MARKSMAN™ HEOR & RWE SOLUTIONS

Extent of Use of Artificial Intelligence and Machine Learning Protocols in Cancer Diagnosis: A Scoping Review

Amit Dang, Dimple Dang, Vallish BN, Aditya Bhardwaj

MarksMan Healthcare Communications, Hyderabad, India

Background

- A large number of modalities for cancer diagnosis that use different artificial intelligence (AI) and machine learning (ML) protocols are currently in various stages of development and validation across the world[1]
- Highly encouraging results are reported in terms of sensitivity, specificity, and accuracy of most of these AI/ML protocols during validation studies that are conducted under experimental settings that usually use retrospective patient databases
- We wanted to evaluate to what extent these protocols would perform under real-world conditions, and whether the physicians would routinely adopt these AI/ML modalities for clinical decision-making based on their superlative performance in validation tests

Objective

- To systematically map the extent of actual use of AI/ ML protocols for diagnosing cancer in prospective settings across the world
- Research question: 'What is known from published literature about the extent of actual usage of AI/ML protocols in cancer diagnosis in prospective (clinical trial/ real-world) settings, such that the diagnosis by the AI/ML protocol aids in clinical decision making?'

Methodology

- **Type of study:** Systematic Literature Review
- **SLR protocol:** drafted as per PRISMA guidelines, and registered prospectively with Open Science Framework on 3rd January 2020 (https://osf.io/643uq)
- **Databases searched:** PubMed, Google Scholar (first 200 hits)
- **Date of search:** From inception till 17th May 2021

Eligibility Criteria

Facet	Inclusion	Exclusion
Population	• Humans suffering from any type of cancer • Any age, any gender	• No human subjects • No cancer
Intervention	• Papers which had described the actual usage of AI/ML protocol for diagnosis of cancer in such a way that the AI/ML diagnosis resulted in or aided in clinical decision making	• AI/ ML protocol used for any application apart from cancer diagnosis or staging • Robotic surgeries • AI/ ML for estimating cancer prognosis
Comparator	• Any comparator	• No restriction
Outcome	• Any outcome which described the application of AI/ML in cancer diagnosis • AI/ML protocol has been used to newly diagnose a cancer or performing staging of a patient already diagnosed with cancer, thereby facilitating clinical decision-making	• All other outcomes
Study design	• Prospective patient enrolment • Clinical trial or real-world setting	• Retrospective data analysis • Studies describing training, testing, or validation of AI/ ML protocols • Reviews, editorials, commentaries

- Post-hoc analysis: After completing planned data extraction, a post-hoc analysis of all the retrieved records was performed to identify studies that described the validation of AI/ML protocol (either using standardized patient databases or prospectively enrolled patients) without their actual usage. Data pertaining to the types of cancer studied, the nature of AI/ML protocol being employed, the year of publication of the study, the country of the first author, the location of the study site, and the number of patients/ lesions/ images being used for the validation of the AI/ML protocol were extracted.

Result

Identification

Records identified through database searching (PubMed) (n = 951)

Additional records identified through other sources (n = 10)

Screening

Records after duplicates removed (n = 960)

Records excluded, with reasons (n = 793)
- Not cancer patients (n = 11)
- Intervention not related to cancer diagnosis (n = 648)
- Outcomes not related to cancer diagnosis (n = 51)
- Not prospective study design (n = 82)
- Abstract not available (n = 1)

Records screened (n = 960)

Eligibility

Full-text articles assessed for eligibility (n = 167)

Full-text articles excluded, with reasons (n = 149)
- Intervention not related to cancer diagnosis (n = 81)
- Outcomes not related to cancer diagnosis (n = 1)
- Not prospective study design (n = 62)
- Unable to retrieve the full text (n = 5)

Included

Studies included in qualitative synthesis (n = 18)

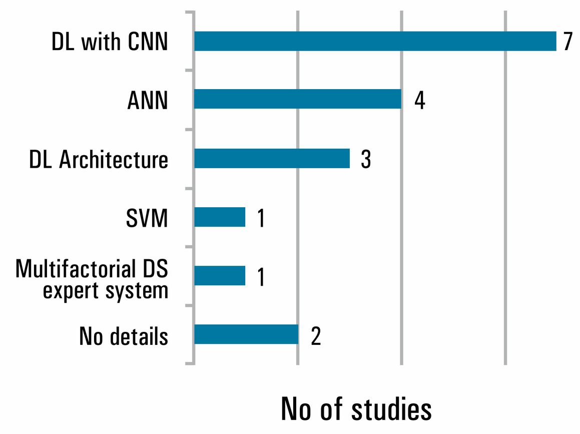
- Total articles included: **18**[3-20]
- First authors were from **10 different countries**
- Year of publication:
  - Before 2000: 2 studies
  - 2000-2010: 4 studies
  - 2011-2020: 12 studies
- Participants of studies:
  - 1 study: 5 different countries
  - Remaining studies: 10 different countries
- All studies were prospective, observational studies
- 8 studies had randomized patients upon recruitment

Main Characteristics of Included Studies

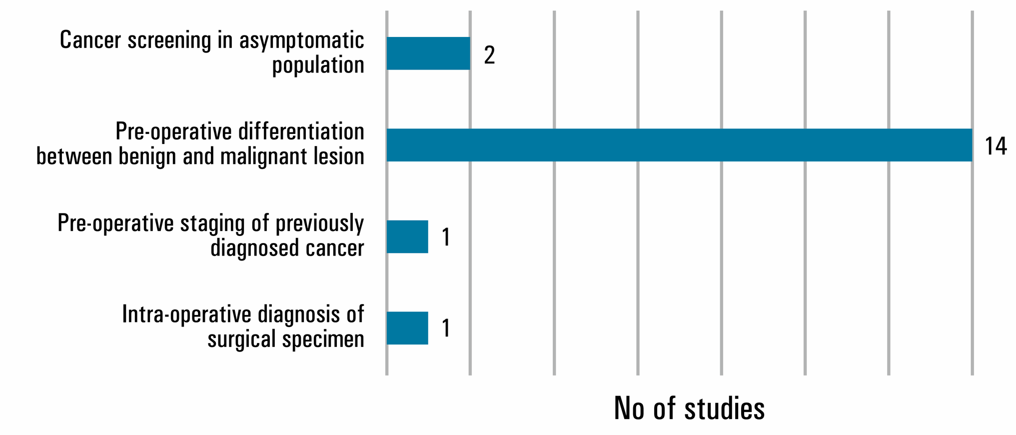
No.	Study, Year[3-20]	1st author country	Cancer studied	Type of lesions studied	AI/ML protocol	No. of patients	Male (%)	Female (%)	No. of lesions studied
1	Mori Y et al, 2018	Japan	Colorectal cancer	Colorectal Polyps	ML, SVM	325	235 (72.3%)	90 (27.7%)	466
2	Wang P et al, 2019*	China	Colorectal cancer	Colorectal Polyps	DL architecture	1,058	512 (48.4%)	546 (51.6%)	767
3	Su JR et al, 2019*	China	Colorectal cancer	Colorectal polyps	CNN, DL	623	307 (49.3%)	316 (50.7%)	442
4	Wang P et al, 2020*	China	Colorectal cancer	Colorectal polyps	DL	369	179 (48.5%)	190 (51.5%)	811
5	Repici A et al, 2020*	Italy	Colorectal cancer	Colorectal polyps	CNN, DL	685	337 (49.2%)	348 (50.8%)	493
6	Gong D et al, 2020*	China	Colorectal cancer	Colorectal polyps	CNN, DL	704	345 (49.0%)	359 (51.0%)	369
7	Wang P et al, 2020*	China	Colorectal cancer	Colorectal polyps	DL	962	495 (51.5%)	467 (48.5%)	809
8	Liu WN et al, 2020*	China	Colorectal cancer	Colorectal polyps	CNN, DL	1026	551 (53.7%)	475 (46.3%)	734
9	Dreiseitl S et al, 2009	Austria	Skin cancer	PSL	ANN-based DS tool	458	NA	NA	3,021
10	Fink C et al, 2017	Germany	Skin cancer	PSL	Not specified	111	59 (53.2%)	52 (46.8%)	346
11	Walker BN et al, 2019	USA	Skin cancer	PSL	CNN, DL	63	34 (54.0%)	29 (46.0%)	63
12	Kok MR et al, 1996	Netherlands	Cervical cancer screening	Cervical smear	ANN-based DS tool	91,294	0	91,294 (100%)	91,294
13	Nieminen P et al, 2002*	Finland	Cervical cancer screening	Cervical smear	ANN-based DS tool	108,686	0	108,686 (100%)	108,686
14	Hollon TC et al, 2020	USA	Brain cancer	Intra-op surgical specimen	CNN, DL	278	NA	NA	278
15	de Veld DC et al, 2004	Netherlands	Cancer of Oral Cavity	Oral mucosal lesion	PCA; ANN	155	NA	NA	176
16	Li L et al, 2019	China	Lung cancer	Lung nodules	CNN, DL	346	221 (63.9%)	125 (36.1%)	1916
17	Lucidarme O et al, 2010	France	Ovarian cancer	TVS image of ovary	Not specified	264	0	264 (100%)	375
18	Chang PL et al, 1999	Taiwan	Prostate cancer	Multiple parameters	Multifactorial DS system	43	43 (100%)	0	043

Note: *Randomization was done in these studies; ANN: Artificial neural network; CNN: Convoluted neural network; DL: Deep learning; DS: Decision support; ML: Machine learning; PCA: Principal Component Analysis; PSL: Pigmented skin lesions; SVM: support vector machine

AI/ ML protocol



Purpose of using AI/ ML protocol



Diagnostic Performance of Included Studies

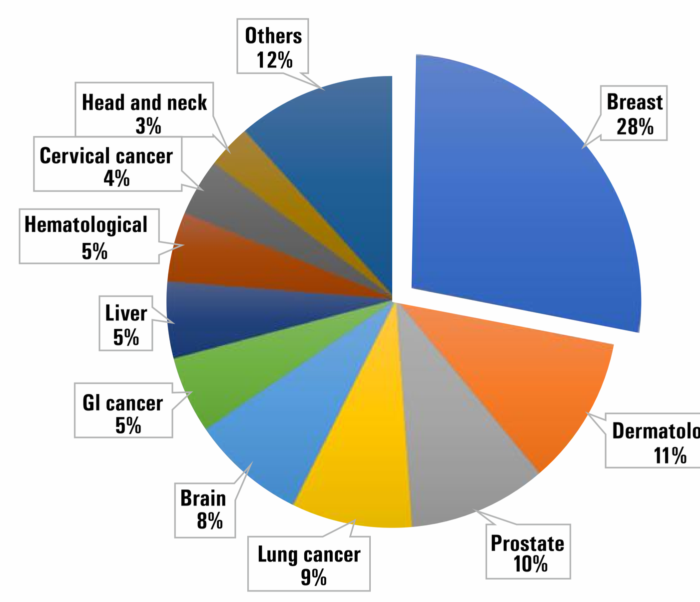
No	Study, Year[3-20]	Performance of AI/ML diagnosis as compared to human diagnosis	Sensitivity of the AI/ML protocol	Specificity of the AI/ML protocol	Accuracy of the AI/ML protocol	PPV of the AI/ML protocol	NPV of the AI/ML protocol
1	Chang PL et al, 1999	AI improves human diagnosis	92%	84%	88.40%	NA	NA
2	Lucidarme O et al, 2010	AI improves human diagnosis	98%	88%	NA	NA	NA
3	Wang P et al, 2019	AI improves human diagnosis	NA	NA	NA	NA	NA
4	Su JR et al, 2019	AI improves human diagnosis	NA	NA	NA	NA	NA
5	Repici A et al, 2020	AI improves human diagnosis	NA	NA	NA	NA	NA
6	Gong D et al, 2020	AI improves human diagnosis	NA	NA	NA	NA	NA
7	Wang P et al, 2020	AI improves human diagnosis	NA	NA	NA	NA	NA
8	Liu WN et al, 2020	AI improves human diagnosis	NA	NA	NA	NA	NA
9	Mori Y et al, 2018	AI is better than human diagnosis	NA	NA	98.10%	NA	93.7% to 96.5%
10	Li L et al, 2019	AI is better than human diagnosis	86.20%	NA	NA	57.00%	NA
11	Hollon TC et al, 2020	AI is better than human diagnosis	NA	NA	94.60%	NA	NA
12	Wang P et al, 2020	AI is better than human diagnosis	NA	NA	NA	NA	NA
13	Kok MR et al, 1996	AI is similar to human diagnosis	NA	NA	NA	NA	NA
14	Nieminen P et al, 2002	AI is similar to human diagnosis	NA	92.50%	NA	55%	NA
15	de Veld DC et al, 2004	Comparison not performed	NA	NA	NA	NA	NA
16	Fink C et al, 2017	Comparison not performed	100%	68.50%	2.30%	2.80%	100%
17	Walker BN et al, 2019	Comparison not performed	86% (system B); 91% (system A)[1]	69% (system B)[1]	NA	88.90%	88.90%
18	Dreiseitl S et al, 2009	Depends on the user's background	72%	82%	NA	NA	NA

Note: [1]System A is a deep learning classifier whose outputs from image processing of pigmented skin lesions were converted into sound waves, which were once again classified by System B.
PPV: Positive predictive value; NPV: Negative predictive value

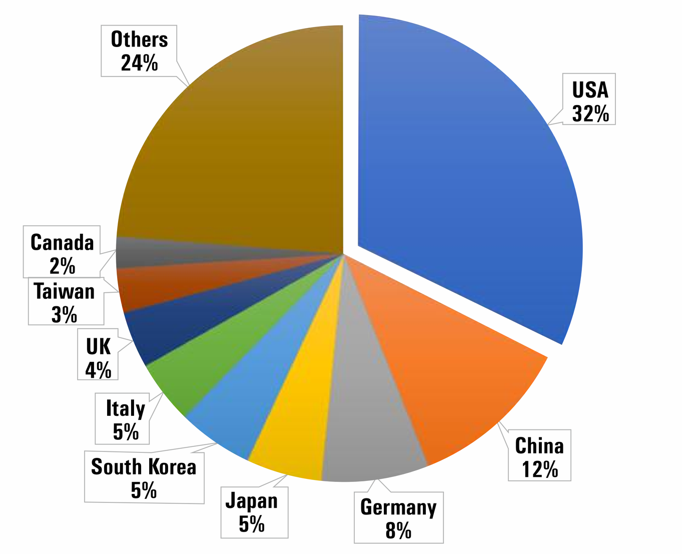
Post-hoc Analysis

- 223 studies described validation of an AI/ML protocol in cancer diagnosis
- A huge variation in the number of samples/ patients/ lesions/ images included for validation of the AI/ML protocol was observed
  - Patient numbers ranged from 8 to 84,424
  - Image/ lesion numbers ranged from 15 to 10,36,496
- Most frequent cancer for which:
  - AI/ ML validation was done: Breast cancer
  - AI/ ML protocol was actually used: colorectal cancer

Type of cancers in which AI/ML protocols were validated



Country of first author of studies validating AI/ML protocols in cancer diagnosis



Discussion

- Only 18/96 (1.9%) of initial hits on AI/ ML have actually used AI/ML protocols for diagnostic decision making in cancer; most excluded studies focused on validation of AI/ ML protocols
- Most studies concluded that AI/ML protocol is able to improve the human diagnosis, especially that made by the less experienced clinician: AI/ML protocols have a potential to significantly improve upon the prevailing diagnostic capabilities
- Meaningful translation of AI/ ML research into oncology diagnosis is lacking
- Performance of AI/ML protocols in validation studies is much better than that in real world studies
- Large number of validation tests, but few number of actual usage studies
- Disconnect between most frequent cancer in validation studies (breast) vs actual use (colorectal cancer)
- Large variations in the number of sample sizes in validation tests: lack of regulation in new diagnostic tests, unlike the stringent drug approval regulations

Study limitations

- Literature search restricted to PubMed and English language articles

Conclusions

- A meaningful translation from validation of AI/ML protocols to their actual usage in cancer diagnosis is lacking
- Development of regulatory framework specific for AI/ML usage in healthcare is essential

References

1 Kann BH et al. Oncology (Williston Park). 2019;33:46-53.
2 Whiting PF et al. Ann Intern Med. 2011;155:529-536.
3 Mori Y et al. Ann Intern Med. 2018;169:357-366.
4 Dreiseitl S et al. Melanoma Res. 2009;19:180-4.
5 Kok MR, Boon ME. Cancer. 1996;78:112-117.
6 Chang PL et al. Med Decis Making. 1999;19:419-427.
7 Nieminen P et al. Int J Cancer. 2003;103:422-426.
8 de Veld DC et al. J Biomed Opt. 2004;9:940-950.
9 Lucidarme O et al. Eur Radiol. 2010;20:1822-1830.
10 Fink C et al. J Dtsch Dermatol Ges. 2017;15:414-419.
11 Walker BN et al. EBioMedicine. 2019;40:176-183.
12 Wang P et al. Gut. 2019;68:1813-1819.
13 Su JR et al. Gastrointest Endosc. 2020;91:415-424.e4.
14 Li L et al. Thorac Cancer. 2019;10:183-192.
15 Hollon TC et al. Nat Med. 2020;26:52-58.
16 Wang P et al. Gastroenterology. 2020;159:1252-1261.e5.
17 Repici A et al. Gastroenterology. 2020;159:512-520.e7.
18 Gong D et al. Lancet Gastroenterol Hepatol. 2020;5:352-61
19 Wang P et al. Lancet Gastroenterol Hepatol. 2020;5:343-51
20 Liu WN et al. Saudi J Gastroenterol. 2020;26:13-19.

Poster presented at ISPOR Europe 2022, 6-9 November 2022, Vienna, Austria