Peer Review File

Article Information:

<mark>Reviewer A</mark>

Comment 1: Authors should address implications for other types of cancer such as skin lesions PMID: 36619609 and implications for melanoma PMID: 36983983 in order to aid earlier detection.

If the above are addressed and references included, paper could be of interest.

Reply 1: Thank you for the insight on the relevance of other skin cancers to our manuscript. The authors have reviewed the papers, PMID: 36619609 and PMID: 36983983, and believe they are a great fit for our manuscript. The authors have updated the manuscript to include the importance of early detection for skin cancers such as melanoma and the implications of racial biases present in dermatological predictive models. The update can be seen on page 5, lines 107-112. The numerical references for PMID: 36619609 and PMID: 36983983 are 27 and 29, respectively.

Changes in the text 1: The modified text reads "For example, ML provides opportunities for early detection of skin cancers such as melanoma, which allow 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)".

<mark>Reviewer B</mark>

Comment 1: Abstract, line 42. Recommend changing "more popular" to "more common".

Reply 1: The authors thank you for your recommendations. Your suggestions have been valuable in improving the understanding and coherence of our manuscript. The authors have updated the manuscript to change "more popular" to "more common" and the update can be seen on page 2, line 42.

Changes in the text 1: The sentence now reads "However, as clinical applications of ML become more common, there is concern about the prevalence…"

Comment 2: Abstract, line 43. Recommend changing to "improper algorithm training".

Reply 2: The authors have updated the manuscript to change "improper training on imbalanced datasets" to "improper algorithm training" and the update can be seen on page 2, line 43.

Changes in the text 2: The full sentence now reads "However, as clinical applications of ML become more common, there is concern about the prevalence of ethnoracial biases due to improper algorithm training."

Comment 3: Abstract, line 49. Recommend changing to "two researchers independently reviewed 115 articles and…"

Reply 3: The authors have updated the manuscript to change "Literature search was conducted by two independent researchers" to "Two researchers independently reviewed 115 articles" and the update can be seen on page 2, line 49.

Changes in the text 3: The full sentence now reads "Two researchers independently reviewed 115 articles and evaluated them for incorporation and inclusion of demographic information in the algorithm."

Comment 4: Highlight box, second to last line. Recommend changing "other datasets" to "other populations".

Reply 4: The authors have updated the manuscript to change "other datasets" to "other populations" and the update can be seen on the second to last line within the highlight box on page 3.

Changes in the text 4: The full sentence now reads "AI built on datasets that do not consider race may be built entirely on homogeneous patient populations that do not generalize well to other populations."

Comment 5: Page 4, line 84. Recommend changing "grown in popularity" to "become more common".

Reply 5: The authors have updated the manuscript to change "grown in popularity" to "become more common" and the update can be seen on page 4, line 84.

Changes in the text 5: The full sentence now reads "However, as ML has become more common, so has more attention been directed to the impact of racial biases due to imbalanced datasets."

Comment 6: Page 5, lines 106-108. I don't see the relevance of COVID here. Certain populations had lower screening rates during COVID, but I don't see how that fact would or should influence ML beyond the race/ethnicity being included in the training dataset. Recommend deleting.

Reply 6: We appreciate the reviewer for bringing this oversight in relevance to our attention. The authors have updated the manuscript and the sentence, originally on page 5, lines 106-108, has been deleted.

Changes in the text 6: The sentence has been deleted from the manuscript.

Comment 7: Page 5, line 116. Recommend changing "targets" to "targeted".

Reply 7: The authors have updated the manuscript to change "targets" to "targeted" and the update can be seen on page 5, line 116.

Changes in the text 7: The sentence now reads " It is important to also consider if algorithms' targeted at cancer prediction…"

Comment 8: Page 5, line 117. Recommend changing to "…account different racial/ethnic groups present in their training samples and the adequacy of those samples".

Reply 8: The authors have updated the manuscript to change "account from the population of different racial/ethnic groups present in their training samples." to "account different racial/ethnic groups present in their training samples and the adequacy of those samples". The update can be seen on page 5, lines 116-117.

Changes in the text 8: The full sentence now reads "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."

Comment 9: Page 6, lines 128-129. Variation is naturally solved after agreement. Please restate.

Reply 9: Thank you for bringing this redundancy to our attention. The authors have restated the phrase "variation was solved after agreement" to "any variations were subsequently resolved" to address this. The update can be seen on page 6, lines 128-129.

Changes in the text 9: The full sentence now reads "Literature search was conducted by two independent researchers and any variations were subsequently resolved."

Comment 10: Page 7, lines 149-150. Recommend changing to "…impact of at least one demographic variable on the algorithm…"

Reply 10: The authors have changed "…impact of a piece of demographic information on the algorithm performance…" to ""…impact of at least one demographic variable on the algorithm…". The update can be seen on page 7, lines 150-151.

Changes in the text 10: The full sentence now reads "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)."

Comment 11: Pages 8-9. The numerical references need to be added to the manuscript rather than just "Yala et al." for example.

Reply 11: We thank the reviewer for drawing attention to this reference oversight. The authors have added the numerical references of the three studies mentioned on pages 8-9. Numerical

references for "Yala et al.", "Beig et al.", and "Schaffter et al." were added on lines 183, 191, and 193 and numbered 31, 32, and 33, respectively.

Changes in the text 11: The full sentences, including the numerical references, are below:

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 (31).

In contrast, Beig et al. found that race was not a statistically significant factor in algorithm performance ($P = 0.97$) (32).

Additionally, Schaffter et al. did not report any statistical testing of race on algorithm performance (33).

Comment 12: Table 2. I think it would be useful to also report in this table the number of studies by cancer site-modality combination. For example, while there are 26 studies for breast overall, what are the numbers from mammogram, MRI, etc.?

Reply 12: The authors appreciate your suggestion to modify table 2 and agree upon the benefits for reporting the number of studies by cancer-site modality combination. By providing the specific number of each modality by cancer site, it provides greater clarity to the readers for interpretation. The modified table can be seen on page 15, Table 2.

Changes in the text 12: The modified table seen on the following page.

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

Total Studies			84				
Cancer Site			**Imaging Method**				**Total**
	Computed Tomography (CT)	Mammogram	Magnetic Resonance Imaging (MRI)	Shear Wave Elasticity Images	Ultrasound	X-Ray	
Breast	1	8	6	0	9	2	26
CNS (Brain, Spine)	0	0	8	0	0	0	8
GI	4	0	4	1	0	0	9
GU (Prostate, Bladder, Kidney, Adrenal)	5	0	6	0	0	0	11
Gyn (Ovaries, Uterus)	0	0	9	0	0	0	9

Head and Neck (Lymph Thyroid)	1	0	0	0	5	1	7
Sarcoma (Soft Tissue, Fatty Tissue)	1	1	2	0	0	0	4
Thorax (Lungs, Chest)	7	0	0	0	0	3	10