Research article

Tumor burden of lung metastases at initial staging in breast cancer patients detected by artificial intelligence as a prognostic tool for precision medicine

Madison R. Kocher [a], Jordan Chamberlin [a], Jeffrey Waltz [a], Madalyn Snoddy [a], Natalie Stringer [a], Joseph Stephenson [a], Jacob Kahn [a], Megan Mercer [a], Dhiraj Baruah [a], Gilberto Aquino [a], Ismail Kabakus [a], Philipp Hoelzer [b], Pooyan Sahbaee [b], U. Joseph Schoepf [a], Jeremy R. Burt [a,*]

[a] Medical University of South Carolina, Department of Radiology, 96 Jonathan Lucas Street Suite 210, MSC 323 Charleston, SC 29425, USA
[b] Siemens Healthineers, Princeton, NJ, USA

HIGHLIGHTS

- Automated detection software can quantify lung metastases on initial staging chest CT in breast cancer patients.
- AI-detected lung metastases number and max diameter on CT at initial cancer staging were strong predictors of mortality.
- AI detection and segmentation tool contributes to accurate individualized prognostication in breast cancer patients.

ARTICLE INFO

ABSTRACT

Background: Determination of the total number and size of all pulmonary metastases on chest CT is time-consuming and as such has been understudied as an independent metric for disease assessment. A novel artificial intelligence (AI) model may allow for automated detection, size determination, and quantification of the number of pulmonary metastases on chest CT.
Objective: To investigate the utility of a novel AI program applied to initial staging chest CT in breast cancer patients in risk assessment of mortality and survival.
Methods: Retrospective imaging data from a cohort of 226 subjects with breast cancer was assessed by the novel AI program and the results validated by blinded readers. Mean clinical follow-up was 2.5 years for outcomes including cancer-related death and development of extrapulmonary metastatic disease. AI measurements including total number of pulmonary metastases and maximum nodule size were assessed by Cox-proportional hazard modeling and adjusted survival.
Results: 752 lung nodules were identified by the AI program, 689 of which were identified in 168 subjects having confirmed lung metastases (Lmet+) and 63 were identified in 58 subjects without confirmed lung metastases (Lmet-). When compared to the reader assessment, AI had a per-patient sensitivity, specificity, PPV and NPV of 0.952, 0.639, 0.878, and 0.830. Mortality in the Lmet + group was four times greater compared to the Lmet-group (p = 0.002). In a multivariate analysis, total lung nodule count by AI had a high correlation with overall mortality (OR 1.11 (range 1.07–1.15), p < 0.001) with an AUC of 0.811 (R2 = 0.226, p < 0.0001). When total lung nodule count and maximum nodule diameter were combined there was an AUC of 0.826 (R2 = 0.243, p < 0.001).
Conclusion: Automated AI-based detection of lung metastases in breast cancer patients at initial staging chest CT performed well at identifying pulmonary metastases and demonstrated strong correlation between the total number and maximum size of lung metastases with future mortality.
Clinical impact: As a component of precision medicine, AI-based measurements at the time of initial staging may improve prediction of which breast cancer patients will have negative future outcomes.

---

* Corresponding author.
E-mail address: burtje@musc.edu (J.R. Burt).

1. Introduction

Breast cancer remains the most common cancer in females as recent estimates report over 3.8 million women living with a history of invasive breast cancer [1]. In the staging of a newly diagnosed breast cancer patient, prognostic factors such as cancer receptor type and distant organ metastases are important in treatment decisions and survival [2]. The specific organ site of any metastases is also important, with lung-only involvement shown to have a 32% mortality within the first year [3]. Computed tomography (CT) is a primary tool for the initial staging of pulmonary metastases; however, there is interobserver variability in the detection of lung metastases with overall low sensitivities [4, 5, 6, 7]. As documentation of the exact number and size of each pulmonary metastases can be an arduous, time-consuming task, there have been limited attempts to apply these metrics to clinical outcomes [8]. Although the incorporation of artificial intelligence (AI) into chest CT examinations has substantially improved the rate of detection of lung nodules, it still lacks accuracy in overall nodule classification and morphology characterization [9, 10, 11, 12]. By developing AI algorithms to detect lung nodules and to quantify overall nodule volume, valuable time could be saved and a more complete assessment of the overall volume of disease could be elucidated aiding in characterization.

Given the prevalence of breast cancer and the varying treatment options based on clinical, imaging, and cytologic markers, more information is needed to help identify trends and prognostic indicators in this population. Prior studies have suggested that imaging biomarkers (radiomics) may play a greater role in disease characterization and prognostication although many metrics are untapped due to lack of reliable AI algorithms [10, 13, 14, 15]. This study will assess a novel AI program for automated detection of the total number of lung metastases and maximum size in breast cancer patients at initial staging CT and correlate this with future mortality.

2. Methods

This retrospective cohort study was approved by the local Institutional Review Board with a waiver of informed consent.

2.1. Subjects

Clinical and imaging data were obtained from the single institution electronic medical record (EPIC, Madison Wisconsin) between January 2014 and January 2019 from a population-based cohort of subjects diagnosed as having breast cancer with (study) and without (control) lung nodules on CT. Exclusion criteria included lack of available initial staging chest CT, age less than 18-years-old, or pregnancy. Demographic information including age, sex, ethnicity, and comorbidities such as hypertension, diabetes, lung disease, and smoking status were recorded. Other information recorded included breast cancer type, breast biopsy results, lymph node biopsy results, hormone receptor status and American Joint Committee on Cancer (AJCC) TNM staging at initial diagnosis. Data from the clinical imaging reports at initial staging (CT, MRI, PET/CT, bone scan) were recorded including presence and location of metastatic disease, and lung biopsy results, if performed, were collected. The specific oncologic treatments were not recorded. Finally, patient outcomes including cancer-related death and development of extrapulmonary metastatic disease were noted.

2.2. Image acquisition

All CT examinations were performed on one of two clinical scanner types (SOMATOM Flash or Force; Siemens, Forchheim, Germany) from the lung apices through the bases, with or without intravenous contrast, and during breath-hold at end-inspiration. Acquisition parameters included 100-110kVp tube-voltage, CareDose mA, 192 × 0.6mm collimation, gantry-rotation time of 0.5 s, pitch of 0.7, and effective slice thickness of 0.5 mm. Images were reconstructed with a sharp body kernel to achieve a lung-window setting and reformatted to an axial slice thickness of 1 mm. Axial thick slice (10mm) lung-window maximum intensity projection (MIP) images were reconstructed.

2.3. Image interpretation and reference standard

All CT studies were assessed by one of four experienced, blinded radiologists reading independently, with one quarter of the subjects randomly assigned to each interpreter. The radiologists were tasked with measuring up to 5 identified lung metastases ≥4 mm per patient, quantifying the total number of metastasis per patient, and recording the location and greatest axial 2D-diameter of the five largest metastasis per patient.

2.4. AI algorithm

A deep convolutional neural network software prototype (AI-RAD Companion Chest CT VA10A, Siemens Healthineers, Malvern PA) provided by Siemens Healthineers was used to detect lung nodules on chest CTs performed for initial staging and follow-up. This algorithm is a software-platform that provides automatic AI-based multi-organ image analysis, visualization, and quantification and has been previously tested and validated on chest CT scans from multiple centers across the USA, Europe, and Asia [16, 17, 18, 19]. No pre-processing of the CT images was required prior to algorithm application for nodule detection. The AI algorithm reported the lobar location and axial 2D-diameter of the five largest AI-detected nodules for each subject and collected up to 30 total nodules. A pre-data analysis of numbers of nodules associated with simple mortality was conducted. There was no statistically significant difference in survival curves in any tested AI nodule number threshold, thus the AI reported number of nodules was capped at 30 for ease of reporting. Only the top 5 largest nodules were evaluated for inter-observer agreement.

2.5. AI validation

AI results were reviewed and validated based on ground-truth observation. Validation was performed on a per-patient and per-nodule basis. The following definitions were used on a per-nodule basis:

- True positive (TP): Both AI and reader identified the same nodule
- True negative (TN): Both AI and reader did not identify the nodule
- False positive (FP): AI identified a nodule but was determined by the reader to not be a nodule
- False negative (FN): Nodule was identified by the reader but not by AI

2.6. Definition of pulmonary metastasis

Based on criteria determined from a large meta-analysis, subjects were labelled as "lung metastases" if they had two or more noncalcified solid nodules ≥6 mm or at least one noncalcified solid nodule >10 mm on initial staging chest CT [20].

2.7. Statistical analyses

A power analysis was performed (Figure 1). All univariate calculations were performed in XLSTAT, and data visualization was completed in R version 3.6.3. Univariate statistics were calculated using appropriate parametric and nonparametric tests. Medians and interquartile ranges were reported for variables with non-normal distributions and all survival functions. Unweighted Cohen's kappa and diagnostic parameters were calculated on contingency tables with 95% confidence intervals using the efficient-score method. Correlation statistics were calculated using spearman's method and reported with spearman's rho and Intraclass-correlation coefficients for agreement between observers.
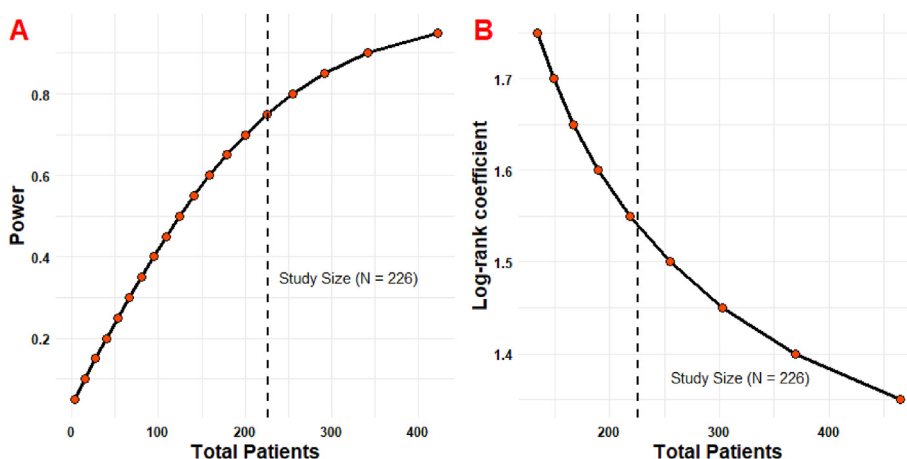
Figure 1. Post-hoc power analysis for log-rank testing of secondary outcomes (Survival by nodule characteristics). A. Power as a function of the total patients. A post-hoc power of 0.75 was achieved for secondary outcomes at a sample size of 226. B. Log-rank coefficient limit of detection assuming a power of 0.8. A sample size of 226 patients gives the lower limit of detection to be a log-rank coefficient of 1.54.

Kaplan-Meier survival calculations and cox-proportional hazard modelling were performed in R with the Survminer package. Tests of significance were achieved using the log-rank test. Cox-proportional hazard modeling and adjusted survival was determined based on AI measurements including total number of pulmonary metastases and maximum nodule size. Survival prediction classification by hormone positivity status was also performed.

3. Results

A total of 226 breast cancer subjects with lung nodules were included in the analysis after 14 subjects were excluded for failed AI processing. Another 24 were also excluded due to lack of adequate follow-up (n = 17) or lack of breast cancer histopathological diagnosis in the electronic medical record (n = 7). The study group consisted of 168 confirmed subjects with pulmonary metastases (total of 689 metastases) and 58 without pulmonary metastases. A power analysis to detect univariate mortality with a type 1 error of 0.05 and type II error of 0.20 revealed 659 nodules were needed, indicating the study was appropriately powered. The median clinical follow-up was 28 months (range = 3–58 months). Subjects ranging from 25-87 years old at initial diagnosis were included in the study (See Table 1 for full demographics).

No statistically significant difference was found regarding size of the primary breast cancer (p = 0.466) or hormone status (HER2+ p = 0.570, ER + p = 0.702, PR + p = 0.213) between the Lmet+ and Lmet-groups. Subjects with lung metastases were more likely to have a pathological diagnosis of infiltrating ductal carcinoma (IDC; p = 0.019) and be older (median age 62 vs 53, p < 0.001). The populations were similar concerning the presence of extrapulmonary metastases (including liver, bone, and brain metastases) except for the extent of axillary lymph node metastases, which was higher in the population without lung metastases p = 0.049.

3.1. AI performance

The AI model had a per-patient sensitivity, specificity, PPV, and NPV of 0.952, 0.639, 0.878, and 0.830, respectively, with a Cohen's Kappa of 0.637 and an intraclass correlation for nodule size of 0.76 (Table 2 and Figure 2A). On a nodule-to-nodule basis, correlation between the true positive, AI-measured nodule maximal diameter and expert nodule diameter demonstrated a Spearman's rho = 0.79 (Figure 2A) and a mean nodule size differences of -1.44 mm (95% CI for any size difference -20.38–17.50) (Figure 2B). Of note, nodule size discrepancy increased after a size cutoff of 20 mm. Nodule-level concordance for detection of lung lesions with a true negative lesion defined as a control patient where

Table 1. Patient demographics and pathologic diagnoses.

N = 226	Lung Metastases (N = 168)		No Lung Metastases (N = 56)		P (α = 0.05)
	Median	IQR	Median	IQR	
Age	62	53.5–67	53	43–63	**<0.001**
BMI	29.8	24.0–33.3	29.3	24.3–34.5	0.779
BSA	1.87	1.70–1.99	1.83	1.69–2.14	0.630
Initial Breast Lesion Size	2.5	1.75–4.75	2.8	2.1–5.1	0.466
	Count	Frequency	Count	Frequency	
Sex (Female)	162	96.4	54	93.1	0.244
Race					
White	118	70.2	35	60.3	0.133
Black	45	26.8	21	36.2	
Hispanic	1	0.6	2	3.4	
Other	4	2.4	0	0	
Hypertension	88	54.0	22	38.6	**0.064**
Hyperlipidemia	35	21.2	19	32.8	**0.002**
Diabetes	31	19.5	12	20.6	1.000
Current Smoker	7	4.2	4	6.9	0.486
COPD	8	4.9	4	6.9	0.515
ILD	3	1.9	1	1.7	1.000
Hx Lung Cancer	2	1.2	0	0	1.000
IDC	131	90.4	43	76.8	**0.019**
ER+	127	81.4	43	75.4	0.702
PR+	100	67.6	36	63.2	0.213
HER2+	40	28.4	12	22.2	0.570

Clinical variables, demographics, pathologic variables, and imaging characteristics associated with mortality. α = 0.05, Bonferroni correction not applied. Continuous variables assessed with Mann-Whitney U Test and categorical variables with Fisher's Exact Test. (BMI = body mass index, BSA = body surface area, COPD = chronic obstructive pulmonary disease, ILD = interstitial lung disease, IDC = invasive ductal carcinoma, ER+ = estrogen receptor positive, PR+ = progesterone receptor positive, HER2+ = human epidermal growth factor receptor 2 positive). Bolded values indicate statistical significance.

both the AI and expert determined there was no lesion had a sensitivity of 0.975 and specificity of 0.612 with a Kappa of 0.626 (Table 3).

3.2. Analysis of false positives

Analysis of the patients with false positives revealed a bias towards increased false positives in patients with lower BMI (26.9 vs 29.7, p =

Table 2. Patient-level concordance for detection of lung nodules.

N = 227	Expert – Nodule	Expert – No Nodule
AI - Nodule	158	22
AI - No nodule	8	39
	Value	**95% CI**
Sensitivity	0.952	0.904–0.977
Specificity	0.639	0.506–0.755
PPV	0.878	0.819–0.920
NPV	0.830	0.687–0.919
Cohen's Kappa	0.637	0.516–0.758

If a patient had any falsely positive or falsely negative nodules as determined by expert review, the patient was classified as a false positive or false negative, respectively. Unweighted Cohen's kappa = 0.637 (0.516–0.758) 95% confidence intervals for testing parameters calculated using the efficient-score method. (PPV = positive predictive value, NPV = negative predictive value).

0.015). Table 4 contains the cohort characteristic analysis of the false positives. Table 5 is a frequency table describing the sources of the false positive measurements. The most common causes of false positive nodules included vessels, atelectasis, and osteophytes (20.8%, 16.7%, 16.7%, respectively). The most common locations for false positive nodules were the right upper lobe, right middle lobe, and left upper lobe (29.2%, 20.8%, 20.8%, respectively). Generally, false positives were more common in the upper segments and decreased towards the lower segments. Figure 3 contains a few images of the false positive results.

3.3. Outcomes

Compared with the Lmet-group, Lmet + group was more than twice as likely to develop extrapulmonary metastases, including bone (p < 0.001) and brain (p < 0.001). There was a negative correlation between survival from initial imaging and increasing total number of nodules as detected by AI (R = -0.32, p < 0.00004)).

The Lmet + group had a four times greater mortality in the follow-up period (p = 0.002) as compared to the Lmet-group (Figure 4).

Presence of AI-detected lung metastases was negatively associated with survival in the setting of ER+, PR + disease (p = 0.00059) (Figures 5A-D). The presence of AI-detected lung metastases was not associated with a difference in survival in subjects with triple negative breast cancer (TNBC), triple positive breast cancer (TPBC), and HER2+ disease (P = 0.55, 0.18, 0.079, respectively). Presence of AI-detected lung metastases was negatively associated with survival in subjects with non-TNBC, non-TPBC, and non-HER2+ Breast cancer (p = 0.00011, 0.01, 0.015, respectively).

Figures 6A–D demonstrate the Cox-proportional hazard modelling and adjusted survival from imaging based on quantitative AI measurements including total metastases and maximum metastasis size. Subjects

were divided into groups based on number of AI lung metastases detected at initial staging (0–10, 10–20, and 20–30 nodules) overall demonstrating decreased survival time based on number of metastases. There was a disproportionate decrease in survival rates as number of metastases increased. All groups demonstrated a plateau in survival rate at approximately 18 months (Figure 6A). All groups demonstrated a proportional decrease in survival rate based on maximal metastasis size (Figure 6B). Maximum AI diameter (mm) of metastases was independently associated with an increased risk of death. For every 1 mm increase in maximum lung metastasis diameter there was a 3% increase in mortality. For every added lung metastasis identified by AI, there was a 7% increase in total mortality (Figure 6C). These findings coincided with results for expert numbered and measured lung metastases (Figure 6D).

Table 3. Nodule-level concordance for detection of lung lesions.

N = 752	Expert - Nodule	Expert - No Nodule
AI - Nodule	672	24
AI - No nodule	17	39
	Value	**95% CI**
Sensitivity	0.975	0.960–0.985
Specificity	0.612	0.488–0.736
PPV	0.966	0.948–0.977
NPV	0.696	0.557–0.808
Kappa	0.626	0.521–0.737

Unweighted Cohen's Kappa = 0.626 (95% CI 0.516–0.758). 95% confidence intervals calculated using the efficient-score method.

Table 4. Cohort characteristic analysis of patients with false positives.

	False Positive Patient (N = 22)		All other patients (N = 204)		P
Age	60	12.7	58.5	12.3	0.593
BMI	26.9	4.7	29.7	6.9	**0.015**
BSA	1.8	0.2	1.9	0.3	0.136
Pack Years	2.7	6	3.2	9.5	0.783
Female Sex	21	95.5	195	95.6	1
Hypertension	6	28.6	104	52.3	0.066
Hyperlipidemia	7	31.8	47	23.4	0.560
Diabetes	2	9.5	42	21.0	0.334
COPD	2	9.5	10	5.0	0.720
Current Smoker	2	9.1	9	4.5	0.667
Lung Cancer	0	0	2	1.0	1

Patients with false positives nodules were more likely to have a lower BMI than all other patients. Bolded value indicates statistical significance.
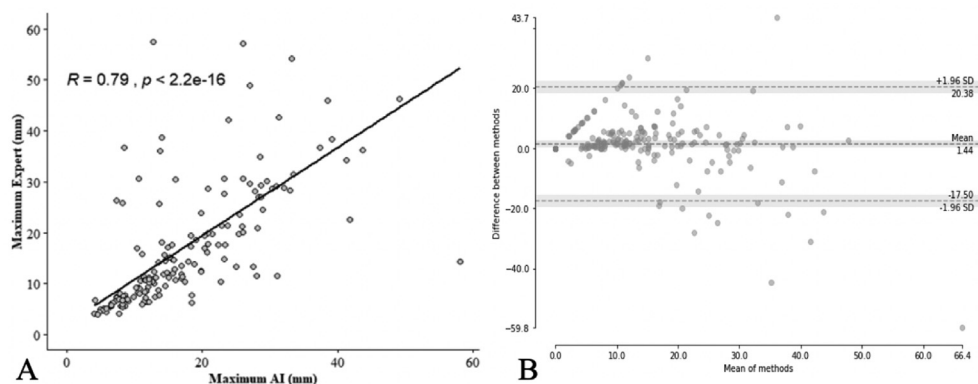
Figure 2. (A) Correlation of true positive maximum AI and expert metastasis measurements. FP were removed from analysis to determine accurate lung metastasis size concordance. There is a strong correlation between both methods (Spearman's rho = 0.79). 2-way average fixed raters ICC = 0.76 (p = 3.0e-18, 95% CI = 0.69–0.82). (B) Bland-Altman plot for quantitative comparison of difference between all AI and expert maximum metastasis size. Mean metastasis size difference was 1.44mm (95% CI for any size difference 20.38–17.50). Metastasis size discrepancy notably increases with sizes of greater than 20 mm as evidenced by bland-altman dispersion.

Table 5. Frequency table of the false positives.

False Positive nodules (N = 24)		
Identity	N (%)	Mean size (mm) (SD)
Osteophyte	4 (16.7)	13.0 (6.1)
Bowel	3 (12.5)	22.0 (9.6)
Vessel	5 (20.8)	14.0 (4.5)
Fluid in Fissure	1 (4.2)	23.7
Azygos Vein	3 (12.5)	24.1 (5.8)
Atelectasis	4 (16.7)	12.3 (6.6)
Diaphragm	1 (4.2)	16.1
Scar	1 (4.2)	20.1
Other	2 (8.4)	11.0 (4.0)
Location	**N (%)**	**Mean size (mm) (SD)**
Right upper lobe	7 (29.2)	17.5 (8.3)
Right middle lobe	5 (20.8)	17.0 (9.7)
Right lower lobe	4 (16.7)	15.6 (4.5)
Left upper lobe	5 (20.8)	14.9 (6.7)
Left lower lobe	3 (12.5)	15.6 (6.5)

The most common identities of false positive nodules were vessels (20.8%), osteophytes (16.7%), and atelectasis (16.7%). The right upper lobe had the highest predominance of false positive nodules (29.2%). Generally, the lowest frequency of false positives occurred in the lower lobes bilaterally.

Using classification and regression trees (CART), Figure 7 delineates deceased subjects via initial staging (clinical, pathologic and imaging), hormone receptor status, and AI-determination of the number/size of metastasis (es). Only significant (p < 0.05) CART nodes were included in the graphical depiction. The most important classifier was total number of lung metastases. Risk-groups were found to be < 7 metastases (11% mortality), 7–13 metastases (31% mortality), and >13 metastases (82% mortality). A best fit logistic regression model for prediction of mortality in this cohort is shown in Figure 8A. Variables included in the final model included axillary LN+, other LN+, liver metastases, AI determined total number of metastases and max lung metastasis diameter (mm) (AUC 0.857, McFadden R^2 0.323). Figure 8B describes the logistic regression model using only AI generated parameters for prediction of mortality (total metastases and max diameter (mm)) (AUC = 0.826, McFadden R^2 = 0.243), and Table 6 explains the model parameters in detail.

4. Discussion

The utility of biomarkers for detecting future outcomes at initial staging cannot be understated. AI detection of maximum axial diameter and number of lung metastases in the context of breast cancer at initial
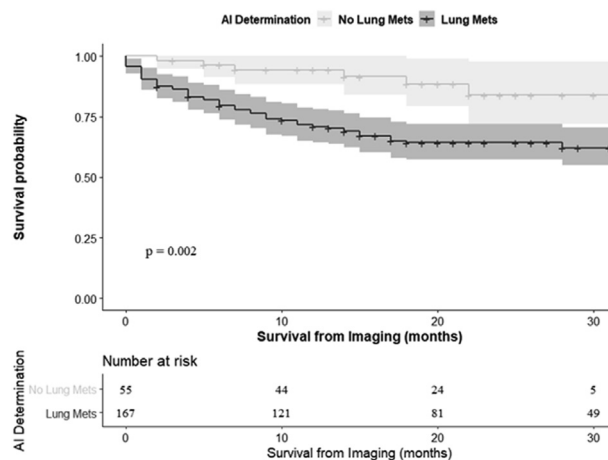

Figure 4. Survival from imaging between subjects with and without lung metastases as determined by AI. Significant survival difference detected using the log-rank test (p = 0.002). Follow-up period defined as 2.5 years. This figure reflects cancer-related mortality.

staging CT is a strong predictor for future mortality. The use of AI to detect and quantify pulmonary metastases in breast cancer patients has the potential to improve characterization efficiently and accurately. The AI model facilitates a rapid assessment, especially in direct comparison to the manual detection and measurement of up to 30 lung nodules. On a per-patient level, the AI model performed with a sensitivity of 0.952 and a specificity of 0.639. This is similar to multiple prior studies, for example Cui et al reported a sensitivity of 0.934 with a third-party database [21], Jin et al reported a sensitivity of 0.912 [22], and Gupta et al reported a sensitivity of 0.856 [23]. Armato III et al reported that mean expert thoracic radiologist nodule-detection sensitivities range from 0.51-0.83% and mean FP rates range from 0.33-1.39 per case [4]. While not highly specific, the utility of the algorithm lies in overall detection of all pulmonary nodules present. Additionally, our results are concordant with prior studies that lung nodule size is an independent predictor of survival [24, 25].

After analysis of the false positive nodules, it was clear that the low specificity comes with the low number of patients with no nodules. The pre-test probability in this cohort was high (78% of patients with nodules) and it would be more accurate to describe a false positive rate of 9.7% per patient. Most false positives were identified as atelectasis, vessels, and osteophytes. Atelectasis and infection were commonly misidentified as nodules, likely because of their relative mass-like and hyperdense appearance with adjacent normal lung parenchyma. Additionally, the lobular contour of the protruding osteophytes from the
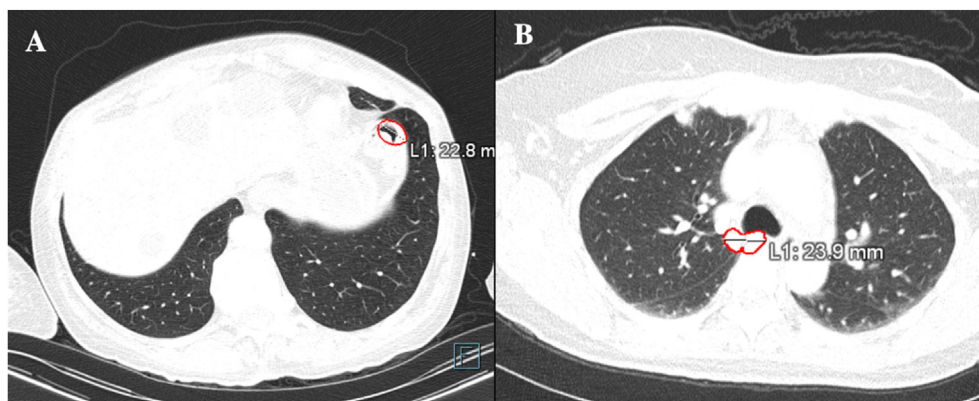

Figure 3. (A) CT with AI-RAD measurements after AI processing demonstrating a false positive result where the algorithm detected a part of the colon and measured it as a 2.3 cm nodule. (B) Additional CT with AI-RAD measurements that measured an osteophyte extending from the thoracic vertebral body.
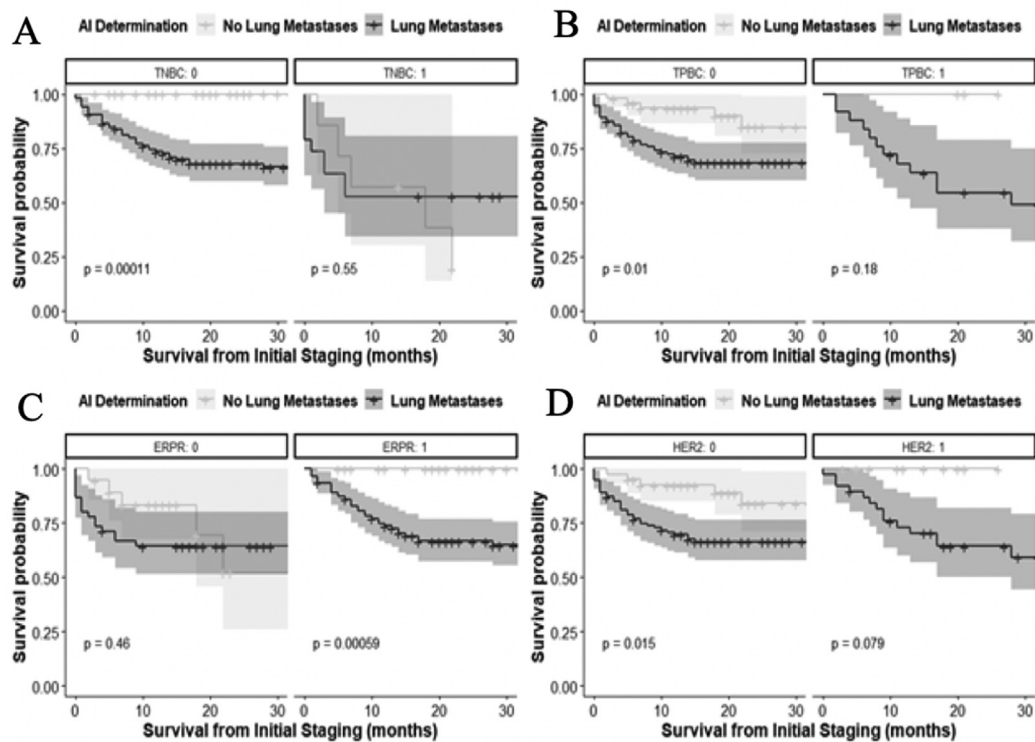
Figure 5. Survival from initial staging in months as classified by presence of AI-detected lung metastases and clinically relevant tumor isotypes. (A) Survival curves adjusted for the presence or absence of lung metastases and triple-negative breast cancer status (TNBC; HER2-, ER-, PR-). (B) Survival curves adjusted for the presence or absence of lung metastases and triple-positive breast cancer status (TPBC; HER2+, ER+, PR+). (C) Survival curves adjusted for the presence or absence of lung metastases and ER+, PR+ (ERPR) status. (D) Survival curves adjusted for the presence or absence of lung metastases and HER2+ status. Presence of a histological subtype indicated by "1", and absence defined by "0". 95% confidence intervals given by the shaded area. P-Values calculated using the log-rank test ($\alpha = 0.05$). (TNBC = triple negative breast cancer, TPBC = triple positive breast cancer, ERPR = estrogen receptor positive receptor).

thoracic vertebral bodies in direct contact with the lung parenchyma likely led to their misidentification as nodules.

In the workup of breast cancer, desirable biomarkers to determine effective treatments are relatively inexpensive, accessible, highly reproducible, and cause no harm to the patient. Initial staging CT examinations also fit these criteria, as they are a pre-existing step in the cancer work-up and their findings can be easily reproduced – especially with the application of AI and a standardized protocol. Radiomics allows for the combination of patient data and clinical features in addition to extracted imaging features to extrapolate prognostic outcomes and predict response to treatments [15]. Prior studies have shown the utility of including radiomic data when assessing response to treatment [13]. Here, the AI algorithm performed at a high level and was comparable to expert radiologist performance.

This study showed that the total number and size of lung metastases at initial staging CT is an effective and readily available biomarker that is not commonly employed in the standard of care, but perhaps could be with the efficiency that AI offers. In our study, number and size of lung metastases was a stronger predictor of mortality than breast mass size with the Lmet + group four times as likely to die in the 2.5-year follow-up period compared to the Lmet-group. There was a strong association of lung metastases and subsequent brain metastases as well.

Survival in metastatic breast cancer patients is influenced by the hormone-receptor positivity status of the tumor, for which there are targeted treatments depending on tumor isotype [26]. The exact influence of individual hormone receptor status on survival in patients with metastatic lung lesions is unspecified. Furthermore, the accurate diagnosis and quantification of metastatic burden in the lungs is prone to type II error and inter-observer variance. We determined that the presence of lung metastases detected by AI specifically impacted survival in patients with ER+, PR + breast cancer (p = 0.00059) and HER2-tumors (p =

0.015) as well as patients with disease other than TNBC, TPBC, and HER2+ tumors in general. This suggests that accurately quantifying the lung metastasis tumor burden is critical in determining an accurate prognosis in patients with ER+, PR + breast cancer and HER2-tumors. Other isotypes do not necessarily reflect this trend and may reflect the lack of efficacy for treatment in such tumor isotypes (e.g. TNBC; ER-, PR-) or the relative rarity of specific isotypes (e.g. TPBC). Survival curves for HER2+ and TPBC tumors are suggestive of an influential effect of AI-detected metastases, but either lack sufficient follow-up, sample size, or combination of both. A 5-year longitudinal follow-up study would be helpful to adequately discriminate for these tumor isotypes.

4.1. Future applications

The use of radiomic biomarkers, specifically number and size of lung nodules, may play a yet unrealized role in cancer imaging and subsequent treatment. The potential benefit of directing more aggressive therapy to those with certain imaging criteria at initial staging requires further research. Radiomic data is currently being explored at a single time point in the patient's work-up, and likely will play an even larger role when applied in a longitudinal fashion (also known as delta radiomics) [14]. In this way, subtle nodule biology and nodule responses to therapy can be assessed and monitored more accurately. Further investigation could be made into the application of an AI algorithm in international and underserved populations that may have a scarcity of radiologists.

4.2. Limitations

There are some limitations to our study. AI algorithm limitations are secondary to model training as well as application. As demonstrated by Figure 1, nodule size measurements and discrepancies became more
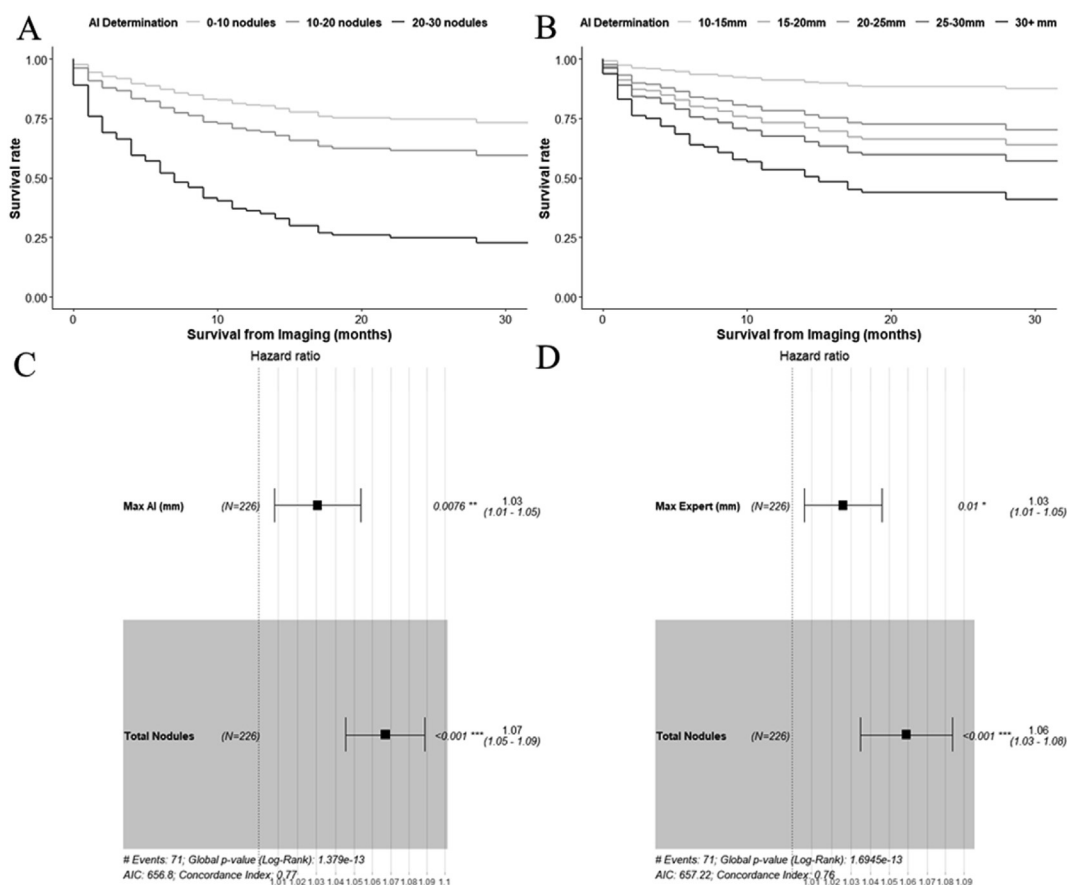
Figure 6. Cox-proportional hazard modelling and adjusted survival from imaging based on quantitative AI measurements including total metastases and maximum metastasis size. A. Adjusted survival curve for AI detected metastasis counts using 0–10, 10–20 and 20–30 nodules as breakpoints. Survival decreases disproportional to total nodule count. B. Adjusted survival curve for AI detected metastasis size. Survival decreases roughly proportional to metastasis size. C. Cox-proportional hazards model controlled for presence of lung metastases for AI-based measurements of total metastases and maximum first-dimension size (MaxFAI). Increase in MaxFAI and total nodule count is independently associated with probability of death (HR-MaxFAI = 1.03 (95% CI 1.01–1.05, p = 0.0076); HR-total metastases = 1.07 (95% CI 1.05–1.09, p < 0.001). D. Cox-proportional hazards model controlled for presence of lung metastases for expert-based measurement of maximum first-dimension metastasis size (MaxFExp) and AI collection of metastases count. Increase in MaxFExp and total metastasis count is independently associated with probability of death (HR-MaxFExp = 1.03 (95% CI 1.01–1.05, p = 0.01), HR-total metastases = 1.06 (95% CI 1.03–1.08, p < 0.001)).
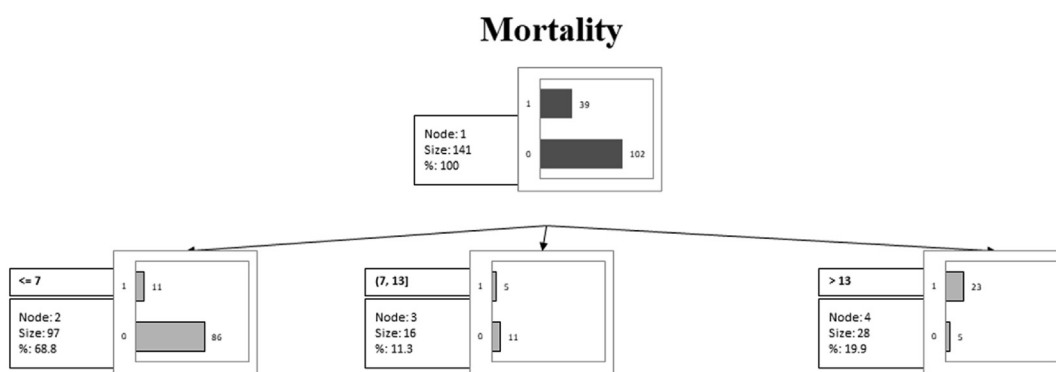

Figure 7. Classification and regression tree for explanation of deceased subjects using initial clinical, pathologic, and imaging staging, hormone receptor status, and AI-determination of the number/size of metastasis(es). The first factor establishing mortality risk is total nodule count. >13 nodules found by AI correlated strongly with overall mortality. 1 = Event (Death), 0 = No event (Alive). Size refers to number of patients who fall into the specified node.

variable, likely because the model was initially trained on lung nodules measuring up to 20–25 mm.

Due to the large number of nodules and the potentially limitless measurements, we only characterized the top 5 nodules by two-dimensional sizes, thus any agreement statistics are extrapolated beyond those five per patient. Therefore, we were unable to analyze the agreement between the experts and AI algorithm in terms of number of

detected nodules per patient or use 3D measurements as this would be time prohibitive to the experts and is not commonly used in practice. In practice radiologists would describe 30 nodules as "innumerable" so the evaluation of 30 nodules automatically could be add value.

Our study was performed at a single center and reflects the findings of the population encountered at our institution. However, it should be noted that this AI algorithm was previously tested and validated at
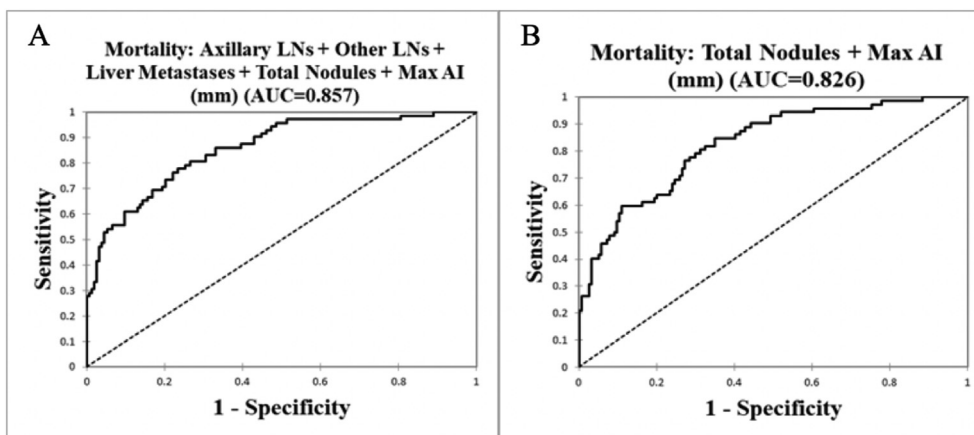
Figure 8. (A) ROC curve for prediction of mortality using initial clinical, pathological, and image-based staging characteristics as well as AI-determined quantitative lung metastasis measurements (McFadden R2 = 0.323). (B) ROC curve for prediction of mortality by MaxFAI and total nodule count (R2 = 0.243).

Table 6. Logistic regression parameters for prediction of mortality at 2.5 years amongst subjects in this cohort.

Variable	P	OR (95% CI)
Total Metastases (AI)	<0.001	1.11 (1.07–1.15)
Max AI (mm)	0.002	1.06 (1.02–1.10)
Axillary LN +	0.044	0.429 (0.189–0.977)
Other Nodes +	0.043	6.97 (1.061–45.8)
Liver Metastases	0.006	8.79 (1.84–42.0)

P-values represent log-likelihood significance testing for importance of the variable inclusion in the model.

multiple centers across the USA, Europe, and Asia. A multi-center study of breast cancer patients would allow generalization of our findings to a wider population.

Due to this being a retrospective analysis, pathologic confirmation was not available for all disease that met our metastatic criteria. In the future it would be ideal to confirm AI findings with tissue samples for all subjects.

There was a significantly increased incidence of axillary lymph node metastases in the patient group without lung nodules (p = 0.033). The patients who presented with distant metastases were less likely to be specifically evaluated for axillary disease likely resulting in this finding. Not assessed does not necessarily equivocate to the absence of disease, but it is noteworthy that there was a predominance in the population without nodules. It is important to note that our findings may represent a reporting bias due to lack of specific assessment of the axilla in certain patients.

4.3. Conclusions

Automated AI-based detection of lung metastases in breast cancer patients at initial staging chest CT performed well at identifying pulmonary metastases and demonstrated strong correlation between the total number and maximum size of lung metastases with future mortality.

Declarations

Author contribution statement

Madison R. Kocher; Jordan Chamberlin; Jeremy R. Burt: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Jeffrey Waltz: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Madalyn Snoddy; Ismail Kabakus; Natalie Stringer; Philipp Hoelzer; Pooyan Sahbaee: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Joseph Stephenson: Performed the experiments; Wrote the paper.

Jacob Kahn; Megan Mercer: Performed the experiments; Analyzed and interpreted the data.

Dhiraj Baruah; Gilberto Aquino; U. Joseph Schoepf: Analyzed and interpreted the data; Wrote the paper.

Data availability statement

The data that has been used is confidential.

Declaration of interests statement

The authors declare the following conflict of interests: Philipp Hoelzer and Pooyan Sahbaee are employed by Siemens Healthcare. Jeremy Burt received research grants from Siemens Healthcare; owner YellowDot Innovations. U Joseph Schoepf received research grants from Siemens Healthcare.

Additional information

No additional information is available for this paper.

References

[1] K.D. Miller, L. Nogueira, A.B. Mariotto, et al., Cancer treatment and survivorship statistics, CA A Cancer J. Clin. 69 (5) (2019) 363–385, 2019.

[2] A. Soni, Z. Ren, O. Hameed, et al., Breast cancer subtypes predispose the site of distant metastases, Am. J. Clin. Pathol. 143 (4) (2015) 471–478.

[3] A.G. Ording, U. Heide-Jorgensen, C.F. Christiansen, M. Norgaard, J. Acquavella, H.T. Sorensen, Site of metastasis and breast cancer mortality: a Danish nationwide registry-based cohort study, Clin. Exp. Metastasis 34 (1) (2017) 93–101.

[4] S.G. Armato, M.F. McNitt-Gray, A.P. Reeves, et al., The Lung Image Database Consortium (LIDC): an evaluation of radiologist variability in the identification of lung nodules on CT scans, Acad. Radiol. 14 (11) (2007) 1409–1421.

[5] J.K. Leader, T.E. Warfel, C.R. Fuhrman, et al., Pulmonary nodule detection with low-dose CT of the lung: agreement among radiologists, AJR Am. J. Roentgenol. 185 (4) (2005) 973–978.

[6] P.F. Pinsky, D.S. Gierada, P.H. Nath, E. Kazerooni, J. Amorosa, National lung screening trial: variability in nodule detection rates in chest CT studies, Radiology 268 (3) (2013) 865–873.

[7] G.D. Rubin, Lung nodule and cancer detection in computed tomography screening, J. Thorac. Imag. 30 (2) (2015) 130–138.

[8] Y. Ren, M.Y. Tsai, L. Chen, et al., A manifold learning regularization approach to enhance 3D CT image-based lung nodule classification, Int. J. Comput. Assist. Radiol. Surg. 15 (2) (2020) 287–295.

[9] K. Awai, K. Murao, A. Ozawa, et al., Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance, Radiology 230 (2) (2004) 347–352.

[10] S.M. Lee, J.B. Seo, J. Yun, et al., Deep learning applications in chest radiography and computed tomography: current state of the art, J. Thorac. Imag. 34 (2) (2019) 75–85.

[11] M. Nishio, O. Sugiyama, M. Yakami, et al., Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning, PLoS One 13 (7) (2018), e0200721.

[12] S.J. van Riel, F. Ciompi, M.M.W. Wille, et al., Malignancy risk estimation of pulmonary nodules in screening CTs: comparison between a computer model and human observers, PLoS One 12 (11) (2017), e0185032.

[13] W.L. Bi, A. Hosny, M. Schabath, et al., Artificial intelligence in cancer imaging: clinical challenges and applications, CA A Cancer J. Clin. 69 (2) (2019) 127–157.

[14] R.J. Gillies, M.B. Schabath, Radiomics Improves Cancer Screening and Early Detection, Cancer Epidemiol Biomarkers Prev, 2020.

[15] J.R. Ferreira Junior, M. Koenigkam-Santos, F.E. Garcia Cipriano, A.T. Fabro, P.M. de Azevedo-Marques, Radiomics-based features for pattern recognition of lung cancer histopathology and metastases, Comput. Methods Progr. Biomed. 159 (2018) 23–30.

[16] J. Chamberlin, M. Kocher, J. Waltz, et al., Automated detection of lung nodules and coronary artery calcium using artificial intelligence on low-dose CT scans for lung cancer screening: accuracy and prognostic value, BMC Med. 19 (1) (2021) 55.

[17] B. Yacoub, I.M. Kabakus, U.J. Schoepf, et al., Performance of an artificial intelligence-based platform against clinical radiology reports for the evaluation of noncontrast chest CT, Acad. Radiol. (2021).

[18] A.M. Fischer, A. Varga-Szemes, S.S. Martin, et al., Artificial intelligence-based fully automated per lobe segmentation and emphysema-quantification based on chest computed tomography compared with global initiative for chronic obstructive lung disease severity of smokers, J. Thorac. Imag. 35 (Suppl 1) (2020) S28–S34.

[19] A.M. Fischer, B. Yacoub, R. Savage, et al., Machine learning/deep neuronal network: routine application in chest computed tomography and workflow considerations, J. Thorac. Imag. 35 (Suppl 1) (2020) S21–S27.

[20] J.J. van den Broek, T. van Gestel, S.Q. Kol, et al., Dealing with indeterminate pulmonary nodules in colorectal cancer patients; a systematic review, Eur. J. Surg. Oncol. (2021).

[21] S. Cui, S. Ming, Y. Lin, et al., Development and clinical application of deep learning model for lung nodules screening on CT images, Sci. Rep. 10 (1) (2020) 13657.

[22] H. Jin, Z. Li, R. Tong, L. Lin, A deep 3D residual CNN for false-positive reduction in pulmonary nodule detection, Med. Phys. 45 (5) (2018) 2097–2107.

[23] A. Gupta, T. Saar, O. Martens, Y. Le Moullec, Automatic detection of multisize pulmonary nodules in CT images: large-scale validation of the false-positive reduction step, Med. Phys. 45 (3) (2018) 1135–1149.

[24] T. Tanvetyanon, L. Robinson, K.E. Sommers, et al., Relationship between tumor size and survival among patients with resection of multiple synchronous lung cancers, J. Thorac. Oncol. 5 (7) (2010) 1018–1024.

[25] R. Wang, Y. Zhu, X. Liu, X. Liao, J. He, L. Niu, The clinicopathological features and survival outcomes of patients with different metastatic sites in stage IV breast cancer, BMC Cancer 19 (1) (2019) 1091.

[26] A.G. Waks, E.P. Winer, Breast cancer treatment: a review, JAMA 321 (3) (2019) 288–300.