Original Article

Racial/ethnic reporting differences in cancer literature regarding machine learning *vs.* a radiologist: a systematic review and meta-analysis

Rahil Patel[1#^], Destie Provenzano[1#], Sherrie Flynt Wallington[2], Murray Loew[1], Yuan James Rao[3], Sharad Goyal[3]

[1]George Washington University School of Engineering and Applied Science, Washington, DC, USA; [2]George Washington University School of Nursing, Milken Institute School of Public Health, Washington, DC, USA; [3]Department of Radiation Oncology, George Washington University School of Medicine and Health Sciences, Washington, DC, USA

Contributions: (I) Conception and design: S Goyal, R Patel, D Provenzano; (II) Administrative support: SF Wallington, M Loew, YJ Rao, S Goyal; (III) Provision of study materials or patients: R Patel, D Provenzano; (IV) Collection and assembly of data: R Patel, D Provenzano, YJ Rao; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Sharad Goyal, MD. Department of Radiation Oncology, George Washington University School of Medicine and Health Sciences, 2150 Pennsylvania Ave., NW, Washington, DC 20037, USA. Email: shgoyal@mfa.gwu.edu.

Background: Machine learning (ML) has emerged as a promising tool to assist physicians in diagnosis and classification of patient conditions from medical imaging data. However, as clinical applications of ML become more common, there is concern about the prevalence of ethnoracial biases due to improper algorithm training. It has long been known that cancer outcomes vary for different racial/ethnic groups.

Methods: We reviewed 84 studies that reported results of ML algorithms compared to radiologists for cancer prediction to evaluate if algorithms targeted at cancer prediction account for potential ethnoracial biases in their training samples. The search engines used to extract the articles were: PubMed, MEDLINE, and Google Scholar. All studies published before May 2022 were extracted. Two researchers independently reviewed 115 articles and evaluated them for incorporation and inclusion of demographic information in the algorithm. Exclusion criteria were if an inappropriate imaging type was used, if they did not report benign *vs.* malignant cancer results, if the algorithm was not compared to a board-certified radiologist, or if they were not in English.

Results: Of the 84 studies included, 87% (n=73) reported demographic information and 38% (n=32) evaluated the effect of demographic information on model performance. However, only about 11% (n=9) of the articles reported racial/ethnic groups and about 4% (n=3) incorporated racial/ethnic information into their models. Of the nine studies that reported racial/ethnic information, the specified racial/ethnic minorities that were included the most were White/Caucasian (n=9/9) and Black/African American (n=8/9). Asian (n=4/9), American Indian (n=3/9), and Hispanic (n=2/9) were reported in less than half of the studies.

Conclusions: The lack of inclusion of not only racial/ethnic information but also other demographic information such as age, gender, body mass index (BMI), or patient history is indicative of a larger problem that exists within artificial intelligence (AI) for cancer imaging. It is crucial to report and consider demographics when considering not only AI for cancer, but also overall care of a cancer patient. The findings from this study highlight a need for greater consideration and evaluation of ML algorithms to consider demographic information when evaluating a patient population for training the algorithm.

Keywords: Racial biases; cancer; machine learning (ML); artificial intelligence (AI)

^ ORCID: 0009-0000-8362-086X.

Received: 14 May 2023; Accepted: 18 September 2023; Published online: 10 November 2023.
doi: 10.21037/jmai-23-31
View this article at: https://dx.doi.org/10.21037/jmai-23-31

Introduction

Background

Machine learning (ML) has emerged as a promising tool to assist physicians in diagnosis and classification of patient conditions from medical imaging data. ML algorithms have been used to diagnose many conditions including breast cancer (1), diabetes (2), heart disease (3), Parkinson's (4), and Coronavirus Disease 2019 (COVID-19) (5). However, as ML has become more common, so has more attention been directed to the impact of racial biases due to imbalanced datasets (6). A recent study found that despite the evident differences in how diabetes affects different populations, artificial intelligence (AI) algorithms often did not take racial/ethnic differences into account or even report what percentages persisted in their training sets (7). This is important as it has been well documented that health disparities and complications within diabetes exist for minority subgroups (8). A recent study found that diabetes occurs earlier and at lower body mass index (BMI) in Asian, Hispanic, and Black/African American populations than in White populations, indicating the need for increased screening (9). Results from ML studies are increasingly emphasizing the need for proper racial evaluation of models and datasets, as many studies including different racial/ethnic groups into their models are showing significant variation in performance by race, increasing the likeliness of potential biases and disparities in care (10-12).

Highlight box

Key findings

- This study found a lack of inclusion of racial/ethnic and other important demographic information in cancer literature comparing artificial intelligence (AI) to radiologists.

What is known and what is new?

- Machine learning (ML) holds promise in aiding diagnosis of patient conditions. However, studies are increasingly highlighting racial/ethnic biases in ML models despite evidence showing variation in condition status amongst different population groups.
- This study evaluates if AI algorithms targeted at cancer prediction account for ethnoracial biases in their model training.

What is the implication, and what should change now?

- Cancer outcomes and diagnoses show significant disparities across races. ML algorithms need to consider demographic information when training on a patient population. AI built on datasets that ignore race may lack generalizability to other patient populations. It is important to not only report this information for AI for cancer imaging, but to also consider it when training a model.

Rationale and knowledge gap

It has long been known that cancer outcomes vary for different racial/ethnic groups (13). For example, there are known disparities in cancer outcomes for Black/African American and Hispanic minorities suffering from breast cancer (14-16), prostate cancer (17), lung cancer (18), colorectal cancer (19), and pancreatic cancer (20). Beyond racial/ethnic status, there are known disparities by gender for cancers such as liver cancer, where men die at a much higher rate than women (21). It has been found that health disparities amongst racial/ethnic subgroups can be further exacerbated due to socioeconomic status (SES) (22), English proficiency (23), environmental pollution based on community locations (24), cultural diets and habits (25), and genetic factors (26). Lack of inclusion of minority subgroups in ML models for cancer diagnoses can have an even more direct impact on decreased performance and further health disparities. For example, ML provides opportunities for early detection of skin cancers such as melanoma, which allows for optimal efficacy of treatment (27). However, recent studies have found an underrepresentation of certain demographic groups such as Black/African American patients in dermatological predictive models which can lead to models unable to accurately predict cancers for darker skin types (28,29).

Objective

We previously conducted a review of 61 articles that compared the performance of radiologists to ML algorithms in regard to cancer diagnoses and prediction (30). It is important to also consider if algorithms targeted at cancer prediction account for different racial/ethnic groups present in their training samples and the adequacy of those samples. We surveyed literature regarding ML prediction

Table 1 Key terms used to identify literature included in this meta-analysis of literature

Subject	Key terms
Article type	Clinical study, clinical trial, controlled clinical trial, journal article, randomized controlled trial
Machine learning keywords	Machine learning, artificial intelligence, neural networks (NN), support vector machine (SVM), naïve Bayes, logistic regression, convolutional neural network (CNN), deep learning, random forest, decision tree
Clinical keywords	Cancer, radiologist, physician, clinician
Exclusion criteria	Does not contain prediction of benign vs. malignant histology, not in English, does not use board certified radiologist, does not compare algorithm to a radiologist
Additional filters	Radiologist deemed imaging technology inappropriate for type of cancer diagnosis (e.g., DEXA for breast cancer)

DEXA, dual energy X-ray absorptiometry.

of cancer and its performance compared to a radiologist's prediction to see if these articles also reported any racial/ethnic information. Cancer outcomes can vary greatly based on different racial/ethnic groups; as such, it is important to draw attention to the need for more diversity and better reporting within ML datasets. The study is presented in accordance with the PRISMA reporting checklist (available at https://jmai.amegroups.com/article/view/10.21037/jmai-23-31/rc) (31).

Methods

Literature review

Online literature databases (PubMed, MEDLINE, and Google Scholar) were searched for studies that reported results of ML algorithms compared to radiologists for cancer prediction. Literature search was conducted by two independent researchers and any variations were subsequently resolved. All studies published before May 2022 were considered. Papers were excluded if an inappropriate imaging type was used to train a model [such as dual energy X-ray absorptiometry (DEXA) to identify breast cancer], if they did not report results on prediction of cancer (benign *vs.* malignant), if they did not compare algorithm to a radiologist, if they did not use a board-certified radiologist, or if they were not in English. Key terms used to identify literature are summarized in *Table 1*. Papers were independently reviewed by the same two researchers for demographic and other reported information.

Quality assessment of literature

Alongside the quality assessment of reported patient information in this study, manuscripts were assessed for completeness and overall quality through a modified CLAIM checklist pertaining to ML within radiology (30,32). Manuscripts were reviewed for quality of ML algorithm, results, reported data, and risk of bias through an evaluation of reported metrics, use of correct validation datasets and methods, and reproducibility. Manuscripts were specifically searched for reported metrics [area under the receiver operating characteristic curve (AUC), accuracy, positive predictive value (PPV) negative predictive value (NPV), etc.], inclusion of a separate training and testing set and cross validation or hold-out sample, features included, and whether sufficient information was provided for replication. Manuscripts were assessed for image quality analysis, including whether it was conducted, mentioned, or if images were excluded or included based on quality assessment. Information regarding the quality assessment of literature is reported in Table S1.

Meta analysis

Studies included in this meta-analysis were evaluated for inclusion of demographic information in addition to information of algorithm and radiologist performance. Total papers identified, imaging type, and cancer site were summarized. Studies included in this review are cited in Table S2.

This study considered demographic information to be

age, racial/ethnic status, and gender. This study also looked to see if additional history or patient information (such as smoking status or BMI) was reported by the original publication. SES was not considered for this meta-analysis due to the overall lack of reporting in literature. Papers were evaluated first for inclusion of any demographic information regarding the patient population in the study overall. Papers were then evaluated for incorporation and inclusion of demographic information in the algorithm. A study was considered to have included demographic information into a model if an analysis was performed that looked at the statistical impact of at least one demographic variable on the algorithm performance (for instance, if the study included a table of P values for demographic information). For demographic information to be considered included in the model, the study had to include demographic information as a variable input, had to report a P value regarding performance of the demographic subgroup on the model, or had to evaluate model performance by demographic subgroup.

Literature that reported racial/ethnic information was evaluated for total makeup of patient population and inclusion of which categories of reported information.

Statistical analysis

Statistical analysis was performed to determine the prevalence and distribution of studies categorized by various reported demographic information. For studies presenting racial/ethnic information, we conducted further analysis to examine the frequency and distribution of each racial group. Manuscripts that incorporated racial/ethnic information into the algorithms were then reviewed for presence of statistical analyses regarding demographic variables and appropriate use of chosen statistical tests.

Results

Literature search results

The literature search resulted in 115 studies (*Figure 1*) (31). After screening for full text eligibility based on the inclusion criteria, 31 studies were excluded. The remaining 84 studies were identified for inclusion in this meta-analysis. Information regarding total studies, cancer sites, and imaging type are reported in *Table 2*. The most common cancer site identified in the review was the breast (26 studies), and the most common imaging method utilized was magnetic resonance imaging (MRI) (35 studies).

Breakdown of total demographic information reported and type of demographic information reported is summarized in *Figure 2*. A total of 73 studies reported demographic information, and 32 evaluated the effect of demographic information on model performance.

Only nine reported racial/ethnic groups, and only three of those studies incorporated racial/ethnic groups into their models. Of the nine studies that reported this information, the specified racial/ethnic minorities that were the most included were White/Caucasian (n=9/9) and Black/African American (n=8/9). Asian (n=4/9), American Indian (n=3/9), and Hispanic (n=2/9) were reported in less than half of the studies. Percentage of racial/ethnic minorities included in each patient population and which racial/ethnic minorities were reported are summarized in *Figure 3*.

Two of the three studies that incorporated racial/ethnic status into the studies' respective analyses reported statistical significance of race on the algorithm performance. Yala *et al.* demonstrated that inclusion of race/ethnicity caused significant diagnostic improvements for breast cancer of a hybrid deep learning model built on mammogram imaging and traditional risk factors (33). Performance of this deep learning model was improved compared to the current clinical standard for breast cancer risk prediction: the Tyrer-Cusick model. Yala *et al.* also evaluated the differences in the model improvements for different racial subtype groups. Inclusion of race/ethnicity caused statistically significant increases in the AUC for White/Caucasian (P<0.001) and Black/African American (P<0.001), which resulted in a model that performed well for both ethnicities.

In contrast, Beig *et al.* found that race was not a statistically significant factor in algorithm performance (P=0.97) (34). Incorporated demographic variables and reported P values for Yala *et al.* and Beig *et al.* are summarized in *Table 3*.

Additionally, Schaffter *et al.* did not report any statistical testing of race on algorithm performance (35). Schaffter *et al.* instead used a stratified sample for training the model that ensured racial groups were evenly distributed amongst training and testing groups.

Discussion

Key findings

The majority (73/84) of the manuscripts reviewed in this meta-analysis included some form of demographic

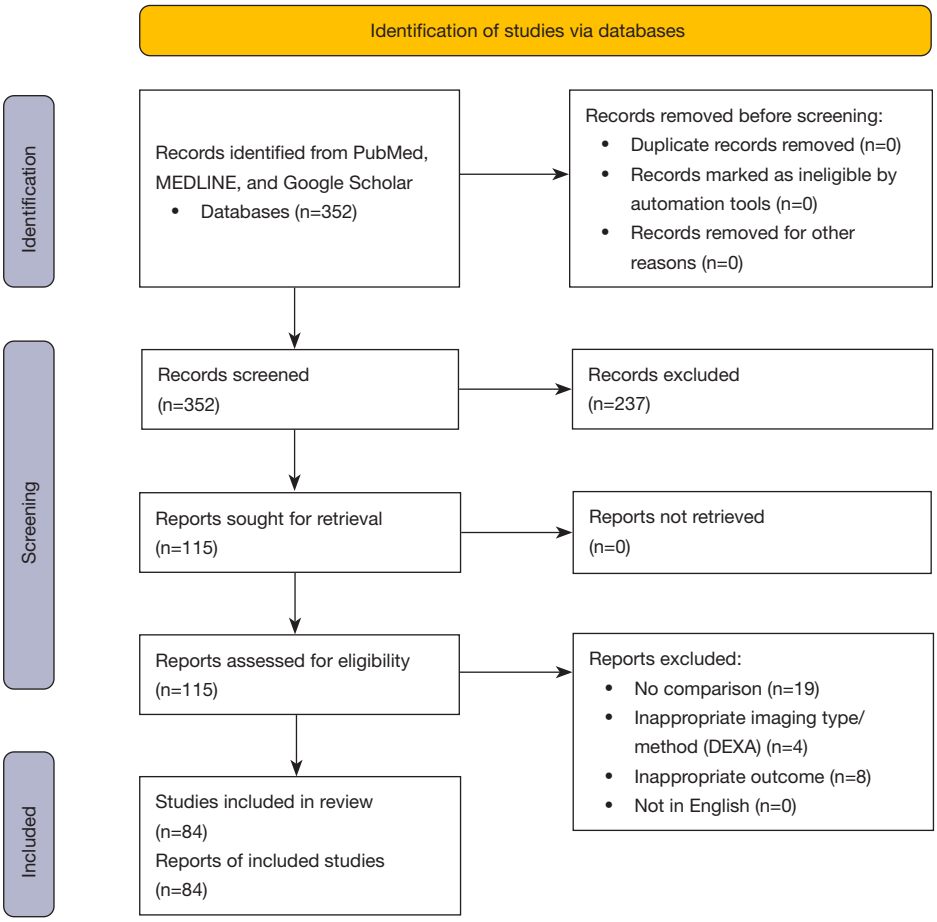

Figure 1 PRISMA flow diagram of the selection of studies to be included in the meta-analysis. DEXA, dual energy X-ray absorptiometry.

Table 2 Total studies, type of cancer, and type of imaging identified in studies included in this meta-analysis

Cancer site	Total (n=84)	Imaging method					
		CT	Mammogram	MRI	Shear wave elasticity images	Ultrasound	X-ray
Breast	26	1	8	6	0	9	2
Central nervous system (brain, spine)	8	0	0	8	0	0	0
Gastrointestinal	9	4	0	4	1	0	0
Genitourinary (prostate, bladder, kidney, adrenal)	11	5	0	6	0	0	0
Gynecology (ovaries, uterus)	9	0	0	9	0	0	0
Head and neck (lymph thyroid)	7	1	0	0	0	5	1
Sarcoma (soft tissue, fatty tissue)	4	1	1	2	0	0	0
Thorax (lungs, chest)	10	7	0	0	0	0	3

CT, computed tomography; MRI, magnetic resonance imaging.

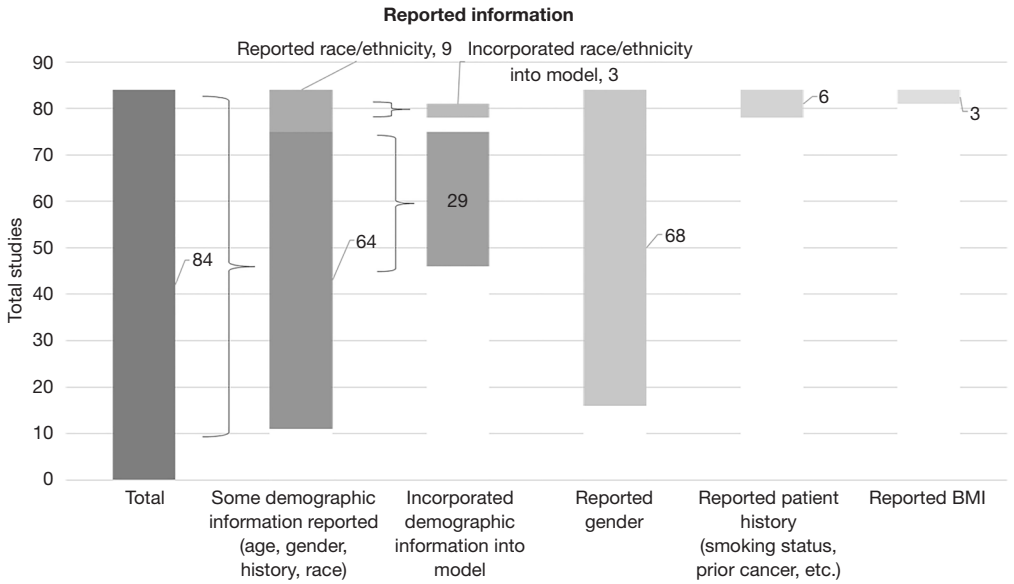

Figure 2 Demographic information reported by studies in this meta-analysis. BMI, body mass index.

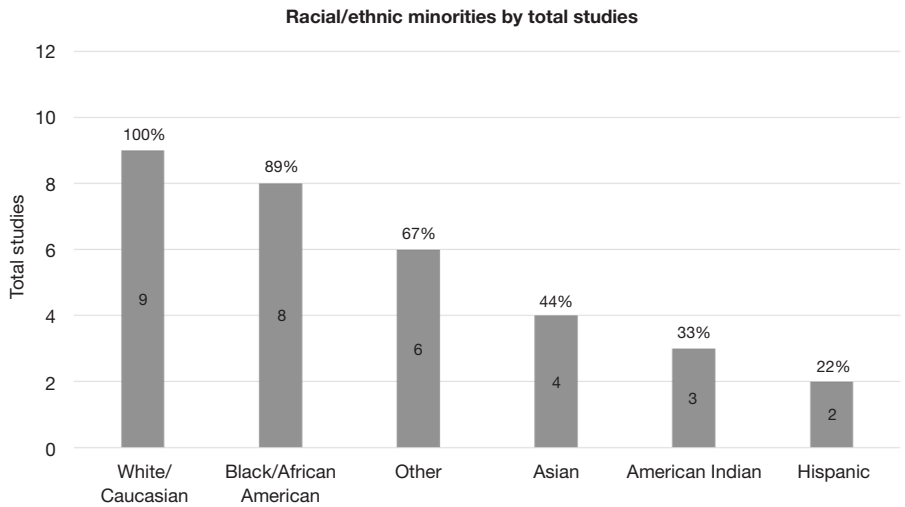

Figure 3 Distribution of racial/ethnic minorities by number of studies that reported racial/ethnic information.

information. However, only nine of these attempted to describe racial/ethnic information for the reported patient population. The lack of inclusion of information of not only race, but also demographic information such as age, gender, BMI, or patient history is indicative of a larger problem that exists within AI for cancer imaging. There are significant disparities in cancer outcomes and diagnoses for different races. Additionally, it has been shown that patient race can be identified from medical imaging (36). Many algorithms do not consider information such as race within consideration of prediction of clinically significant cancer. However, race is often correlated with different physical attributes such as BMI, which makes evaluation of race as a unique demographic variable important for downstream treatment of a patient as well. It is important to report and consider demographic information when considering not only AI for cancer, but overall care of a cancer patient. ML algorithms need to consider demographic information when evaluating a patient population for training the algorithm. AI built on datasets that do not consider race may be built

Table 3 Studies reporting statistical significance tests

Studies	Incorporated demographic variables	Reported P value
Yala *et al.*	Patient history	
	Premenopausal women	0.40
	Postmenopausal women	<0.001
	Race	
	White	<0.001
	Black/African American	<0.01
Beig *et al.*	Gender	0.38
	Age	<0.01
	Patient history (smoking status)	<0.01
	Race	0.97

entirely on homogeneous patient populations that do not generalize well to other datasets.

Implications

It is vital to account for racial/ethnic groups in ML models as carefully addressing these disparities can lead to improved treatment outcomes (chemotherapy, radiation therapy, etc.), reduction of inequalities present in the lack of technology access of certain minorities, prevention of biases, and further integration of relevant social determinants of health. Race has been shown to affect outcomes of cancer treatment such as chemotherapy (37) and radiation therapy (38,39); further, patient attributes that matter to treatment such as BMI are often strongly correlated with race/ethnicity status. By not including multiple races/ethnicities into the modeling process, the lack of heterogenous genetic information in open source datasets can further exacerbate a lack of access of certain races/ethnicities to personalized medicine and treatments. Improper training of models on non-representative datasets can cause racial biases that skew performance (40). Algorithms try to maximize overall prediction accuracy by optimizing for those individuals which appear frequently in the training data. This can cause variable performance for different racial/ethnic groups. Additionally, the performance of the predictors used in the models could substantially vary across different populations. There is a need to establish a diversity standard and prioritization of racial and social determinants of health data collection as well as the need for thorough evaluation of ML algorithms in race subgroups before clinical deployment to reduce bias. ML algorithms should be carefully designed to be ethical and reliable so that all demographic populations obtain equal benefit, with equal performance amongst groups and proper allocation of resources during clinical usage.

Limitations

This study was limited in that it only evaluated manuscripts that compared an AI to a radiologist. Many of these were feasibility studies that also did not report more information regarding the algorithm itself. Additionally, this study only evaluated demographic information for manuscripts evaluating AI for cancer. A broader review might reveal even more problems with demographic reporting present in the wider AI for medicine community.

Conclusions

An evaluation of 84 manuscripts regarding AI in cancer found that only 9 reported racial/ethnic information for the patient population involved in the study. Only 3 of these studies incorporated racial/ethnic information into the final model. It is important to not only report this information for AI for cancer imaging, but to also consider it when training a model.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the PRISMA reporting checklist. Available at https://jmai.amegroups.com/article/view/10.21037/jmai-23-31/rc

Peer Review File: Available at https://jmai.amegroups.com/article/view/10.21037/jmai-23-31/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://jmai.amegroups.com/article/view/10.21037/jmai-23-31/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related

to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Aljuaid H, Alturki N, Alsubaie N, et al. Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. Comput Methods Programs Biomed 2022;223:106951.
2. Naz H, Ahuja S. Deep learning approach for diabetes prediction using PIMA Indian dataset. J Diabetes Metab Disord 2020;19:391-403.
3. Bemando C, Miranda E, Aryuni M. Machine-learning-based prediction models of coronary heart disease using naïve Bayes and Random Forest Algorithms. 2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM). 2021:232-7.
4. Sherly Puspha Annabel L, Sreenidhi S, Vishali N. A novel diagnosis system for parkinson's disease using K-means clustering and decision tree. In: Sharma H, Gupta MK, Tomar GS, et al. editors. Communication and Intelligent Systems. Singapore: Springer; 2021:607-15.
5. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. Pattern Anal Appl 2021;24:1207-20.
6. The Lancet Digital Health. Can artificial intelligence help create racial equality in the USA? Lancet Digit Health 2021;3:e135.
7. Pham Q, Gamble A, Hearn J, et al. The Need for Ethnoracial Equity in Artificial Intelligence for Diabetes Management: Review and Recommendations. J Med Internet Res 2021;23:e22320.
8. Haw JS, Shah M, Turbow S, et al. Diabetes Complications in Racial and Ethnic Minority Populations in the USA. Curr Diab Rep 2021;21:2.
9. Aggarwal R, Bibbins-Domingo K, Yeh RW, et al. Diabetes Screening by Race and Ethnicity in the United States: Equivalent Body Mass Index and Age Thresholds. Ann Intern Med 2022;175:765-73.
10. Nayan M, Salari K, Bozzo A, et al. Predicting survival after radical prostatectomy: Variation of machine learning performance by race. Prostate 2021;81:1355-64.
11. Chen IY, Szolovits P, Ghassemi M. Can AI Help Reduce Disparities in General Medical and Mental Health Care? AMA J Ethics 2019;21:E167-179.
12. Allen A, Mataraso S, Siefkas A, et al. A Racially Unbiased, Machine Learning Approach to Prediction of Mortality: Algorithm Development Study. JMIR Public Health Surveill 2020;6:e22400.
13. Zavala VA, Bracci PM, Carethers JM, et al. Cancer health disparities in racial/ethnic minorities in the United States. Br J Cancer 2021;124:315-32.
14. Iqbal J, Ginsburg O, Rochon PA, et al. Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States. JAMA 2015;313:165-73.
15. Hirko KA, Rocque G, Reasor E, et al. The impact of race and ethnicity in breast cancer-disparities and implications for precision oncology. BMC Med 2022;20:72.
16. Petersen SS, Sarkissyan M, Wu Y, et al. Time to Clinical Follow-up after Abnormal Mammogram among African American and Hispanic Women. J Health Care Poor Underserved 2018;29:448-62.
17. Yamoah K, Lee KM, Awasthi S, et al. Racial and Ethnic Disparities in Prostate Cancer Outcomes in the Veterans Affairs Health Care System. JAMA Netw Open 2022;5:e2144027.
18. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. Cancer stat facts: Lung and bronchus cancer. Cancer Stat 2018. Available online: https://seer.cancer.gov/statfacts/html/lungb.html
19. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020;70:7-30.
20. Scarton L, Yoon S, Oh S, et al. Pancreatic Cancer Related Health Disparities: A Commentary. Cancers (Basel) 2018;10:235.
21. Rumgay H, Arnold M, Ferlay J, et al. Global burden of primary liver cancer in 2020 and predictions to 2040. J Hepatol 2022;77:1598-606.
22. Wisniewski JM, Walker B. Association of Simulated Patient Race/Ethnicity With Scheduling of Primary Care Appointments. JAMA Netw Open 2020;3:e1920010.
23. Cataneo JL, Meidl H, Ore AS, et al. The Impact of

Limited Language Proficiency in Screening for Breast Cancer. Clin Breast Cancer 2023;23:181-8.
24. Cheng I, Tseng C, Wu J, et al. Association between ambient air pollution and breast cancer risk: The multiethnic cohort study. Int J Cancer 2020;146:699-711.
25. Pinheiro PS, Callahan KE, Stern MC, et al. Migration from Mexico to the United States: A high-speed cancer transition. Int J Cancer 2018;142:477-88.
26. Charan M, Verma AK, Hussain S, et al. Molecular and Cellular Factors Associated with Racial Disparity in Breast Cancer. Int J Mol Sci 2020;21:5936.
27. Diaz MJ, Mark I, Rodriguez D, et al. Melanoma Brain Metastases: A Systematic Review of Opportunities for Earlier Detection, Diagnosis, and Treatment. Life (Basel) 2023;13:828.
28. Bhatt H, Shah V, Shah K, et al. State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: A comprehensive review. Intell Med 2023;3:180-90.
29. Kleinberg G, Diaz MJ, Batchu S, et al. Racial underrepresentation in dermatological datasets leads to biased machine learning models and inequitable healthcare. J Biomed Res (Middlet) 2022;3:42-7.
30. Provenzano D, Rao YJ, Goyal S, et al. Radiologist vs. machine learning: A comparison of performance in cancer imaging. 2021 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). 2021:1-10.
31. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
32. Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. Radiol Artif Intell 2020;2:e200029.
33. Yala A, Lehman C, Schuster T, et al. A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. Radiology 2019;292:60-6.
34. Beig N, Khorrami M, Alilou M, et al. Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas. Radiology 2019;290:783-92.
35. Schaffter T, Buist DSM, Lee CI, et al. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. JAMA Netw Open 2020;3:e200265.
36. Gichoya JW, Banerjee I, Bhimireddy AR, et al. AI recognition of patient race in medical imaging: a modelling study. Lancet Digit Health 2022;4:e406-14.
37. Hoeh B, Würnschimmel C, Flammia RS, et al. Effect of chemotherapy in metastatic prostate cancer according to race/ethnicity groups. Prostate 2022;82:676-86.
38. Lee W, Nelson R, Akmal Y, et al. Racial and ethnic disparities in outcomes with radiation therapy for rectal adenocarcinoma. Int J Colorectal Dis 2012;27:737-49.
39. Lee DJ, Zhao Z, Huang LC, et al. Racial variation in receipt of quality radiation therapy for prostate cancer. Cancer Causes Control 2018;29:895-9.
40. Zou J, Schiebinger L. AI can be sexist and racist - it's time to make it fair. Nature 2018;559:324-6.

doi: 10.21037/jmai-23-31
Cite this article as: Patel R, Provenzano D, Wallington SF, Loew M, Rao YJ, Goyal S. Racial/ethnic reporting differences in cancer literature regarding machine learning *vs.* a radiologist: a systematic review and meta-analysis. J Med Artif Intell 2023;6:25.

Supplementary

Table S1 Overview of quality assessment of literature

Subject of assessment	Number of studies
Total abstracts reviewed	352
Total full articles retrieved	115
Total studies included	84
Used training/testing set	82
Did not use or mention training/testing set	2
Inclusion of separate validation/hold-out/cross validation sample	68
Inclusion of only one measure of performance	37
Algorithms compared across accuracy only	9
Algorithms compared across AUC only	25
Algorithms compared across sensitivity/specificity only	3
Did not mention image quality	45
Lack of image quality assessment	11
Evaluated and excluded images	22
Evaluated but included poor quality images	6

https://dx.doi.org/10.21037/jmai-23-31

Table S2 Studies included in this meta-analysis

1. Akselrod-Ballin A, Chorev M, Shoshan Y, *et al.* Predicting Breast Cancer by Applying Deep Learning to Linked Health Records and Mammograms. Radiology. 2019;292(2):331-342. doi:10.1148:radiol.2019182622
2. Aldoj N, Lukas S, Dewey M, Penzkofer T. Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network. Eur Radiol. 202030(2):1243-1253.
3. Antonelli M, Johnston EW, Dikaios N, *et al.* Machine learning classifiers can predict Gleason pattern 4 prostate cancer with greater accuracy than experienced radiologists Eur Radiol. 201929(9)-4754-4764
4. Ariji Y, Fukuda 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.
5. Ashraf AB, Gavenonis SC, Daye D, Mies C, Rosen MA, Kontos D. A multichannel Markov random field framework for tumor segmentation with an application to classification of gene expression-based breast cancer recurrence risk. IEEE Trans Med Imaging. 2013.
6. Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A. Deep Learning in Mammography: Diagnostic Accuracy of a Multipurpose Image Analysis Software in the Detection of Breast Cancer. Invest Radiol. 2017
7. Becker AS, Mueller M, Stoffel E, Marcon M, Ghafoor S, Boss A. Classification of breast cancer in ultrasound imaging using a generic deep learning analysis software- a pilot study. Br J Radiol. 2018;91(1083)-20170576. doi-10.1259:bjr.20170576
8. Beig N, Khorrami M, Alilou M, *et al.* Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas. Radiology. 2019;290(3)-783-792. doi-10.1148:radiol.2018180910
9. Bickelhaupt S, Paech D, Kickingereder P, *et al.* Prediction of malignancy by a radiomic signature from contrast agent-free diffusion MRI in suspicious breast lesions found on screening mammography. J Magn Reson Imaging. 2017
10. Bonekamp D, Kohl S, Wiesenfarth M, *et al.* Radiomic Machine Learning for Characterization of Prostate Lesions with MRI- Comparison to ADC Values. Radiology. 2018;289(1)-128-137. doi-10.1148:radiol.2018173064
11. Buda M, Wildman-Tobriner B, Hoang JK, *et al.* Management of Thyroid Nodules Seen on US Images: Deep Learning May Match Performance of Radiologists. Radiology. 2019;292(3):695-701. doi:10.1148:radiol.2019181343
12. Byra M, Galperin M, Ojeda-Fournier H, *et al.* Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion. Med Phys. 2019;46(2):746-755. doi:10.1002:mp.13361
13. Cha MJ, Chung MJ, Lee JH, Lee KS. Performance of Deep Learning Model in Detecting Operable Lung Cancer With Chest Radiographs. J Thorac Imaging. 2019;34(2)-86-91. doi-10.1097
14. Chen K, Nie Y, Park S, *et al.* Development and Validation of Machine Learning-based Model for the Prediction of Malignancy in Multiple Pulmonary Nodules: Analysis from Multicentric Cohorts. Clin Cancer Res. 2021;27(8):2255-2265. doi:10.1158/1078-0432.CCR-20-4007
15. Choi JS, Han BK, Ko ES, *et al.* Effect of a Deep Learning Framework-Based Computer-Aided Diagnosis System on the Diagnostic Performance of Radiologists in Differentiating between Malignant and Benign Masses on Breast Ultrasonography. Korean J Radio
16. Corral JE, Hussein S, Kandel P, Bolan CW, Bagci U, Wallace MB. Deep Learning to Classify Intraductal Papillary Mucinous Neoplasms Using Magnetic Resonance Imaging. Pancreas. 2019;48(6):805-810. doi:10.1097:MPA.0000000000001327
17. Daimiel Naranjo I, Gibbs P, Reiner JS, *et al.* Breast Lesion Classification with Multiparametric Breast MRI Using Radiomics and Machine Learning: A Comparison with Radiologists' Performance. Cancers (Basel). 2022;14(7):1743. Published 2022 Mar 29. doi:10.3390/cancers14071743
18. Elmohr MM, Fuentes D, Habra MA, *et al.* Machine learning-based texture analysis for differentiation of large adrenal cortical tumours on CT. Clin Radiol. 2019;74(10):818.e1-818.e7. doi:10.1016:j.crad.2019.06.021.pdf
19. Fujioka T, Kubota K, Mori M, *et al.* Distinction between benign and malignant breast masses at breast ultrasound using deep learning method with convolutional neural network. Jpn J Radiol. 2019;37(6):466-472. doi:10.1007:s11604-019-00831-5
20. Gatos I, Tsantis S, Spiliopoulos S, *et al.* Temporal stability assessment in shear wave elasticity images validated by deep learning neural network for chronic liver disease fibrosis stage assessment. Med Phys. 2019;46(5).pdf
21. Gong J, Liu J, Hao W, Nie S, Wang S, Peng W. Computer-aided diagnosis of ground-glass opacity pulmonary nodules using radiomic features analysis. Phys Med Biol. 2019;64(13):135015. Published 2019 Jul 5. doi:10.1088:1361-6560:ab2757.pdf
22. Hamm CA, Wang CJ, Savic LJ, *et al.* Deep learning for liver tumor diagnosis part I- development of a convolutional neural network classifier for multi-phasic MRI. Eur Radiol. 2019;29(7)-3338-3347. doi-10.1007:s00330-019-06205-9
23. Jiang H, Liu X, Chen J, *et al.* Man or machine? Prospective comparison of the version 2018 EASL, LI-RADS criteria and a radiomics model to diagnose hepatocellular carcinoma. Cancer Imaging. 2019;19(1)84. Published 2019 Dec 5
24. Jun Y, Eo T, Kim T, *et al.* Deep-learned 3D black-blood imaging using automatic labelling technique and 3D convolutional neural networks for detecting metastatic brain tumors. Sci Rep. 2018;8(1)-9450. Published 2018 Jun 21. doi-10.1038:s41598-01.
25. Juntu J, Sijbers J, De Backer S, Rajan J, Van Dyck D. Machine learning study of several classifiers trained with texture analysis features to differentiate benign from malignant soft-tissue tumors in T1-MRI images. J Magn Reson Imaging. 2010;31(3
26. Kniep HC, Madesta F, Schneider T, *et al.* Radiomics of Brain MRI- Utility in Prediction of Metastatic Tumor Type. Radiology. 2019;290(2)-479-487. doi-10.1148:radiol.2018180946
27. Kooi T, Litjens G, van Ginneken B, *et al.* Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal. 2017;35:303-312. doi:10.1016:j.media.2016.07.007
28. Land WH, Margolis D, Gottlieb R, Krupinski EA, Yang JY. Improving CT prediction of treatment response in patients with metastatic colorectal carcinoma using statistical learning theory. BMC Genomics. 2010;11 Suppl 3(Suppl 3)-S15. Published 20
29. Lee E, Ha H, Kim HJ, *et al.* Differentiation of thyroid nodules on US using features learned and extracted from various convolutional neural networks. Sci Rep. 2019;9(1)-19854. Published 2019 Dec 27. doi-10.1038:s41598-019-56395-x.
30. Lee HJ, Nguyen AT, Ki SY, *et al.* Classification of MR-Detected Additional Lesions in Patients With Breast Cancer Using a Combination of Radiomics Analysis and Machine Learning. Front Oncol. 2021;11:744460. Published 2021 Dec 2. doi:10.3389/fonc.2021.744460
31. Li L, Liu Z, Huang H, Lin M, Luo D. Evaluating the performance of a deep learning-based computer-aided diagnosis (DL-CAD) system for detecting and characterizing lung nodules Comparison with the performance of double reading by radiologists.
32. Li Y, Jian J, Pickhardt PJ, *et al.* MRI-Based Machine Learning for Differentiating Borderline From Malignant Epithelial Ovarian Tumors: A Multicenter Study [published online ahead of print, 2020 Feb 11]. J Magn Reson Imaging. 2020
33. Lu Y, Liu L, Luan S, Xiong J, Geng D, Yin B. The diagnostic value of texture analysis in predicting WHO grades of meningiomas based on ADC maps: an attempt using decision tree and decision forest. Eur Radiol. 2019
34. Luo Y, Chen X, Chen J, *et al.* Preoperative Prediction of Pancreatic Neuroendocrine Neoplasms Grading Based on Enhanced Computed Tomography Imaging- Validation of Deep Learning with a Convolutional Neural Network. Neuroendocrinology. 20
35. Ma X, Hadjiiski LM, Wei J, *et al.* U-Net based deep learning bladder segmentation in CT urography. Med Phys. 2019;46(4)-1752-1765. doi-10.1002:mp.13438
36. Ma X, Wei J, Zhou C, *et al.* Automated pectoral muscle identification on MLO-view mammograms- Comparison of deep neural network to conventional computer vision. Med Phys. 2019;46(5)-2103-2114. doi-10.1002:mp.13451
37. Ma, J., He, N., Yoon, J. H., Ha, R., Li, J., Ma, W., Meng, T., Lu, L., Schwartz, L. H., Wu, Y., Ye, Z., Wu, P., Zhao, B., & Xie, C. (2021). Distinguishing benign and malignant lesions on contrast-enhanced breast cone-beam CT with deep learning neural architecture search. European journal of radiology, 142, 109878. https://doi.org/10.1016/j.ejrad.2021.109878
38. Malek M, Gity M, Alidoosti A, *et al.* A machine learning approach for distinguishing uterine sarcoma from leiomyomas based on perfusion weighted MRI parameters. Eur J Radiol. 2019;110:203-211. doi:10.1016:j.ejrad.2018.11.009
39. Mao N, Yin P, Wang Q, *et al.* Added Value of Radiomics on Mammography for Breast Cancer Diagnosis: A Feasibility Study. J Am Coll Radiol. 2019;16(4 Pt A):485-491. doi:10.1016:j.jacr.2018.09.041
40. Michael E, Ma H, Li H, Qi S. An Optimized Framework for Breast Cancer Classification Using Machine Learning. Biomed Res Int. 2022;2022:8482022. Published 2022 Feb 18. doi:10.1155/2022/8482022
41. Molloi S, Ding H, Feig S. Breast density evaluation using spectral mammography, radiologist reader assessment, and segmentation techniques- a retrospective study based on left and right breast comparison. Acad Radiol. 2015;22(8)-1052-1059.
42. Nakagawa M, Nakaura T, Namimoto T, *et al.* A multiparametric MRI-based machine learning to distinguish between uterine sarcoma and benign leiomyoma: comparison with 18F-FDG PET:CT. Clin Radiol. 2019;74(2):167.e1-167.e7. doi:10.1016:j.crad.2018.10.010
43. Nakagawa M, Nakaura T, Namimoto T, *et al.* Machine learning based on multi-parametric magnetic resonance imaging to differentiate glioblastoma multiforme from primary cerebral nervous system lymphoma. Eur J Radiol. 2018
44. Nakagawa M, Nakaura T, Namimoto T, *et al.* Machine Learning to Differentiate T2-Weighted Hyperintense Uterine Leiomyomas from Uterine Sarcomas by Utilizing Multiparametric Magnetic Resonance Quantitative Imaging Features. Acad Radiol. 2019
45. Nam JG, Park S, Hwang EJ, *et al.* Development and Validation of Deep Learning-based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. Radiology
46. Nikpanah M, Xu Z, Jin D, *et al.* A deep-learning based artificial intelligence (AI) approach for differentiation of clear cell renal cell carcinoma from oncocytoma on multi-phasic MRI. Clin Imaging. 2021;77:291-298. doi:10.1016/j.clinimag.2021.06.016
47. Nishio M, Nishizawa M, Sugiyama O, *et al.* Computer-aided diagnosis of lung nodule using gradient tree boosting and Bayesian optimization. PLoS One
48. O'Connell AM, Bartolotta TV, Orlando A, Jung SH, Baek J, Parker KJ. Diagnostic Performance of an Artificial Intelligence System in Breast Ultrasound. J Ultrasound Med. 2022;41(1):97-105. doi:10.1002/jum.15684
49. Patel BK, Ranjbar S, Wu T, *et al.* Computer-aided diagnosis of contrast-enhanced spectral mammography: A feasibility study. Eur J Radiol. 2018;98:207-213. doi:10.1016:j.ejrad.2017.11.024
50. Payabvash S, Aboian M, Tihan T, Cha S. Machine Learning Decision Tree Models for Differentiation of Posterior Fossa Tumors Using Diffusion Histogram Analysis and Structural MRI Findings. Front Oncol. 2020;10-71. Published 2020 Feb 7. doi-10.3389
51. Prior FW, Fouke SJ, Benzinger T, *et al.* Predicting a multi-parametric probability map of active tumor extent using random forests. Conf Proc IEEE Eng Med Biol Soc. 2013;2013:6478-6481. doi:10.1109:EMBC.2013.6611038.
52. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz CP, Patel BN, Yeom KW, Shpanskaya K, Blankenberg FG, Seekins J, Amrhein TJ, Mong DA, Halabi SS, Zucker EJ, Ng AY, Lungren MP. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med. 2018 Nov 20;15(11):e1002686. doi: 10.1371/journal.pmed.1002686. PMID: 30457988; PMCID: PMC6245676.
53. Ren T, Cattell R, Duanmu H, Huang P, Li H, Vanguri R, Liu MZ, Jambawalikar S, Ha R, Wang F, Cohen J, Bernstein C, Bangiyev L, Duong TQ. Convolutional Neural Network Detection of Axillary Lymph Node Metastasis Using Standard Clinical Breast MRI. Clin Breast Cancer. 2020 Jun;20(3):e301-e308. doi: 10.1016/j.clbc.2019.11.009. Epub 2019 Dec 5. PMID: 32139272.
54. Reyad YA, Berbar MA, Hussain M. Comparison of statistical, LBP, and multi-resolution analysis features for breast mass classification. J Med Syst. 2014 Sep;38(9):100. doi: 10.1007/s10916-014-0100-7. Epub 2014 Jul 19. PMID: 25037713.
55. Romeo V, Cuocolo R, Apolito R, *et al.* Clinical value of radiomics and machine learning in breast ultrasound: a multicenter study for differential diagnosis of benign and malignant lesions. Eur Radiol. 2021;31(12):9511-9519. doi:10.1007/s00330-021-08009-2
56. Romeo V, Maurea S, Cuocolo R, Petretta M, Mainenti PP, Verde F, Coppola M, Dell'Aversana S, Brunetti A. Characterization of Adrenal Lesions on Unenhanced MRI Using Texture Analysis: A Machine-Learning Approach. J Magn Reson Imaging. 2018 Jul;48(1):198-204. doi: 10.1002/jmri.25954. Epub 2018 Jan 17. PMID: 29341325.
57. Saha A, Grimm LJ, Ghate SV, Kim CE, Soo MS, Yoon SC, Mazurowski MA. Machine learning-based prediction of future breast cancer using algorithmically measured background parenchymal enhancement on high-risk screening MRI. J Magn Reson Imaging. 2019 Aug;50(2):456-464. doi: 10.1002/jmri.26636. Epub 2019 Jan 16. PMID: 30648316; PMCID: PMC6625842.
58. Schaffter T, Buist DSM, Lee CI, Nikulin Y, Ribli D, Guan Y, Lotter W, Jie Z, Du H, Wang S, Feng J, Feng M, Kim HE, Albiol F, Albiol A, Morrell S, Wojna Z, Ahsen ME, Asif U, Jimeno Yepes A, Yohanandan S, Rabinovici-Cohen S, Yi D, Hoff B, Yu T, Chaibub Neto E, Rubin DL, Lindholm P, Margolies LR, McBride RB, Rothstein JH, Sieh W, Ben-Ari R, Harrer S, Trister A, Friend S, Norman T, Sahiner B, Strand F, Guinney J, Stolovitzky G; and the DM DREAM Consortium, Mackey L, Cahoon J, Shen L, Sohn JH, Trivedi H, Shen Y, Buturovic L, Pereira JC, Cardoso JS, Castro E, Kalleberg KT, Pelka O, Nedjar I, Geras KJ, Nensa F, Goan E, Koitka S, Caballero L, Cox DD, Krishnaswamy P, Pandey G, Friedrich CM, Perrin D, Fookes C, Shi B, Cardoso Negrie G, Kawczynski M, Cho K, Khoo CS, Lo JY, Sorensen AG, Jung H. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. JAMA Netw Open. 2020 Mar 2;3(3):e200265. doi: 10.1001/jamanetworkopen.2020.0265. Erratum in: JAMA Netw Open. 2020 Mar 2;3(3):e204429. PMID: 32119094; PMCID: PMC7052735.
59. Shin I, Kim YJ, Han K, Lee E, Kim HJ, Shin JH, Moon HJ, Youk JH, Kim KG, Kwak JY. Application of machine learning to ultrasound images to differentiate follicular neoplasms of the thyroid gland. Ultrasonography. 2020 Jul;39(3):257-265. doi: 10.14366/usg.19069. Epub 2020 Feb 29. PMID: 32299197; PMCID: PMC7315296.
60. Stanzione A, Cuocolo R, Del Grosso R, Nardiello A, Romeo V, Travaglino A, Raffone A, Bifulco G, Zullo F, Insabato L, Maurea S, Mainenti PP. Deep Myometrial Infiltration of Endometrial Cancer on MRI: A Radiomics-Powered Machine Learning Pilot Study. Acad Radiol. 2021 May;28(5):737-744. doi: 10.1016/j.acra.2020.02.028. Epub 2020 Mar 28. PMID: 32229081.
61. Suh HB, Choi YS, Bae S, Ahn SS, Chang JH, Kang SG, Kim EH, Kim SH, Lee SK. Primary central nervous system lymphoma and atypical glioblastoma: Differentiation using radiomics approach. Eur Radiol. 2018 Sep;28(9):3832-3839. doi: 10.1007/s00330-018-5368-4. Epub 2018 Apr 6. PMID: 29626238.
62. Sun C, Zhang Y, Chang Q, Liu T, Zhang S, Wang X, Guo Q, Yao J, Sun W, Niu L. Evaluation of a deep learning-based computer-aided diagnosis system for distinguishing benign from malignant thyroid nodules in ultrasound images. Med Phys. 2020 Sep;47(9):3952-3960. doi: 10.1002/mp.14301. Epub 2020 Jun 25. PMID: 32473030.
63. Sun W, Zheng B, Qian W. Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis. Comput Biol Med. 2017 Oct 1;89:530-539. doi: 10.1016/j.compbiomed.2017.04.006. Epub 2017 Apr 13. PMID: 28473055.
64. Sun XY, Feng QX, Xu X, Zhang J, Zhu FP, Yang YH, Zhang YD. Radiologic-Radiomic Machine Learning Models for Differentiation of Benign and Malignant Solid Renal Masses: Comparison With Expert-Level Radiologists. AJR Am J Roentgenol. 2020 Jan;214(1):W44-W54. doi: 10.2214/AJR.19.21617. Epub 2019 Sep 25. PMID: 31553660.
65. Trivizakis E, Ioannidis GS, Melissianos VD, Papadakis GZ, Tsatsakis A, Spandidos DA, Marias K. A novel deep learning architecture outperforming 'off-the-shelf' transfer learning and feature-based methods in the automated assessment of mammographic breast density. Oncol Rep. 2019 Nov;42(5):2009-2015. doi: 10.3892/or.2019.7312. Epub 2019 Sep 12. PMID: 31545461; PMCID: PMC6787954.
66. Uhlig J, Biggemann L, Nietert MM, *et al.* Discriminating malignant and benign clinical T1 renal masses on computed tomography A pragmatic radiomics and machine learning approach. Medicine (Baltimore). 202099(16)e19725.pdf
67. Urushibara A, Saida T, Mori K, *et al.* The efficacy of deep learning models in the diagnosis of endometrial cancer using MRI: a comparison with radiologists. BMC Med Imaging. 2022;22(1):80. Published 2022 Apr 30. doi:10.1186/s12880-022-00808-3
68. Urushibara, A., Saida, T., Mori, K., Ishiguro, T., Sakai, M., Masuoka, S., Satoh, T., & Masumoto, T. (2021). Diagnosing uterine cervical cancer on a single T2-weighted image: Comparison between deep learning versus radiologists. European journal of radiology, 135, 109471. https://doi.org/10.1016/j.ejrad.2020.109471
69. von Schacky CE, Wilhelm NJ, Schäfer VS, *et al.* Development and evaluation of machine learning models based on X-ray radiomics for the classification and differentiation of malignant and benign bone tumors [published online ahead of print, 2022 Apr 9]. Eur Radiol. 2022;10.1007/s00330-022-08764-w. doi:10.1007/s00330-022-08764-w
70. Vos M, Starmans MPA, Timbergen MJM, van der Voort SR, Padmos GA, Kessels W, Niessen WJ, van Leenders GJLH, Grünhagen DJ, Sleijfer S, Verhoef C, Klein S, Visser JJ. Radiomics approach to distinguish between well differentiated liposarcomas and lipomas on MRI. Br J Surg. 2019 Dec;106(13):1800-1809. doi: 10.1002/bjs.11410. PMID: 31747074; PMCID: PMC6899528.
71. Wang R, Cai Y, Lee IK, Hu R, Purkayastha S, Pan I, Yi T, Tran TML, Lu S, Liu T, Chang K, Huang RY, Zhang PJ, Zhang Z, Xiao E, Wu J, Bai HX. Evaluation of a convolutional neural network for ovarian tumor differentiation based on magnetic resonance imaging. Eur Radiol. 2021 Jul;31(7):4960-4971. doi: 10.1007/s00330-020-07266-x
72. Wang T, Gong J, Li Q, Chu C, Shen W, Peng W, Gu Y, Li W. A combined radiomics and clinical variables model for prediction of malignancy in T2 hyperintense uterine mesenchymal tumors on MRI. Eur Radiol. 2021 Aug;31(8):6125-6135. doi: 10.1007/s00330-020-07678-9.
73. Wang Y, Choi EJ, Choi Y, Zhang H, Jin GY, Ko SB. Breast Cancer Classification in Automated Breast Ultrasound Using Multiview Convolutional Neural Network with Transfer Learning. Ultrasound Med Biol. 2020 May;46(5):1119-1132. doi: 10.1016/j.ultrasmedbio.2020.01.001. Epub 2020 Feb 12. PMID: 32059918.
74. Wei L, Yang Y, Nishikawa RM, Jiang Y. A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications. IEEE Trans Med Imaging. 2005 Mar;24(3):371-80. doi: 10.1109/tmi.2004.842457. PMID: 15754987.
75. Wu E, Hadjiiski LM, Samala RK, Chan HP, Cha KH, Richter C, Cohan RH, Caoili EM, Paramagul C, Alva A, Weizer AZ. Deep Learning Approach for Assessment of Bladder Cancer Treatment Response. Tomography. 2019 Mar;5(1):201-208. doi: 10.18383/j.tom.2018.00036. PMID: 30854458; PMCID: PMC6403041.
76. Wu M, Krishna S, Thornhill RE, Flood TA, McInnes MDF, Schieda N. Transition zone prostate cancer: Logistic regression and machine-learning models of quantitative ADC, shape and texture features are highly accurate for diagnosis. J Magn Reson Imaging. 2019 Sep;50(3):940-950. doi: 10.1002/jmri.26674. Epub 2019 Jan 30. PMID: 30701625.
77. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. Radiology. 2019 Jul;292(1):60-66. doi: 10.1148/radiol.2019182716. Epub 2019 May 7. PMID: 31063083.
78. Yanagawa M, Niioka H, Hata A, Kikuchi N, Honda O, Kurakami H, Morii E, Noguchi M, Watanabe Y, Miyake J, Tomiyama N. Application of deep learning (3-dimensional convolutional neural network) for the prediction of pathological invasiveness in lung adenocarcinoma: A preliminary study. Medicine (Baltimore). 2019 Jun;98(25):e16119. doi: 10.1097/MD.0000000000016119. PMID: 31232960; PMCID: PMC6636940.
79. Zhang XP, Wang ZL, Tang L, Sun YS, 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 Jan 11;11:10. doi: 10.1186/1471-2407-11-10. PMID: 21223564; PMCID: PMC3025970
80. Zhao CK, Ren TT, Yin YF, *et al.* A Comparative Analysis of Two Machine Learning-Based Diagnostic Patterns with Thyroid Imaging Reporting and Data System for Thyroid Nodules: Diagnostic Performance and Unnecessary Biopsy Rate. Thyroid. 2021;31(3):470-481. doi:10.1089/thy.2020.0305
81. Zhao HB, Liu C, Ye J, *et al.* A comparison between deep learning convolutional neural networks and radiologists in the differentiation of benign and malignant thyroid nodules on CT images. Endokrynol Pol. 2021;72(3):217-225. doi:10.5603/EP.a2021.0015
82. Zhao SS, Feng XL, Hu YC, Han Y, Tian Q, Sun YZ, Zhang J, Ge XW, Cheng SC, Li XL, Mao L, Shen SN, Yan LF, Cui GB, Wang W. Better efficacy in differentiating WHO grade II from III oligodendrogliomas with machine-learning than radiologist's reading from conventional T1 contrast-enhanced and fluid attenuated inversion recovery images. BMC Neurol. 2020 Feb 7;20(1):48. doi: 10.1186/s12883-020-1613-y. PMID: 32033580; PMCID: PMC7007642.
83. Zhao X, Zhou Y, Zhang Y, *et al.* Radiomics Based on Contrast-Enhanced MRI in Differentiation Between Fat-Poor Angiomyolipoma and Hepatocellular Carcinoma in Noncirrhotic Liver: A Multicenter Analysis. Front Oncol. 2021;11:744756. Published 2021 Oct 13. doi:10.3389/fonc.2021.744756
84. Zhou ZG, Liu F, Jiao LC, Wang ZL, Zhang XP, Wang XD, Luo XZ. An evidential reasoning based model for diagnosis of lymph node metastasis in gastric cancer. BMC Med Inform Decis Mak. 2013 Nov 6;13:123. doi: 10.1186/1472-6947-13-123. PMID: 24195733; PMCID: PMC3827004

 https://dx.doi.org/10.21037/jmai-23-31