----------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|----------------|--------------------------------------------------------|---------------------------|--------------------------------|
| Xiaojun Yang et al. (2019)[48]*       |                | Preoperative contrast-enhanced CT images within 2 weeks before surgery; histologically confirmed primary invasive breast cancer; SLN biopsy (and ALND); pathologically results after operation confirmed SLN metastasis | Nestadjuvant therapy before CT examination and surgery; poor visualization of the tumor for segmentation due to serious artifacts caused by metallic foreign bodies on the breast; tumor was too small to be seen on CT images; incomplete clinicopathological data | 348/348        | Training set:71(71)/113(113) Testing set:63(63)/101(101) | Training set: SLN-P:52(9;NR); SLN – N:50(11;NR) Testing set: SLN-P:50(10;NR); SLN – N:53(10;NR) | NR                            |
| Yuan Gao et al. (2019) [49]           | NR             |                                                                                     | No metastatic LNs revealed by CT; with preoperative neoadjuvant radio-chemotherapy; complicated with abdominal infection; pathological grouping different from CT grouping; LN adhesions | 602/38,495     | NR                                                     | 62(NR;20 – 91)          | 72%                           |
| David Coronado-Gutierrez et al. (2019)[50]* |                | Positive metastatic nodes by ultrasound-guided FNA or CNB; Negative metastatic nodes determined by histopathology | Surgical biopsy showed positive result after not suspicious nodes in ultrasound exam or negative results of ultrasound-guided FNA or CNB; Patients refused to receive SLNB | 127/118        | NR[53](NR;65)                                         | 54.6 (NR;26–91)          | NR                            |
| Yukinori Okada et al. (2019)[51]      | NR             |                                                                                     |                                                                                     | 56/NR          | 56(NR)/0(0)                                           | 59 (12.7;NR)              | 0                             |
| Jeong Hoon Lee et al. (2019) [52]*    | NR             |                                                                                     |                                                                                     | 202/995        | NR                                                     | 202(995)/20(995)         | NR                            |
| Jansen et al. (2019)[53]              | NR             |                                                                                     |                                                                                     | 111/111        | 72(NR)/39(NR)                                         |                          | NR                            |
| Chuangming Li et al. (2019) [54]*     | NR             | Patients with oropharyngeal squamous cell carcinoma; underwent neck dissections; had preoperative PET and CT imaging | MRI examination data were incomplete, or image quality was poor | 62/62          | 35(NR)/27(NR)                                         | SLN-P:48.14 (8.35; NR) SLN—N:49.78 (12.53; NR) | NR                            |
| M. Dohopolski et al. (2019)[55]       | NR             |                                                                                     |                                                                                     | 129/543        | NR                                                     |                          | NR                            |
| Yige Peng et al. (2019) [56]*         | NR             | Definitive diagnosis by histopathology                                              | No detailed metastases information Neoadjuvant chemotherapy or radiotherapy or endoscopic resection; end-stage disease or severe complications precluding surgery; disease that could not be detected on imaging; poor imaging quality or poor gastric resection | 48/NR          | 24(NR)/24(NR)                                         |                          | NR                            |
| Qiuxia Feng MD et al. (2019) [57]*    | NR             |                                                                                     |                                                                                     | 490/NR         | 279(NR)/211(NR)                                       | 61.8(10.4; NR)            | Training and validation set: 73% Test set: 77% |
| First author and year | Participants inclusion criteria | Exclusion criteria | Patient/Sample | Positive Patients(samples)/ Negative Patients(samples) | Mean age (SD; range), year | Percentage of male participants |
|-----------------------|-------------------------------|-------------------|----------------|--------------------------------------------------------|--------------------------|-------------------------------|
| Thoma Schnelldorfer et al. (2019) [58] | Underwent a laparoscopic operation with the initial intent for either resection or palliation of the underlying malignancy; Video recordings of the operation were available | Malignancy originating from esophageal, hepatic and colorectal malignancies | 35/35 | 20(20)/15(15) | 67 (NR; 44–85) | 66% |
| Samir D. Mehta et al. (2019) [59]* | Underwent CT of the abdomen and pelvis or radiographs of the lumbar spine and DEXA studies; CT studies/ lumbar spine radiographs were performed not more than 1 year prior to the DEXA study | NR | 200/NR | 45(NR)/155(NR) | Case: 70.5 (NR; 63.9–76.7) Control: 62 (NR; 53.5–69) | Case: 78% Control: 83% |
| Yoshiko Ariji, et al. (2019) [60] | Underwent intravenous contrast enhanced CT and dissection of cervical lymph nodes | NR | 45/441 | NR(127)/(NR(314) | 63 (NR; 33–95) | 53% |
| Yunpeng Zhou et al. (2019) [61] | Definite lymph node metastasis reported by preoperative imaging; Received radical colectomy with lymph node dissection; Patients with colon cancer diagnosis; Patients with no history of previous or coexisting other malignancies; Patients who underwent preoperative enhanced CT for local colon cancer staging and for liver metastasis diagnosis; | With a history of abdominal pelvic surgery, and pelvic radiochemotherapy | 301/12,060 | 301(NR)/0(NR) | 59.5(NR; NR) | 75% |
| Yu Li et al. (2019) [62]* | Training set: NR (91)/NR (287) Test set: NR (25)/NR (234) | NR | 75/257 | NR | 69(9; NR) | NR |
| Zhiguo Zhou et al. (2019) [63]* eMine acar et al. (2019) [64] | Sclerotic lesions >2 cm in patients with at least three sclerotic metastatic lesions; sclerosis areas of the bones that located on the surface of the joint and/or on the surface of the other side of the joint; osteophytes not considered as metastasis. | No bone metastasis; <3 bone metastasis; no sclerotic metastasis; uptake-liver uptake | 129/543 | Training set: NR (91)/NR (287) Test set: NR (25)/NR (234) | NR | NR |
| Fang Hou et al. (2019) [65]* | Oral squamous cell carcinoma; underwent neck dissection; pathology confirms cervical lymph node metastasis | NR | 28/573 | Training set: NR (21)/NR (293) Test set: NR (25)/NR (234) | NR | NR |
| Yoshiko Ariji et al. (2019) [66]* | Oral squamous cell carcinoma; underwent neck dissection; pathology confirms cervical lymph node metastasis | NR | 54/143 (LN) 703 (image) | Training set: NR (33)/NR (110) Test set: NR (25)/NR (234) | NR | 52.94% |
| First author and year                  | Participants Inclusion criteria                                                                 | Exclusion criteria                                                                 | Patient/Sample | Positive Patients(samples)/Negative Patients(samples) | Mean age (SD; range), year | Percentage of male participants |
|---------------------------------------|-------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|----------------|-------------------------------------------------------|---------------------------|---------------------------------|
| Xiaojuan Xu et al. (2019) [67]        | Patients who received standard FIGO surgical staging for endometrial cancer between January 2011 and December 2017 | Patients without DCE-MRI 2 weeks before surgery; patients with serious MR artifacts and without uniform MR scanner; patients missing clinical characteristics data and endometrial biopsy histological information; patients with any preoperative therapy; patients suffering from other malignant tumor diseases concurrently | 200/NR         | 67/NR/133(NR)                                        | Training cohort: pN(+):55.7(NR;NR) pN(-):55.7 (NR;NR) Test Cohort: pN(+):57.4(NR;NR) pN(-):51.7(NR;NR) | NR                              |
| Jiaxiu Luo et al. (2018) [68]*         |                                                                                                 |                                                                                      | 172/NR         | 74(NR)/98(NR)                                        | NR                        | NR                              |
| Richard Ha et al. (2018) [69]          |                                                                                                 |                                                                                      | 275/275        | 133(133)/142(142)                                    | NR                        | NR                              |
| B.H. Kann et al. (2018) [70]*          |                                                                                                 |                                                                                      | 270/653        | NR(380)/NR(273)                                      | NR                        | NR                              |
| Jeong Hoon Lee et al. (2018) [71]*     |                                                                                                 |                                                                                      | 804/812        | Training set: NR (286)/NR (263) Validation set: NR (33)/NR (30) Test set: NR (100)/NR (100) Training set:35(1)/28(0)/0(0) Test set: NR (0) | Training & Validation set:44 (NR;13 – 84) Training set:55(NR;10 – 81) Test set:27% | NR                              |
| Yun Lu et al. (2018) [72]              |                                                                                                 |                                                                                      | 41/36,000      | LNM: Test set:23(NR)/29(NR) Validation set:9(NR)/7(NR) DM: Test set:8(NR)/44(NR) Validation set:5(NR)/11(NR) | Test set:66.6(9.1; 41 – 85) Validation set:64.88(9.1; 41 – 79) | NR                              |
| Jose Raniery Ferreira Junior et al. (2018) [73]* | No standard contrast-enhanced CT protocol; did not present all clinical data; presented other opacities attached to the tumor |                                                                                      | 70/75          | 70(75)/0(0)                                          | Test set:57.7% Validation set:62.5% | NR                              |
| Tzu-Yun Lo et al. (2018) [74]          |                                                                                                 |                                                                                      | 70/619         | Original data: NR(307)/NR(312) augmented data: NR(1535)/NR (1560) | NR                        | NR                              |
| Jin Li et al. (2018) [75]              |                                                                                                 |                                                                                      | 70/153         | NR(87)/NR(66)                                        | NR                        | NR                              |
| Mohamed Amine Lahmann et al. (2018) [76] | Underwent surgical resection and systematic LN dissection according to the American Thoracic Society criteria; had no enlargement of the hilar or mediastinal LNs at CT (enlargement defined as short axis of a node ≥ 10 mm on axis images) and clinical NO; no distant metastasis |                                                                                      | 492/492        | 78(78)/41(414)                                       | 61.4(9.7; NR) N-LNM:61.28(9.8; NR) LNM:61.7(9.6; NR) | 35% N-LNM:32% LNM:50% |
| Yan Zhong et al. (2018) [77]*          |                                                                                                 |                                                                                      | 168/1397       | NR(127)/NR(1270)                                     | 61(NR;38 – 81)            | NR                              |
| Wang, H et al. (2017) [78]*            |                                                                                                 |                                                                                      | 265/265        | 124(124)/101(101)                                    | NR                        | NR                              |
| Mitsuhiro Koizumi et al. (2017) [79]*   |                                                                                                 |                                                                                      | 26/NR          | 26(NR)/0(NR)                                         | 58(14; NR)                | NR                              |
| Juan Wang et al. (2017) [80]           |                                                                                                 |                                                                                      | 26/NR          | 26(NR)/0(NR)                                         | 58(14; NR)                | 54%                             |
| First author and year | Participants | Inclusion criteria | Exclusion criteria | Patient/Sample | Positive Patients(samples)/Negative Patients(samples) | Mean age (SD; range), year | Percentage of male participants |
|-----------------------|--------------|--------------------|-------------------|----------------|------------------------------------------------------|---------------------------|-------------------------------|
| Zhi-Long Wang et al. (2017) [81] | NR | Pathologically proven adenocarcinoma, small cell carcinoma, mixed cancer, or other diseases; other preoperative therapies simultaneously; esophageal multiple primary carcinoma; death within 30 days after surgery; enhanced CT data before preoperative chemotherapy not obtained or images not interpretable; non-suitability for radical esophagectomy | 131/NR | 51(NR)/80(NR) | 58(NR;42 – 75) | 77.90% |
| Tuan D. Pham et al. (2017) [82] | Biopsy-proven primary lung malignancy with pathological mediastinal nodal staging; Underwent axilla conventional US and RTE simultaneously | NR | Patients with nodal biopsy more than three months from CT | 148/NR | Test set: NR (133)/NR (138) | 69.4(NR;36 – 84) | 63% |
| Qi Zhang et al. (2017) [83] | Underwent axilla conventional US and RTE simultaneously | NR | Take neoadjuvant therapy before SLNB or ALND | 158/161 | NR (92)/NR (69) | 55.2(5.2;21 – 81) | NR |
| Yu-wen Wang et al. (2016) [84] | Underwent axilla conventional US and RTE simultaneously | NR | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | NR (24)/NR (81) | 57(NR;30 – 87) | 60% |
| Ali Aslantas et al. (2016) [85] | Underwent axilla conventional US and RTE simultaneously | NR | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | NR (24)/NR (81) | 57(NR;30 – 87) | 60% |
| Aneta Chmielewski et al. (2015) [86] | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | NR (24)/NR (81) | 57(NR;30 – 87) | 60% |
| Mitsuru Koizumi et al. (2015) [87] | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | NR (24)/NR (81) | 57(NR;30 – 87) | 60% |
| Mitsuru Koizumi et al. (2015) [88] | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | NR (24)/NR (81) | 57(NR;30 – 87) | 60% |
| Nesrine Trabelsi et al. (2015) [89] | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | NR (24)/NR (81) | 57(NR;30 – 87) | 60% |
| Xuan Gao et al. (2015) [90] | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | NR (24)/NR (81) | 57(NR;30 – 87) | 60% |
| Osamu Tokuda, et al. (2014) [91] | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | NR (24)/NR (81) | 57(NR;30 – 87) | 60% |
| Ari Seff et al. (2014) [92] | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | NR (24)/NR (81) | 57(NR;30 – 87) | 60% |
| Zhi-Guo Zhou et al. (2013) [93] | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | NR (24)/NR (81) | 57(NR;30 – 87) | 60% |
| Seungwook Yang et al. (2013) [94] | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | Underwent surgical treatment for invasive breast cancer with axillary lymph node evaluation | NR | NR (24)/NR (81) | 57(NR;30 – 87) | 60% |

(continued on next page)
| First author and year | Participants | Inclusion criteria | Exclusion criteria | Patient/Sample | Positive Patients(samples)/Negative Patients(samples) | Mean age (SD; range), year | Percentage of male participants |
|-----------------------|--------------|-------------------|-------------------|---------------|-----------------------------------------------------|---------------------------|-------------------------------|
| Jianfei Liu et al. (2013) [95] | NR | NR | 50/NR | Training set: NR; Test set:44/102 (NR) | 28 (95)/0(NR) | NR | NR |
| Yoshihiko Nakamura et al. (2013) [96] | NR | NR | 28/NR | | | | |
| Chuan-Yu Chang et al. (2013) [97] | NR | NR | 6/177 | | | | |
| Johannes Feulner et al. (2013) [98] | NR | NR | 54/1086 | | | | |
| Chao Li et al. (2012) [99] | NR | NR | 38/NR | 27/11(NR) | | | |
| Hongmin Cai et al. (2012) [100] | NR | NR | 228/NR | | | | |
| Shao-Jer Chen et al. (2012) [101] | NR | NR | 38/NR | | | | |
| Xiao-Peng Zhang et al. (2011) [102]* | Patients received radical gastrectomy and D2 lymph nodes dissection; Preoperatively examined with multi-detector row CT; Confirmed as gastric cancer by postoperative histopathology | Received preoperative neoadjuvant therapy; Distant metastasis was found in the preoperative examination or in the operation | 175/NR | 134(NR)/41(NR) | | 60.6 (12.1; 25~87) | 61% |
| Matthias Dietzel et al. (2010) [103] | Invasive breast lesions with histopathological verification after bMRI | With a history of breast biopsy/ interventions (surgical or minimally invasive) and chemotherapy/radiation therapy up to 12 months before bMRI; Histopathological grading not possible | 194/NR | 97(NR)/97(NR) | | 60.6 (12.1; 25~87) | NR |
| May Sadik et al. (2008) [104]* | Underwent whole-body bone scintigraphy with a dual-detector r-camera; Patients with a complete set of technically sufficient images; At least 1 yr follow-up bone scan | With a urine catheter, large bladder, sternotomy or fracture that could be misleading for the CAD system | NR/869 | NR/297(NR)/572 | Training set: 66 (NR;25~92) | Training: 65% Test: 69% All: 62% |
| Junji Shiraishi et al. (2008) [105] | NR | NR | 97/103 | NR(26);NR(77) | | | |
| Junhua Zhang et al. (2008) [106]* | NR | NR | 112/210 | NR(114);NR(96) | | 53 (17;17~81) | NR |
| Rie Tagaya et al. (2008) [107]* | NR | NR | 91/91 | | | | NR |
| K. Marten et al. (2004) [108] | Patients with pulmonary metastasis; undergoing clinical staging and follow-up CT examinations of the chest | NR | 20/135 | | | | 62.4(NR)/NR | |

Abbreviation: NR=not reported. CT=computed tomography. GSI=Gemstone spectral imaging. LN=Lymph node. US=ultrasound. 3D-T1-MPRAGE images=Three-dimensional T1 magnetization prepared rapid acquisition gradient echo. SNL=sentinel lymph node. ALND=axillary lymph node dissection. FDG-PET/CT=fluoro-deoxy glucose positron emission tomography with CT. MRT=magnetic resonance imaging. FNA= fine needle aspiration. CNB= core needle biopsy. DW-MRI=diffusion-weighted magnetic resonance imaging. DCE-MRT=contrast-enhanced magnetic resonance imaging. OPSSC= oropharyngeal squamous cell carcinoma. DEXA=Dual-energy X-ray absorptiometry. HNC=head and neck cancer. DCE-MRI= dynamic contrast enhanced MRI. FIGO=International Federation of Gynecology and Obstetrics. RTE=real-time elastography. NPC=nasopharyngeal carcinoma. CAD=computer-assisted diagnosis.

* 34 studies included in the meta-analysis.
Table 2
Model training and validation for the 69 included studies.

| First author and year | Metastasis type | Target condition | Primary tumor | Reference standard | Type of internal validation | External validation |
|------------------------|-----------------|------------------|---------------|--------------------|----------------------------|---------------------|
| Mitsuru Koizumi et al. [40] | DM | Disseminated skeletal metastasis | prostate cancer (n = 12), GC= (n = 15), breast cancers (n = 10) | Expert consensus | NR | YES |
| Jing Li et al. [41] | LNM | LNM in GC | GC | Histopathology; follow up | Resampling method | NO |
| L. Zhang et al. [42] | DM | Lung metastasis in STS | STS | Histopathology | Random split sample validation | NO |
| Li-Qiang Zhou et al. [43] | LNM | Clinically negative axillary lymph node metastasis in primary breast cancer | Breast cancer | Histopathology | NR | YES |
| Endre Grøvik et al. [44] | DM | Detection and Segmentation of Brain Metastases | Lung (n = 99), breast (n = 33), melanoma (n = 7), gastrointestinal (n = 5), and miscellaneous-cancers (n = 5) | Expert consensus | NR | NO |
| Yu Zhao et al. [45] | DM & LNM | Bone metastasis, lymph node metastasis in prostate cancer | Metastatic castration-resistant prostate cancer | Expert consensus | NR | NO |
| Jie Xue et al. [46] | DM | Detection and Segmentation of Brain Metastases | Lung, Breast, Kidney, Other organs (rectum, colon, melanoma, ovary and liver) | Expert consensus | Resampling method | NO |
| Bettina Baessler et al. [47] | LNM | LNM in NSTGCT patients | Based on CT, MRI and clinical findings: expert consensus | NR |NR |
| Xiaojun Yang et al. [48] | LNM | SLNM in Breast Cancer | Breast cancer | Histopathology | Resampling method | NO |
| Yuan Gao et al. [49] | LNM | PCMLNs in GC | Breast cancer | Histopathology; expert consensus | Resampling method | NO |
| David Coronado-Gutierrez et al. [50] | DM | Metastasis in the axillary lymph node | Breast cancer | Based on CT, MRI and clinical findings: expert consensus | Random split sample validation | NO |
| Yukinori Okada et al. [51] | DM | Bone metastasis | Breast cancer | Histopathology by FNA and/or surgery | Resampling method | NO |
| Jeong Hoon Lee et al. [52] | LNM | Metastasis in the cervical lymph node | Thyroid cancer | Histopathology by FNA and/or surgery | Random split sample validation | NO |
| Jansen et al. [53] | DM | Liver metastasis | NR | Expert consensus | NR | NO |
| Chuangming Li et al. [54] | LNM | Sentinel lymph node metastasis | Breast cancer | Histopathology; expert consensus | NR | NO |
| M. Dohopolski et al. [55] | LNM | Small Lymph node metastasis | Oropharyngeal squamous cell carcinoma | Histopathology | NR | NO |
| Yige Peng et al. [56] | DM | Distant metastasis in STS | Gastric adenocarcinoma: 19. Pancreatic adenocarcinoma: 11; Gastric adenocarcinoma: 1; Gallbladder carcinoma: 2. Metastatic pancreatic neuroendocrine tumor, jejunal adenocarcinoma, ampullary adenocarcinoma: 1 each | Histopathology | NR | NO |
| Qiuxia Feng MD et al. [57] | LNM | LNM in GC | Biopsy or CT and/or PET images | Histopathology | NR | NO |
| Thoma Schnelldorfer et al. [58] | DM | Distinguish metastasis in the peritoneal from the benign lesions | Gastric adenocarcinoma: 19. Pancreatic adenocarcinoma: 11; Gallbladder carcinoma: 2. Metastatic pancreatic neuroendocrine tumor, jejunal adenocarcinoma, ampullary adenocarcinoma: 1 each | Histopathology | NR | NO |

(continued on next page)
| First author and year | Metastasis type | Target condition | Primary tumor | Reference standard | Type of internal validation | External validation |
|-----------------------|----------------|-----------------|--------------|--------------------|---------------------------|--------------------|
| eMine acar et al. (2019) [64] | DM | Differentiating metastatic and completely responded sclerotic bone lesion in prostate cancer | Prostate cancer | Expert consensus | Resampling method | NO |
| Fang Hou et al. (2019) [65]* | LNM | LNM | NR | Histopathology | NR | NO |
| Yoshiko Arji et al. (2019) [66]* | LNM | LNM in Oral squamous cell carcinoma | Oral squamous cell carcinoma | Histopathology | NR | NO |
| Xiaojuan Xu et al. (2019) [67] | LNM | LNM in EC | EC | Histopathology | NR | NO |
| Jiaxiu Luo et al. (2018) [68]* | LNM | LNM in breast cancer | Breast cancer | Biopsy; follow up | Resampling method | NO |
| Richard Ha et al. (2018) [69] | LNM | LNM in breast cancer | Breast cancer | Biopsy | Resampling method | NO |
| B.H. Kann et al. (2018) [70]* | LNM | LNM in HNC | HNC | Histopathology | Resampling method | NO |
| Jeong Hoon Lee et al. (2018) [71]* | LNM | LNM in thyroid tumor | Thyroid tumor | FNA and/or laboratory tests | Random split sample validation | NO |
| Yun Lu et al. (2018) [72] | LNM | Pelvis LNM in rectal cancer | Rectal cancer | Expert consensus | Random split sample validation | YES |
| José Raniery Ferreira Junior et al. (2018) [73]* | DM& LNM | LNM and distant metastasis in lung cancer | Lung cancer | Clinical notes | Resampling method | NO |
| Tzu-Yun Lo et al. (2018) [74] | LNM | LNM in HNC | HNC | Clinical notes | Resampling method | NO |
| Jin Li et al. (2018) [75] | LNM | LNM in Colorectal Cancer | Colorectal Cancer | Expert consensus | Resampling method | NO |
| Mohamed Amine Larhmam et al. (2018) [76] | DM | Spine metastasis | NR | Single expert | Resampling method | NO |
| Yan Zhong et al. (2018) [77]* | LNM | Occult mediastinal LNM of lung adenocarcinoma | Lung adenocarcinoma | Histopathology | Resampling method | NO |
| Wang, H et al. (2017) [78]* | LNM | Mediastinal LNM of non-small cell lung cancer | Non-small cell lung cancer | Histopathology | Resampling method | NO |
| Mitsuru Koizumi et al. (2017) [79]* | DM | Skeletal metastasis in prostate cancer | Prostate cancer | BSS&CT expert consensus; follow up; and/or biopsy | NR | YES |
| Juan Wang et al. (2017) [80] | DM | Spinal metastasis | 15 lung, 5 thyroid, two liver, 1 breast, 1 prostate, 1 esophageus, 1 urinary tract | Biopsy | Resampling method | NO |
| Zhi-Long Wang et al. (2017) [81] | LNM | LNM in esophageal cancer with preoperative chemotherapy | Esophageal cancer | Postoperative pathological results | Random split sample validation | NO |
| Tuan D. Pham et al. (2017) [82]* | LNM | Mediastinal lymph nodes in lung Cancer | Lung cancer | Histopathology | Resampling method | NO |
| Qi Zhang et al. (2017) [83]* | LNM | Auxiliary lymph node metastasis in breast cancer | Breast cancer | Histopathology | Resampling method | NO |
| Yu-wen Wang et al. (2016) [84]* | LNM | Metastasis in the retropharyngeal lymph nodes | NPC | MRI follow-up | Random split sample validation | NO |
| Ali Aslantas et al. (2016) [85]* | DM | Bone metastatic | Chest, prostate, lung cancers | Single expert (laboratory tests, and other accessible radiographic images) | Resampling method | NO |
| Aneta Chmielewski et al. (2015) [86]* | LNM | Axillary lymph node metastasis in breast cancer patients | Breast cancer | Imaging-pathology gold standards: FNA, biopsy, LND, normal image with long term follow-up | Resampling method | NO |
| Mitsuru Koizumi et al. (2015) [87]* | DM | Metastasis in bone | Prostate cancer, lung cancer, breast cancer, and other cancers | Radiology (CT, MR or PET/CT), follow-up scan and patients' clinical course | NR | YES |
| Mitsuru Koizumi et al. (2015) [88] | DM | Metastasis in bone | Prostate cancer, lung cancer, breast cancer, and other cancers | Radiology (CT, MR or PET/CT), follow-up scan and patients' clinical course | NR | YES |
| Nesrine Trabelsi et al. (2015) [89] | DM | Metastasis in liver | NR | NR | NR | NO |
| Xuan Gao et al. (2015) [90] | LNM | Mediastinal lymph nodes in lung cancer | Lung cancer | Histopathology | Random split sample validation | NO |
| First author and year                  | Metastasis type | Target condition                          | Primary tumor                                           | Reference standard                                                                 | Type of internal validation | External validation |
|----------------------------------------|-----------------|-------------------------------------------|---------------------------------------------------------|-------------------------------------------------------------------------------------|-----------------------------|---------------------|
| Osamu Tokuda et al. (2014) [91]        | DM              | Bone metastasis                          | Prostatic cancer (N = 71), breast cancer (N = 109), other cancers (N = 226) | All bone-scan images, including the follow-up scans, expert consensus; laboratory tests; (OR) biopsy | NR                          | YES                 |
| Ari Seff et al. (2014) [92]            | LNM             | LNM                                      | NR                                                      | Expert consensus                                                                    | Resampling method           | NO                  |
| Zhi-Guo Zhou et al. (2013) [93]*       | LNM             | LNM in GC                                | GC                                                      | Surgery and histopathology                                                           | Resampling method           | NO                  |
| Seungwook Yang et al. (2013) [94]*     | DM              | Brain metastases                         | NR                                                      | Single expert                                                                       | NR                          | NO                  |
| Jianfei Liu et al. (2013) [95]         | DM              | Ovarian Cancer Metastases                | Ovarian Cancer                                          | Single expert                                                                       | NR                          | NO                  |
| Yoshihiko Nakamura et al. (2013) [96]  | LNM             | Abdominal Lymph Node                     | 5 colorectal; 23 stomach cancer                         | Single expert 26 cases: single expert 2 cases: experts consensus using a particular medical image | Resampling method           | NO                  |
| Chen Yu Chang et al. (2013) [97]       | LNM             | LNM                                      | NR                                                      | Histopathology                                                                      | NR                          | NO                  |
| Johannes Feulner et al. (2013)         | LNM             | Mediastinal lymph nodes                  | NR                                                      | Single expert                                                                       | Resampling method           | NO                  |
| Chao Li et al. (2012) [99]             | LNM             | LNM in GC                                | GC                                                      | Histopathology                                                                      | NR                          | NO                  |
| Hongmin Cai et al. (2012) [100]        | LNM             | Regional LNM                             | Rectal cancer                                           | Histopathology                                                                      | Resampling method           | NO                  |
| Shao-Jer Chen et al. (2012) [101]      | LNM             | LNM                                      | NR                                                      | Histopathology; follow up                                                           | Resampling method           | NO                  |
| Xiao-Peng Zhang et al. (2011) [102]*   | LNM             | LNM in GC                                | GC                                                      | Histopathology                                                                      | Resampling method           | NO                  |
| Matthias Dietzel et al. (2010) [103]   | LNM             | Metastasis to the ipsilateral axilla lymph node | Breast cancer                                           | Surgicopathology                                                                    | Random split sample validation | NO                  |
| May Sadik et al. (2008) [104]*         | DM              | Metastasis to bone                       | Testing: Breast/prostate cancer                         | Training: Clinical reports and the bone scan images Testing: Final clinical assessments made by the same experienced physician | NR                          | NO                  |
| Junji Shiraishi et al. (2008) [105]    | DM              | Metastasis to the liver                  | NR                                                      | Biopsy or surgical specimens                                                       | NR                          | NO                  |
| Junhua Zhang et al. (2008) [106]*      | LNM             | Metastasis to the cervical lymph nodes   | NR                                                      | Histopathology                                                                      | Resampling method           | NO                  |
| Rie Tagaya et al. (2008) [107]*        | LNM             | Diagnosis of LNM by B-Mode Images from Convex-Type Echobronchoscopy | 66 lung cancer,25 sarcoidosis                          | Histopathology or cytologic testing                                                 | NR                          | NO                  |
| K. Marten et al. (2004) [108]          | DM              | Pulmonary nodules                         | NR                                                      | Expert consensus                                                                    | NR                          | NO                  |

Characteristics only be described in 1 or 2 studies are classified to others. Abbreviation: NR=not reported. LNM=Lymph node metastasis. DM=distant metastasis. BS=bone scintigraphy. GC=gastric cancers. STS=soft-tissue sarcoma. NSTGCT=Non-seminomatous testicular germ cell tumor. PGMLNs=peri-gastric metastatic lymph nodes. EC=Endometrial cancer. FNA=fine needle aspiration.

* 34 studies included in the meta-analysis.
Table 3: Indicator, algorithm, and data source for the 69 included studies.

| First author and year | Method for predictor measurement | Exclusion of poor-quality imaging | Heatmap provided | Extracted features | Algorithm architecture name | Algorithm architecture | Transfer learning applied | Source of data | Number of images for training/testing | Data range | Open access data |
|-----------------------|----------------------------------|----------------------------------|------------------|-------------------|-----------------------------|------------------------|--------------------------|----------------|---------------------------------------|------------|-----------------|
| Mitsuru Koizumi et al. (2020)[40] | IS | NR | NR | NO | NR | ANN | NR | Retrospective clinical data from cancer institute hospital, Tokyo, Japan | NR/54 | 2013.1–2019.8 | NO |
| Jing Li et al. (2020) [41] | dual-energy CT | YES | NR | YES | DCNNs; ANN; K SVM | CNN; ANN; SVM | NR | Retrospective cohort | 136/68 | 2012.1–2018.11 | NO |
| L. Zhang et al. (2020) [42] | MRI, CT | NR | NR | NO | Inception V3 | CNN; Inception | YES | Data collected from Cancer Imaging Archive | 25/15 | NR | YES |
| Li-Qiang Zhou et al. (2020) [43] | US image | YES | YES | NO | Inception V3; Inception-ResNet V2; ResNet-101 | CNN; Inception; Residual Network | NR | Cohort 1: retrospective cohort collected from Tongji Hospital; Cohort 2: retrospective cohort collected from Hubei Cancer Hospital (Hubei, China) 1: 2016.5–2018.10; Cohort 2: 2018.10–2019.4 | 877/97 (internal test) +81 (external test) | Cohort |
| Endre Grøvik et al. (2020) [44] | Multisequence MRI | NR | YES | NO | GoogLeNet | CNN | NR | Retrospective cohort | 100/51 | 2016.6–2018.6 | NO |
| Yu Zhao et al. (2019) [45] | PSMA PET/CT, CT | NR | NR | NR | triple combing 2.5D U-NET | CNN | NR | Retrospective cohort from medical centers of Technical University of Munich, University of Munich and University of Bern | 130/63 | NR | NR |
| Jie Xue et al. (2019) [46] | 3D-T1-MPRAGE images | YES | NR | NO | 3D CNN | CNN | NR | Dataset 1: Retrospective clinical data from the Shandong Provincial Hospital Affiliated to Shandong University; Dataset 2: Retrospective clinical data from the Affiliated Hospital of Qingdao University Medical College; Dataset 3: Retrospective clinical data from the Second Hospital of Shandong University 1: 2016.10–2019.5; Dataset 2: 2017.8–2019.5; Dataset 3: 2017.4–2019.4 | 1201/451 | Dataset |
| Bettina Raessler et al. (2019) [47] | CT | YES | NR | YES | logistic regression | logistic regression | NR | Retrospective cohort | 120/23 (internal test) +61 (external test) | 2008–2017 | NO |
| Xiaojun Yang et al. (2019) [48] | CT | YES | NR | YES | CNN-F; multivariable logistic regression | CNN; logistic regression | YES | Retrospective cohort | 184/164 | 2016.1–2018.11 | NO |
| Yuan Gao et al. (2019) [49] | CT | YES | NR | YES | FR-CNN | CNN | NR | Cohort 1: retrospective cohort collected from Tongji Hospital; Cohort 2: retrospective cohort collected from Hubei Cancer Hospital (Hubei, China) | 32,495/6000 | 2011.1–2018.5 | No |
| David Coronado-Gutierrez et al. (2019) [50] | US | YES | NR | YES | CNN; VGG-M | VGG | NR | Retrospective cohort | NR/NR | 2015.4–2018.8 | NO |

(continued on next page)
| First author and year | Method for predictor measurement | Exclusion of poor-quality imaging | Heatmap provided | Extracted features | Algorithm architecture name | Algorithm architecture | Transfer learning applied | Source of data | Data range | Open access data |
|-----------------------|----------------------------------|----------------------------------|------------------|-------------------|-----------------------------|-----------------------|--------------------------|----------------|-----------|----------------|
| Yukinori Okada et al. (2019)[51] | BS | NR | NR | NO | NR | CNN | NR | Retrospective cohort | NR/NR | 2012.1~2014.11 | NO |
| Jeong Hoon Lee et al. (2019)[52] | CT (Axial) | NR | YES | NO | VGG16; VGG19; Inception; Inception V3; InceptionResNetV2; D3netNet121; DenseNet169; ResNet | CNN; VGG; Inception; Residual Network | NR | Retrospective cohort | 891/104 | 2017.7~2018.1 | NO |
| Jansen et al. (2019)[53] | Contrast-enhanced MRI, diffusion-weighted MRI | YES | NR | NON | Logistic regression; SVM; XGBoost | NR | NR | Retrospective cohort from University Medical Center Utrecht, The Netherlands | 55/17 | 2015.2~2018.2 | NO |
| Chuangming Li et al. (2019)[54] | Contrast-enhanced MRI | NR | NR | NR | AlexNet-like, UNET | CNN | NR | Retrospective cohort from the First Affiliated Hospital of Chongqing Medical University, China | 407454 | NR | NR |
| M. Dobopolski et al. (2019)[55] | PET, CT | YES | NR | YES | 3D deep multi-modality collaborative learning | CNN | NR | Public PET-CT dataset of STS patients | 326/164 | 2014.1~2016.12 | NO |
| Yuqi Peng et al. (2019)[56] | CT | YES | NR | YES | NR | CNN | NR | Retrospective cohort from the University of Texas Southwestern Medical Center | 378/165 | 2009 | NO |
| Zhou M D et al. (2019)[57] | Laparoscopy | NR | NR | NR | DNN | Deep neural network | NR | Retrospective cohort | 16040 | 2014.1.1~2017.9.30 | NO |
| Samir D. Mehta et al. (2019)[58] | Dual X-ray absorptiometry | NR | NR | NR | Radom forest algorithm; SVM | Radom forest algorithm; SVM | NR | Retrospective cohort | 35388 | 2007~2015 | NO |
| Yoshinari Arik et al. (2019)[60] | CT (Contrast-enhanced, axial) | NR | NR | NR | AlexNet | AlexNet | NR | Retrospective cohort | 314/259 | 2017.1~2018.11 | NO |
| Fanpeng Zhou et al. (2019)[61] | PET/CT | NR | NR | YES | Decision tree; discriminant analysis; SVM; KNN; BP-ANN | Decision tree; discriminant analysis; SVM; KNN; ANN | NR | Retrospective cohort from Department of Head and Neck Tumor, Tianjin Medical University Cancer Institute and Hospital, China | 314259 | NR | NO |
| Feng Hou et al. (2019)[62] | OCT | NR | NR | YES | AlexNet | CNN | NR | Retrospective cohort from Aichi-Gakuen University School of Dentistry, Nagoya, Japan | 562141 | 2017~2018 | NO |

(continued on next page)
| First author and year | Indicator definition | Algorithm data source | Number of images for training/testing | Data range | Open access data |
|------------------------|----------------------|----------------------|--------------------------------------|------------|-----------------|
| **Table 3 (Continued)** | | | | | |
| Xiaojuan Xu et al. (2019) [67] | Contrast-enhanced-MRI | Retrospective cohort from National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China | 140/80 | 2011.1–2017.12 | NO |
| Jiaxiu Luo et al. (2018) [68] | Diffusion-weighted MRI | Retrospective cohort | 122/50 | 2014.3–2016.6 | NO |
| Richard Ha et al. (2018) [69] | CT | Retrospective cohort | 522/131 | 2013–2017 | NO |
| B.H. Kann et al. (2018) [70] | US | Retrospective cohort | 612/200 | | |
| Yun Lu et al. (2018) [72] | MRI | Retrospective cohort | 28,080/36,000 | 2008.1–2015.11 | NO |
| José Ramony Ferreira Junior et al. (2018) [73] | CT | Retrospective cohort from Taipei Veterans General Hospital of Taiwan; Test set: Retrospective cohort from 6 Chinese Medical Centers | 52/16 | 2013.1–2016.9 | NO |
| Jin Li et al. (2018) [75] | MRI | Retrospective cohort from Harbin Medical University Cancer Hospital | NR | NR | |
| Mohamed Amine Larhmam et al. (2018) [76] | MRI | Retrospective cohort | 200/1197 | 2009.6–2014.9 | NO |
| Yan Zhong et al. (2018) [77] | CT | Retrospective cohort from Cancer Hospital Affiliated to Harbin Medical University | NR | NR | |
| Tuan D. Pham et al. (2017) [82] | CT | Retrospective cohort | NR | NR | |
| Mitsuru Koizumi et al. (2017) [79] | BS | Clinical data collected from the Peking University Third Hospital | 85,503/NR | 2006.1–2012.1 | NO |
| Juan Wang et al. (2017) [80] | MRI | Clinical data collected from the Peking University Cancer Hospital & Institute, Beijing, China | NR | NR | |
| Zhi-Long Wang et al. (2017) [81] | CT | retrospective cohort | NR/271 | 2010.4–2015.4 | NO |

(continued on next page)
| First author and year | Method for predictor measurement | Indicator definition | Exclusion of poor-quality imaging | Heatmap provided | Extracted features | Algorithm architecture name | Algorithm architecture | Transfer learning applied | Source of data | Number of images for training/testing | Data range | Open access data |
|-----------------------|---------------------------------|---------------------|-------------------------------|----------------|-----------------|--------------------------|---------------------------|----------------------|----------------|--------------------------------------|-----------|----------------|
| Qi Zhang et al. (2017) [83]* | Real-time elastography and B-mode ultrasound | Mayer et al. (2010) [84] | NR | NR | YES | SVM | SVM | NR | Retrospective cohort | NR/NR | 2013.11 – 2014.11 | NO |
| Yu-wen Wang et al. (2016) [85] | Heatmap provided | NR | NR | YES | Feed-forward back-propagation NN | AN | AN | NR | Retrospective cohort | Stage I: 331/332; Stage II: 410/205 | NR | NO |
| Ali Aslantas et al. (2016) [86] | BS | NR | NR | YES | ANN | BODENAVI | ANN | NR | Retrospective cohort | NR/130 | 2013 NO |
| Aneta Chmielewski et al. (2015) [87] | US | NR | NR | YES | SVM | SVM | NR | Retrospective cohort | 80/25 | NR | NO |
| Mitsuru Koizumi et al. (2015) [88] | BS | NR | NR | YES | BODENAVI | ANN | NR | Retrospective cohort | NR/NR | 2013.1–2013.12 | NO |
| Nersme Trabesli et al. (2015) [89] | CT | NR | NO | YES | Neural network | Neural network | NR | Retrospective cohort | 8/3 | NR | NO |
| Xuan Guo et al. (2015) [90] | 18F-FDG PET/CT | NR | NR | YES | RBF; SVM | SVM | NR | Retrospective cohort | 30/30 | 2009/6–2013.7 | NO |
| Osamu Tokuda et al. (2014) [91] | BS | NR | NR | NO | BODENAVI | ANN | NR | Retrospective cohort | NR3/248 | 2006.1–2011.5 | NO |
| Ani Seff et al. (2014) [92] | CT | NR | YES | YES | Random forest; SVM | Random forest; SVM | NR | Retrospective cohort | NR/NR | 2006/4–2008.9 | NO |
| Zhe-Guo Zhou et al. (2013) [93] | MDCT | YES | NR | YES | Conjugate gradient BP-ANN | Conjugate gradient BP-ANN | NR | Retrospective cohort | 37/53 | NR | NO |
| Seungwook Yang et al. (2013) [94] | Magnetic resonance black-blood imaging | NR | NR | YES | 3-D X-ray CT | 3-D X-ray CT | NR | Retrospective cohort | 6/44 | NR | NO |
| Jianfei Liu et al. (2013) [95] | Abdominal contrast-enhanced CT | NR | NR | NO | Joint framework | Joint framework | NR | Retrospective cohort | 37/53 | NR | NO |
| Yoshihiko Nakanuma et al. (2013) [96] | CT | NR | YES | YES | SVM | SVM | NR | Retrospective cohort | 6/44 | NR | NO |
| Chuan-Yu Chang et al. (2013) [97] | GSI-CT | NR | NR | YES | SVM | SVM | NR | Retrospective cohort | 88/89 | 2005–2007 | NO |
| Johannes Feulner et al. (2012) [98] | CT | NR | NO | YES | Spatial prior; AdaBoost | Spatial prior; AdaBoost | NR | Retrospective cohort | 280/1086 | NR | NO |
| Chao Li et al. (2012) [99] | GSI-CT | NR | NR | YES | SVM | SVM | NR | Retrospective cohort | NR/NR | 2010/4 | NO |
| Hongmin Gu et al. (2012) [100] | CT | NR | NO | YES | SVM | SVM | NR | Retrospective cohort | NR/NR | 2007/1–2008.11 | NO |
| Shao-Jer Chen et al. (2012) [101] | US | NR | NR | YES | SVM | SVM | NR | Retrospective cohort | NR/NR | 2007/1–2008.11 | NO |

(continued on next page)
| First author and year | Indicator definition | Algorithm architecture name | Algorithm architecture | Transfer learning applied | Source of data | Number of images for training/testing | Data range | Open access data |
|-----------------------|----------------------|----------------------------|------------------------|--------------------------|----------------|--------------------------------------|------------|----------------|
| Xiao-Peng Zhang et al. (2011) [102]* | Multi-detector row CT | LibSVM 2.89 | SVM | NR | Retrospective cohort | NR/NR | 2006.4~2008.9 | NO |
| Matthias Dietzel et al. (2010) [103] | Breast MRI | ANN | ANN | NR | Retrospective cohort | 123/71 | NR | NO |
| May Sadik et al. (2008) [104]* | BS | ANN | ANN | NR | Retrospective cohort | 810/59 | NR | NO |
| Junji Shirashi et al. (2008) [105] | Contrast-enhanced ultrasonography US | ANN | ANN | NR | Retrospective cohort | NR/NR | NR | NO |
| Junhua Zhang et al. (2008) [106]* | US from convex-type ANN | v-SVM | SVM | NR | Retrospective cohort | NR/NR | 2005.7~2006.6 | NO |
| Rie Tagaya et al. (2008) [107]* | BP-ANN | Retrospective cohort from Marianna University School of Medicine, Tokyo, Japan | 9/82 | 2005.4~2007.3 | NO |
| K. Marten et al. (2004) [108] | MSCT | Retrospective cohort from Klinikum rechts der Isar, Technical University Munich, Germany | NR/NR | NR | NR |

Abbreviation: NR=not reported. BS=bone scintigraphy. GC=gastric cancers. CT=computed tomography. MRI=magnetic resonance imaging. ANN=artificial neural network. SVM=support vector machine. NN=neural networks. CNN=convolutional neural networks. US=ultrasound. PSMA=Prostate specific-membrane antigen. 3D-T1-MPRAGE images=Three-dimensional T1 magnetization prepared rapid acquisition gradient echo. FR-CNN=fast region convolutional neural networks. CNN-F=CNN fast. PET=positron emission tomography. DNN=Deep neural network. MO=multi-objective model. KNN=k nearest neighbors. OCT=Optical coherence tomography. ANN=artificial neural network. BP-ANN=back-propagation artificial neural network. MSCT=multi-slice CT.

* 34 studies included in the meta-analysis.
ROC curves of these 34 studies (123 contingency tables) are shown in Fig. 3a, in which the pooled sensitivity was 82% (95% CI 79–84%) for all studies, and the pooled specificity was 84% (82–87%), with AUC of 0.90 (0.87–0.92). Many studies used more than one algorithm with several different accuracy for each algorithm. So, when selecting the contingency tables reporting the highest accuracy for different algorithms in these 34 studies with 48 tables, the pooled sensitivity was 87% (95% CI 84–89%), and the pooled specificity was 88% (84–92%), with AUC of 0.90 (0.87–0.92) (Fig. 3b).

Considering different algorithms were used in the included studies, we divided them into ML algorithms (ANN, KNN, SVM, RF, logistic regression and decision tree) and DL algorithms (CNN, DNN and DCNN) and did separate analysis for them, which showed a pooled sensitivity of 87% (95% CI 83–90%) for ML and 86% (82–89%) for DL, and a pooled specificity of 89% (82–93%) for ML and 87% (82–91%) for DL (Fig. 4).

30 studies included in the meta-analysis were validated by in-sample dataset with a pooled sensitivity of 86% (95% CI 83–89%) and a pooled specificity of 90% (85–93%). Only 4 studies used out-of-sample dataset to perform an external validation, for which sensitivity was 89% (84–93%) and specificity was 74% (69–79%) (Fig. 5).

Of these 34 studies, 8 compared performance between AI algorithms and health-care professionals using the same sample, with 10 contingency tables for AI algorithm and 16 tables for health-care professionals (Fig. 6). The pooled sensitivity was 89% (95% CI 83–93%) for AI algorithms and 72% (61–81%) for health-care professionals. The pooled specificity was 85% (79–89%) for AI algorithms and 72% (63–79%) for health-care professionals. Only 1 of the 8 studies was validated by out-of-sample dataset, and therefore a comparison between the performance of AI and health-care professionals by the identical external sample could not be performed.

All studies showed that the AI algorithms were beneficial for the diagnosis of tumor metastasis from medical radiology imaging when compared to the reference standard used in each study (OR 22.14 [95% CI 18.52–26.46] P<0.001, I²=79.6%) (Fig. 7), from which we can also see high heterogeneity among these studies. Visual inspection of funnel plots suggested there was no publication bias (P=0.19) (Supplementary figure 2).

To determine the source of heterogeneity, we did several subgroup analyses. In terms of metastasis types, there were DM whose pooled sensitivity was 88% (95% CI 80–93%), pooled specificity was 90% (76–96%), and AUC was 0.94 (0.92–0.97) (n = 15, I²=79.7%, P=0.001) and LNM whose sensitivity was 86% (95% CI 83–88%), specificity was 87% (84–90%), and AUC was 0.93 (0.90–0.95) (n = 33, I²=79.0%, P<0.001) (Fig. 8a). The outcomes were similar regarding the primary tumor types and medical imaging types. When it comes to the primary tumor types, in the breast cancer group, the sensitivity was 85% (95% CI 81–87%), the specificity was 82% (75–87%), and AUC was 0.86 (0.83–0.89) (n = 12, I²=46.4%, P=0.039). In the head and neck cancer group, the sensitivity was 87% (95% CI 81–91%), the specificity was 91% (87–94%), and AUC was 0.95 (0.92–0.96) (n = 10, I²=77.8%, P<0.001). Regarding the

Fig. 2. International research situation.
other primary tumor types, the sensitivity was 88% (95% CI 83–91%), the specificity was 89% (81–94%), and AUC was 0.94 (0.91–0.95) (n = 26, I²=84-2%, P<0.001) (Fig. 8b). As for medical imaging types, there were 16 contingency tables using CT (I²=85.7%, P<0.001), 12 tables using ultra sound (I²=0.0%, P = 0.505), 9 tables using bone scintigraphy (I²=62.3%, P = 0.007), 6 tables using MRI (I²=76.8%, P = 0.001) and 5 tables using other imaging types (I²=65-6%, P = 0.02) (Fig. 8c). Subgroup analysis for different AI algorithms contained ML (n = 32, I²=82.4%, P<0.001) and DL (n = 16, I²=70.8%, P<0.001). While in the studies were externally validated, heterogeneity was acceptable (n = 7, I²=45.1%, P = 0.091). We could not find a reasonable explanation for heterogeneity from subgroup analysis. We also did regression analysis to find the sources of heterogeneity. However, the results also could not make an explanation (regression analysis results are provided in Supplementary table).
4. Discussion

With great attention to the development of AI, more and more people are curious about its performance in medicine. In this systematic review and meta-analysis, we found that AI algorithms may be used for the diagnosis of tumor metastasis from medical radiology imaging material with equivalent or even better performance to health-care professionals, in terms of sensitivity and specificity. Tumor metastasis, as one of the main reasons for tumor-induced death, has a great impact on the treatment plan and prognosis judgment. Tumor metastasis sites may involve lymph nodes and distant organs, such as liver, lung and brain, which may be difficult to diagnose in clinical examination. Medical imaging is an important tool to diagnose tumor metastasis. However, the accurate diagnosis of tumor metastasis without misdiagnosis and missed diagnosis is a challenging task. The excellent performance of AI in image identification with rapid speed, high accuracy and significant manpower reduction excited the public. In 2019, Liu XX, et al. [30], conducted a systematic review and meta-analysis and found the diagnostic performance of deep learning models from medical imaging to be equivalent to that of health-care professionals in classifying diseases, with the sensitivity of 87.0% and specificity of 92.5%, which provided the basis for the
Fig. 7. Forest plot of studies included in the meta-analysis (34 studies).
clinical use of deep learning models. As for the diagnosis of tumor metastasis, there were no other meta-analyses focus on this subject to date, where we also reached a similar positive conclusion.

The first appearance of AI as a term can be dated back to a conference in 1956 [31]. As a branch of computer science, AI attempted to use computers to simulate the thought processes and intelligent behaviors of people, of which machine learning is an important part. The presence of ANN, SVM and other ML algorithms aroused people’s enthusiasm towards ML. It is not until 2006 that Geoffrey Hinton [13] proposed the concept of DL, which was the further development of ML. Twenty-three of the included studies in 2018 and beyond witnessed an increase in DL in contrast to that only 1 study before 2018 involved in DL. Taking into account the different development stages of AI, we did a separate analysis for studies using different algorithms, where no significant difference was observed. This may be attributed to the small dataset of included studies, most of which collected a few hundred data, limiting the advantages of DL.

In our research, we observed statistically significant heterogeneity among the included studies. So, we did several subgroup analyses and meta-regression for different algorithms, type of population, and medical imaging. The heterogeneity of studies validated by external sample was acceptable. 3 of the 4 studies with external validation based on the different version of the same computer assisted diagnosis system, which may contribute to the result. Generally, the results still cannot explain the source of heterogeneity, which may be contributed to the broad nature of the review (accepting any classification task using any imaging types for any metastasis types of any primary tumors).

Although the outcome of our research seems to bring light to the application of AI in detecting tumor metastasis from medical radiology imaging, several common methodological defects should be noted.

First, the design and practice of some included studies may make the research results out of clinical practice, among which the most common is the lack of comparison with health-care professionals in diagnostic accuracy. In the 69 included studies, only 8 studies made a comparison with health-care professionals. Assessing the performance of AI in insolation instead of comparing with the most common way in clinical practice (review the medical imaging by a radiologist) makes the outcomes unreliable when applied in the clinical setting. Even if some studies had the comparison, very few of them made it with humans using the same test dataset, resulting in a lack of comparability. Although we have reached the conclusion that AI models had the equivalent or even better diagnostic performance from medical imaging compared to health-care professionals, some factors still need to be considered. Only 8 studies using the same sample to compare health-care professionals and AI algorithms. Different studies recruited radiologists with different years of experience and different numbers. Some studies did not train radiologists in advance. All of above may influence the result. Furthermore, we included the studies that only used medical imaging to identify the presence of tumor metastasis, and excluded those that used other clinical materials, such as electronic medical record and clinical information of patients. It made our research topic more consistent. With the additional information available in the clinical practice, some prediction models can predict the possibility of metastasis based on the patient’s gender, age and history to assist diagnosis [32–36].

Second, there were no prospective studies. All included studies were retrospective studies, whose participants were selected from hospital medical records. Some studies used online open-access datasets instead of being done in the real clinical environment. And some studies provided poor description of missing data. In terms of the standard to diagnose metastasis, some studies only used the opinion of a single radiologist as a standard, which may not be convincing.

Third, various indicators of diagnostic performance were used in the studies. The value of TP, TN, FP and FN at a specified threshold should at least be provided, but most studies did not give a threshold or explain the reason for choosing this threshold. Most studies set the threshold at the value of 0-5, which is a convention in machine learning development [37,38]. Indicators like the sensitivity, specificity and accuracy were used in most studies. When the number of patients with/without metastasis in the test dataset was reported, sensitivity and specificity can be used to calculate TP, TN, FP and FN for contingency tables construction. Other indicators such as precision, dice ratio, F1 score and recall, which are common in the field of computer science, also appeared as the only measure in some studies. However, these indicators are not comprehensive, only with which we cannot get enough information to construct contingency tables.

Last but not least, in the 69 included studies there were only 4 with external validation, which means testing the model with out-of-sample dataset from one or more other centers. Most studies split the dataset from one center into training set and test set randomly or...
according to different time periods. The performance was evaluated by the test set, which should be called internal validation. Since the goal of validation is to investigate the performance within patients from different population, it is appropriate to collect a new dataset from different center. The absence of external validation made it hard to ensure the generalizability of the model, leading to overestimated results [39]. In our research, studies with external validation had an expectedly worse performance than internally validated studies. It is understandable that better performance can be achieved with the less heterogeneous samples. Strict external validation in the development of diagnostic model is urgently needed.

During the research, we also found some common deficiencies in AI studies. The most obvious point is that some key terminology is not uniformly named. Different studies have different definitions of the same terminology. For instance, for one AI model, the dataset is usually divided into several different parts, including the initial training set and one or more testing sets used to evaluate model effectiveness. While the term “validation” is used causally, some authors used this word to indicate the dataset used to test the diagnostic performance of the final model. Others defined it as a dataset with tuning function during the development process. The naming confusion makes it difficult to judge whether the test set is independent. The independent dataset, which is never learned by the model, is crucial to the credibility of the final model. So, canonical naming is urgently needed. Some scholars [30] have put forward suggestions. They distinguished the dataset used for a model as training set (for training the model), tuning set (for tuning the parameters of the model) and validation test set (for evaluating the performance of the final model), which is also accepted by our article. As for different types of validation test set, Altman and Royston’s suggestion [39] may be adopted. They named dataset for in-sample validation as internal validation, dataset for out-of-sample validation with a temporal split as temporal validation, and dataset for out of sample validation as external validation. Studies on the AI application in the medical field should strive to avoid problems mentioned above in the future.

Diagnosis of tumor metastasis using AI algorithms has great potential. From this meta-analysis, we conservatively draw a conclusion that the AI algorithms may be used for the diagnosis of tumor metastasis from medical radiology imaging with equivalent or even better performance to health-care professionals, in terms of sensitivity and specificity, providing a basis for its clinical application. Its widespread clinical application may alleviate the shortage of medical resources, improve the detection rate and accuracy of tumor metastasis and then the prognosis of patients. However, it should be acknowledged that more high-quality studies on the AI application in the medical field with adaption to the clinical practice and standardization of research routines are needed. In this review, we also put forward some existing problems of design and reporting that the algorithm developers should consider. High-quality studies are always the cornerstone of evaluation for diagnostic performance by various algorithms, which will finally benefit patients and the health care system.

Acknowledgement
This work was supported by College Students’ Innovative Entrepreneurial Training Plan Program.

Data sharing statement
The search strategy was shown in Appendix Section 1, and the contingency tables of 34 studies included in the meta-analysis were shown in Appendix Section 2. The results of risk of bias and publication bias were separately provided in the Supplementary Figure 1 and 2. Additional data are available on request.

Supplementary materials
Supplementary material associated with this article can be found in the online version at doi:10.1016/j.eclinm.2020.100669.

References

[1] Yamashta K, Hosoda K, Ema A, Watanabe M. Lymph node ratio as a novel and simple prognostic factor in advanced gastric cancer. Eur J Surg Oncol 2016;42(9):1253–60.
[2] O’Sullivan B, Brierley J, Byrd D, et al. The TNM classification of malignant tumours—towards common understanding and reasonable expectations. Lancet Oncol 2017;18(7):S849–51.
[3] Ruyenberg T, Verbst BM, Vonk-Van Ooeten J, Aastrineidou E, Sjogren EV, Webb AG. Improvements in high resolution laryngeal magnetic resonance imaging for preoperative transoral laser microsurgery and radiotherapy considerations in early lesions. Front Oncol 2018;8:216.
[4] Tsili AC, Alexiou G, Naka C, Argypoulou MI. Imaging of colorectal cancer liver metastases using contrast-enhanced US, multidetector CT, MRI, and FDG PET/CT: a meta-analysis. Acta Radiol 2020 284(15):120925481.
[5] Zhen L, Liu X, Yang C, et al. Accuracy of multiparametric magnetic resonance imaging for diagnosing prostate Cancer: a systematic review and meta-analysis. BMC Cancer 2019;19(1):1244.
[6] Ozturk M, Selcak MB, Polar AV, Ozbalci AB, Baris YS. The diagnostic value of ultrasound and shear wave elastography in the differentiation of benign and malignant soft tissue tumors. Skeletal Radiol 2020.
[7] Zhang L, Wang H, Li Q, Zhao M-H, Zhan Q-M. Big data and medical research in China. BMJ 2016;360:j5910.
[8] King BF. Artificial intelligence and radiology: what will the future hold? J Am Coll Radiol 2018;15(3 P 8):501–3.
[9] Schlemmer H-P, Bittencourt LR, D’Anastasi M, et al. Global challenges for cancer imaging. J Glob Oncol 2018;4.
[10] Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Cyber Cyber 1980;36(4):193–202.
[11] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273–97.
[12] The random subspace method for constructing decision forests -IEEE J Magazine [Internet]. [cited 2018 Jul 10]. Available from: https://ieeexplore.ieee.org/document/709601/.
[13] Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. Neural Comput 2006;18(7):1527–54.
[14] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006;313(5786):504–7.
[15] Krizhevsky A, Sutskever I, Hinton GJ. ImageNet classification with deep convolutional neural networks. J Digit Imaging 2017;20(1):
[16] Lijns G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis for interstitial lung diseases using a deep convolutional neural network. J Digit Imaging 2017;30(1).
[17] Ruytenberg T, Verbist BM, Vonk-Van Ooeten J, Astreinidou E, Sjogren EV, Webb AG. Improvements in high resolution laryngeal magnetic resonance imaging for preoperative transoral laser microsurgery and radiotherapy considerations in early lesions. Front Oncol 2018;8:216.
[18] Vinyals O, Toshev A, Bengio S, Erhan D. Show and tell: lessons learned from the 2015 MS-COCO image captioning challenge. IEEE Trans Pattern Anal Mach Intell 2017;39(4):652–63.
[19] Cao G, Liu F, Tan H, et al. Deep learning and its applications in biomedicine. Geno Proteomics Bioinform 2018;16(1):17–32.
[20] Anthimopoulos M, Christodoulides S, Ebner L, Christe A, Mougiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans Med Imaging 2016;35(5):1207–16.
[21] Xu Y, Hosny A, Zeleznik R, et al. Deep learning predicts lung cancer treatment response from serial medical imaging. Clin Cancer Res 2018;25(11):3266–75.
[22] Kooi T, Lijns G, van Ginneken B, et al. Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal 2017;35:303–12.
[23] Yang X, Kwitt R, Stener M, Niethammer M. Quicksilver: fast predictive image registration - a deep learning approach. Neuroimage 2017;158:378–96.

Contributors
YL and GL contributed to the conception and design of the study. QZ, LY, JL and BZ contributed to the literature search and data extraction. QZ and KG contributed to risk of bias evaluation. BZ contributed to data analysis and interpretation. QZ wrote the first draft of the report with input from LY. All authors contributed to critical revision of the manuscript. All authors approved the manuscript.

Declaration of Competing Interest
All authors declare no competing interests.
Koizumi M, Motegi K, Koyama M, et al. Diagnostic performance of a computer-assisted diagnosis system for bone scintigraphy of occult mediastinal lymph node metastasis of lung adenocarcinoma. 2019: 2019. p. 3658–88.

Feng Q-L, Liu C, Qi L, et al. An intelligent clinical decision support system for preoperative prediction of lymph node metastasis in gastric cancer. J Am Coll Radiol 2019;16(7):525–60.

Schnellfrohner T, Ware MP, Liu LP, Sarr MG, Birkett DH, Rutherz R. Can we accurately identify peritoneal metastases based on their appearance? An assessment of the current practice of intraoperative gastrointestinal cancer staging. Ann Surg Oncol 2019;26(6):1795–804.

Mehta SD, Sebro R. Random forest classifiers aid in the detection of incidental osteoblastic osseous metastases in DEKX studies. Int J Comput Assist Radiol Surg 2019;14(5):915–8.

Hou F, Yang Z, Gu W, Yu Y, Liang Y. OCT Automatic identification of metastatic lymph nodes in OCT images. In: Fujimoto JG, Izatt JA, editors. Proceedings of SPIE, 2019:12.

Arjì Y, Fukushima M, Kise Y, et al. Contrast-enhanced computed tomography image assessment of cervical lymph node metastasis in patients with oral cancer by using a deep learning system of artificial intelligence. Oral Surg Oral Med Oral Pathol Oral Radiol 2019;128(5):690–5.

Zhou YP, Li S, Zhang XX, et al. High dimensional MRI rectal lymph node aided diagnostic system based on deep neural network. Zhonghua Wai Ke Za Zhi 2019;57 (2):108–13.

Li Y, Ersemen A, Shangguan J, et al. Establishment of a new non-invasive imaging prediction model for liver metastasis in colon cancer. Am J Cancer Res 2019;9(11):2482–92.

Zhiguo Z, Dohopolski M, Liyuan C, et al. Reliable lymph node metastasis prediction from head & neck cancer through automated multi-objective model. 2019: 4.

Acar E, Lebelibci A, Ellidokuzu BE, Basinir Y, Kaya GC. Machine learning for differentiating metastatic and completely responded sclerotic bone lesion in prostate cancer: a retrospective radiomics study. Br J Radiol 2019;92(1096):

Hou F, Yang Z, Gu W, Yu Y, Liang Y. OCT Automatic identification of metastatic lymph nodes in OCT images. In: Fujimoto JG, Izatt JA, editors. Proceedings of SPIE, 2019.

Arjì YA-Oh, Sugita Y, Nagao T, et al. CT evaluation of extranodal extension of cervical lymph node metastases in patients with oral squamous cell carcinoma using deep learning classification. doi:10.1007/s11282-019-00391-4.

Yi X, Li H, Wang S, et al. Combination of MBI-based predictive model for preoperative assessment of lymph node metastasis in endometrial cancer.

Luo J, Ning Z, Zhang S, Feng Q, Zhang Y. Bag of deep features for preoperative prediction of sentinel lymph node metastasis in breast cancer. Phys Med Biol 2018;63(12):4574–86.

Kann BH, Aneja S, Loganadane GV, et al. Successful identification of head and neck cancer (HNC) nodal metastasis (NM) and extranodal extension (ENE) using deep learning neural networks. Int J Radiation Oncol Biol Phys 2018;102S(3):560–5.

Lee JH, Baek JH, Kim JH, et al. Deep learning-based computer-aided diagnosis system for localization and diagnosis of metastatic lymph nodes on ultrasound: a pilot study. THYROID 2018;28(10):1332–8.

Lu Y, Yu Q, Gao Y, et al. Identification of metastatic lymph node metastases in MR imaging with faster region-based convolutional neural networks. Cancer Res 2018;78(17):5334–43.

Ferreira Junior JR, Koenigkam-Santos M, Garcia Cipriano FE, Fabro AT, de Azevedo-Marques PM. Radiomics-based features for pattern recognition of lung cancer histopathology and metastases. Comput Methods Programs Biomed 2018;159:23–30.

Lo T-Y, Wei P-Y, Yen C-H, et al. Prediction of metastasis in head and neck cancer from computed tomography images. 2018: 2018–23.

Jin L, Peng W, Yanzhuo K, Yang Z, XiaoLong L, Kuan L. Transfer learning of pre-trained inception-v3 model for colorectal cancer lymph node metastasis classification. 2018: 1650–4.

Larhammar D, Mahmoudi S, Drisri S, Bengjelloul M. A Texture analysis approach for the cervical lymph node metastasis classification in T1 and T2 MRI. In: Rojas I, Ortuño F, editors; 2018: 198–211.

Zhong Y, Yuan M, Zhang T, Zhang YD, Li H, Yu TF. Radiomics approach to prediction of occult mediastinal lymph node metastasis of lung adenocarcinoma.

Wang H, Zhou Z, Li Y, et al. Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images. EJNMMI Res 2017;7(1):11.

Koizumi M, Motegi K, Koyama M, Terauchi T, Yusa A, Tonne J. Diagnostic performance of an image-assisted diagnostic system for bone scintigraphy of
newly developed skeletal metastasis in prostate cancer patients: search for low-sensitivity subgroups. Ann Nucl Med 2017;31(7):521–8.

[80] Wang J, Fang Z, Lang N, Yuan H, Su M-Y, Baldi P. A multi-resolution approach for spinal metastasis detection using deep Siamese neural networks. Comput. Biol. Med. 2017;84:137–66.

[81] Wang Z-L, Zhou Z-G, Chen Y, Li X-T, Sun Y-S. Support vector machine model of computed tomography for assessing lymph node metastasis in esophageal cancer with neoadjuvant chemotherapy. J Comput Assist Tomogr 2017;41(3):455–60.

[82] Pham TD, Watanabe Y, Higuchi M, Suzuki H. Texture analysis and synthesis of malignant and benign mediastinal lymph nodes in patients with lung cancer on computed tomography. Sci Rep 2017;7(43299).

[83] Zhang Q, Suo J, Chang W, Shi J, Chen M. Dual-modal computer-aided evaluation of axillary lymph node metastasis in breast cancer patients on both real-time elastography and B-mode ultrasound.

[84] Wang Y-W, Wu C-S, Zhang G-Y, et al. Can parameters other than minimal axial diameter in mri and pet/ct further improve diagnostic accuracy for equivocal retropharyngeal lymph nodes in nasopharyngeal carcinoma. PLoS ONE 2016;11(3):e0153741 (10).

[85] Atlantas A, Dandil E, Saglam S, Cakiroglu M. CADBOSS: a computer-aided diagnosis system for whole-body bone scintigraphy scans.

[86] Chmielewski A, Dufort P, Scaranelo AM. A computerized system to assess axillary lymph node malignancy from sonographic images. Ultrasound Med Biol 2015;41(10):2690–9.

[87] Koizumi M, Miyaji N, Murata T, et al. Evaluation of a revised version of computer-aided diagnosis system, BONENAVI version 2, for bone scintigraphy in cancer patients. Ann Nucl Med 2015;29(8):659–65.

[88] Koizumi M, Wagatssuma K, Miyaji N, et al. Evaluation of a computer-aided diagnosis system, BONENAVI version 2, for bone scintigraphy in cancer patients in a routine clinical setting. Ann Nucl Med 2015;29(2):138–48.

[89] Trabelsi N, Ben Sellem D. Comparison of supervised and unsupervised classifications in the detection of hepatic metastases, 2015. 

[90] Gao X, Chu C, Li Y, et al. The method and efficacy of support vector machine classifiers based on texture features and multi-resolution histogram from 18F-FDG PET-CT images for the evaluation of mediastinal lymph nodes in patients with lung cancer.

[91] Tokuda G, Harada Y, Ohishi Y, Matsuenga N, Edenbrandt L. Investigation of computer-aided diagnosis system for bone scans: a retrospective analysis in 406 patients. Ann Nucl Med 2014;28(4):329–39.

[92] Self A, Lu L, Cherry KM, et al. 2D view aggregation for lymph node detection using a shallow hierarchy of linear classifiers. In: Golland P, Hata N, Barillot C, Hornegger J, Howe R, editors. Lecture notes in computer science; 2014. p. 544–52.

[93] Zhou Z-G, Liu F, Jiao L C, et al. An evidential reasoning based model for diagnosis of lymph node metastasis in gastric cancer. BMC Med Inform Decis Mak 2013;13(123).

[94] Yang S, Nam Y, Kim M-O, Kim EY, Park J, Kim D H. Computer-aided detection of metastatic brain tumors using magnetic resonance black-blood imaging. Invest Radiol 2013;48(2):113–9.

[95] Liu J, Wang S, Lingurarug M, Yao J, Summers RM. A Variational framework for joint detection and segmentation of ovarian cancer metastases. In: Sakuma I, Barillot C, Navab N, editors. Lecture notes in computer science; 2013. p. 83–90.

[96] Nakamura Y, Nimura Y, Kitasaka T, et al. Automatic abdominal lymph node detection method based on local intensity structure analysis from 3D X-ray CT images. In: Novak CI, Ayliward S, editors. Proceedings of SPIE; 2013.

[97] Chang C-Y, Lai C-C, Lai C-T, Chen S-J. Integrating PSONN and Boltzmann function for feature selection and classification of lymph nodes in ultrasound images. J Vis Commun Image Represent 2013;24(1):23–30.

[98] Feuiller J, Zhou Sk Fau Hammon M, Hammon M, Fau Hornerjeg J, Hornerjeg J, Fau Comanicu D, Comanicu D. Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior.

[99] Li C, Zhang S, Zhang H, et al. Using the K-nearest neighbor algorithm for the classification of lymph node metastasis in gastric cancer. Comput Math Methods Med 2012;2012(876545).

[100] Cai H, Cui C Fau Tian H, Tian H Fau Zhang M, Zhang M Fau Li L, Li L. A novel approach to segment and classify regional lymph nodes on computed tomography images.

[101] Chen SJ, Lin Ch Fau Chang C-Y, Fau Fau Chang K-Y, et al. Characterizing the major sonographic textural difference between metastatic and common benign lymph nodes using support vector machine with histopathologic correlation.

[102] Zhang X-F, Wang Z-L, Yang L, Sun Y-S, Cao K, Gao Y. Support vector machine model for diagnosis of lymph node metastasis in gastric cancer with multidetector computed tomography: a preliminary study. BMC Cancer 2011;11(10).

[103] Dietzel M, Baltzer PAT, Dietzel A, et al. Application of artificial neural networks for the prediction of lymph node metastases to the ipsilateral axilla - initial experience in 194 patients using magnetic resonance mammography. Acta radiol 2010;51(8):851–8.

[104] Sadik M, Hamadeh I, Nordblom P, et al. Computer-assisted interpretation of planar whole-body bone scans. J Nuclear Med 2008;49(12):1958–65.

[105] Shiraishi J, Sugimoto K, Moriyasu F, Kamiyama N, Doi K. Computer-aided diagnosis for the classification of focal liver lesions by use of contrast-enhanced ultrasonography. Med Phys 2008;35(5):1734–46.

[106] Zhang J, Wang Y, Dong Y, Wang Y. Computer-aided diagnosis of cervical lymph nodes on ultrasonography. Med Phys 2008;35(5):43.

[107] Tagaya R, Kurimoto N, Osada H, Kobayashi A. Automatic objective diagnosis of lymph nodal disease by B-mode images from convex-type echobronchoscopy. Chest 2008;133(1):137–42.

[108] Marten K, Grillhosl A, Seyfarth T, Seyfarth T, Fau Obenauer S, Obenauer S, Fau Rummeny EJ, Rummeny EJ Fau Fau Engelke C, Engelke C. Computer-assisted detection of pulmonary nodules: evaluation of diagnostic performance using an expert knowledge-based detection system with variable reconstruction slice thickness settings.