"Enhancing Medical-Imaging Artificial Intelligence through Holistic Use of Time-Tested Key Imaging and Clinical Parameters: Future Insights": Response to Reviewers (manuscript 75227)

(CONFIDENTIAL – for Editor-in-Chief - first)

We thank the reviewers for their valuable inputs. In the text below, we show a Reviewer's suggestion in **Bold**, and our response in plain text.

Briefly, we agreed with and accommodated all of the suggestions of Reviewers 1 and 3 and the Editorial Staff. For illustrations, we have included a detailed Table 1, which contrasts Symbolic AI, Statistical Learning and Deep Learning, at the end of the paper and a Figure (Tension Pneumothorax on CT- Fig. 1).

Reviewer 2's suggestions were considerably more problematic, however, and we spotted an issue that raises a serious concern about reviewer misconduct.

Reviewer 2 had provided a list of references that should be cited. (We supply this list later.) **All the references that were suggested by Reviewer 2 have the same first author, Mohammad Behdad Jamshidi.**

Apart from clearly divulging the reviewer's identity to us (the reviewer is either Dr. Jamshidi or someone from his group who is a co-author on all these papers), this action raises a serious ethical issue. It suggests a blatant attempt to pressure authors to cite the reviewer's work (thereby increasing the reviewer's citation metrics), irrespective of whether the said papers are relevant to the paper under review, or even to the journal to which the paper has been submitted. (Some of the papers deal with AI applications in areas completely unrelated to medicine, like semiconductor technology/materials science and physics.)

Dr. Nadkarni, first author, has been on the Editorial Board of several journals and continues to be an Associate Editor of the Journal of the American Medical Informatics Association (JAMIA). https://academic.oup.com/jamia/pages/Editorial_Board . He has played this role since 2005.

At JAMIA, Dr. Nadkarni's role involves writing triage reviews and selecting reviewers for those submissions that pass triage. (Last year, he handled 95 submissions, and wrote about 80 triage reviews with an average length of slightly more than 2 single-spaced pages per review).

JAMIA policy is that a reviewer caught pulling this well-known stunt - which has been tried too many times to enumerate - is blacklisted from ever again reviewing for JAMIA.

We leave the appropriate course of action to the Editor-in-Chief: we have simply brought this issue to their notice.

Reviewer #1:

Conclusion: Minor revision

Specific Comments to Authors: …My evaluation is that the paper is publishable with minor revisions.

We agree with all of Reviewer 1's suggestions. In particular, the suggested articles are all very useful: they brought up very important issues that needed to be discussed in depth, and necessitated revision of the paper's structure, with some expansion in content. The outline of the revised paper is reproduced below.

Major points:

1) Introduction: there is discrepancy seen in the abstract and core tip. In the abstract, there is a sentence stating that the AI applications to be trusted to be bias-free. However, the authors stated that AI algorithm can be biased in the core tip. Please reconsider this text.

Our phrasing was ambiguous: we apologize for the confusion. We have rephrased that sentence thus:

"Further, unless an AI application is explainable, radiologists will not trust it to be either reliable or bias-free; we discuss regulatory concerns regarding explainability ("transparency") and some approaches aimed at explaining such applications better."

2) … The disadvantages of the machine learning/deep-learning and the strategy to overcome these problems also should be described. Please reference the following articles.

1) **The "Black-box" prediction (Loyola-Gonzalez et al. IEEE access. 2019 7: 154096-154113)**

We have introduced this article in Section 4.1 ", "black-box" vs "white-box" and follow the author's lead in stating that explainability depends not just on the methodology but also a given model's complexity and whether the input data needs to be transformed algorithmically into a form unfamiliar to the human expert. We also cite it in the new section 4.3 "Making Black-box AI more explainable".

a. **Batch normalization/dropout: Li et al IEEE/CVF conference 2019**

To introduce Li's article, we had to introduce the two main concepts in the paper first. We introduce Batch Normalization (BN) early in section 1.1, where we introduce DNNs, because without BN, DNNs with a large number of layers are infeasible. We introduce dropout in a new subsection 1.2.2, "Sources of Error", noting that dropout is an approach to regularization to prevent overfitting. We then introduce Li's discussion of BN vs dropout and summarize their recommendations.

b. **AI early warning score (xAIEWS) system (Lauritsen et al. Nat Commun. 2020 Jul 31;11(1):3852).**

We introduce Deep Taylor Decomposition, the approach used by this paper, and the paper itself, in the new Section 4.3, "Making black-box AI more explainable". We note that DTD is unwieldy when the features get very numerous or are discovered from the raw data (as in image processing by most DNNs).

c. **Double-descent (Nakkarian et al. ICLR 2020; Nakkarian et al. arXiv. 2019: 1912.02292)**

We introduce this paper in Section 1.1.2. ("Sources of error"). A subsequently published paper (Oakden-Rayner et al 2020, cited 146 times) introduced the concept of "hidden stratification", which explains the double-descent phenomenon elegantly (we believe), so we have put both together.

2) **Domain shift**

a. **① Domain adaptation (Choudhary et al. Yearb Med Inform. 2020 Aug;29(1):129-138.)**

We introduce this paper in Section 1.1, "Neural networks: Deep Learning" As Choudhary et al do, we identify Domain adaptation as a special case of Transfer Learning, where the previously trained DNN can be used as-is or tweaked with an accelerated training phase for better performance.

3) **3) Clinical application (Guidelines and legislation)**

a. **Good machine learning practices, Algorithm change protocol, General Data Protection Regulation (Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. 2021 Jan; U.S Food and Drug Administration)**

b. **④ Proposal for a Regulation laying down harmonized rules on artificial intelligence (Proposal for a Regulation laying down harmonized rules on artificial intelligence. 2021; European commission)**

These have been introduced in a new subsection 4.4 "Regulatory Concerns". We introduce the concept of SaMD by pointing out that FDA-approved EKG-diagnosis software runs on

smartphones, and additionally note that Human Rights Watch has criticized the EC plan (which has much wider scope than the FDA's plan) as being insufficiently protective of the rights of individuals at the economic fringes of society, and we have cited them as well.

Minor points: 1) Page 5; the title of "Statistical supervised learning techniques" should be underlined.
Done.

Reviewer #2:

Conclusion: Major revision

We believe that the reviewer reading of the paper has been superficial. Our paper is partly a review (plus tutorial) of Artificial Intelligence *methodologies* used in medical imaging, and our advocacy for an integrated approach. Instead, the reviewer has presumed that the paper is a review of other people's work applying AI in medical imaging (and beyond), from point #3 onward.

Major Concerns:

1.In order to help readers understand the methods verified in this research, it is necessary to add some tables, charts, and graphical works, showing the understudied papers.
The Editorial Staff also had the suggestion to include graphical elements to assist the reader's perusal. Table 1 contrasts symbolic, statistical, and deep learning methodologies.

2.The contribution of the paper is ambiguous. The authors need to explain how this research can be used for developing AI in cardiac telerehabilitation.

The Reviewer's choice of cardiac telerehabilitation is strange, given that our paper is focused on Imaging applications, as is this journal itself. A literature review showed that cardiac rehab hardly ever, if at all, involves imaging (and tele-rehab not at all, for obvious reasons). Even for in-person rehab, imaging use is limited by cost, plus modest patient hazard for angiographic procedures-unless an adverse event occurs to the patient that would require imaging anyway. Most AI is focused on customizing exercise programs to the patient's rehab status. Some sample citations follow:

- PMID 33860679, "Harnessing artificial intelligence in cardiac rehabilitation, a systematic review" focuses on three scenarios: wearable devices, diagnostic and home monitoring devices, and providing decision support to clinicians.
- "Artificial Intelligence-Based Cardiac Rehabilitation Therapy Exercise Recommendation System", ieeexplore.ieee.org/document/9437568 and PMC7451486 both compare AI-assisted customized exercise therapy vs "usual" or randomly assigned exercise therapy.
- PMID 34217385 "Cardiac rehabilitation in the digital era" (Lancet 2021) and PMID: 28813954 "Human-robot sensor interface for cardiac rehabilitation" both discuss leading-edge innovations. The former advocates use of sensors, plus history and clinical parameters**. (In effect, advocating a holistic approach similar to ours but toward a different goal**). The latter describes a trial of-robot sensor interface to increase patient motivation by providing feedback.

Again, we wish to keep this review focused on imaging applications.

3.Reliability, accuracy, and features of each method or group of methodologies need to be discussed comprehensively. Moreover, the effectiveness of these studies should be explained.

It is unclear what part of the paper the reviewer is referring to. They could be referring to an entire family of methodologies, e.g., statistical AI, symbolic AI, Deep learning - or to the extremely brief list of individual studies cited in references 21 through 39, listed in the first part of "The Need for a Holistic, System-based Approach". (In the revised paper, we have added a sentence stating that this list is not intended to be comprehensive.) We therefore address both possibilities here:

A. Review of a family of methodologies

Reliability, accuracy, and features are attributes of individual applications of a methodology, not of the methodology itself. AI is still an art rather than a pure science; that is why its quality varies so greatly. For example, the symbolic AI program DENDRAL, used for interpreting Mass Spectroscopy data, was high quality AI, while the much-despised "Clippy" feature of Windows 95 exemplified Artificial Stupidity.

The proper application of AI requires in-depth domain knowledge (or collaboration with experts with such knowledge). In statistical AI, for example, the judicious choice of features is critical. So, a symbolic or statistical AI algorithm with a thoughtfully chosen feature set and expert input may outperform a convolutional neural network that has to rediscover, from lots of raw data, well-established "expert knowledge" known for decades.

In this paper, and especially section 2, we have emphasized that no single approach is optimal under all circumstances, and that all available information must be used.

B. Review of individual studies, refs 21-39

This paper is NOT a review of every single-purpose AI application ever developed to assist in diagnoses that require medical imaging. We mention a handful of papers only to illustrate a point - namely, that while they may (or may not) do their chosen task well, they do only that task and nothing else. Therefore, for the purpose of this paper:

- We are not concerned whether they perform well or not: we only care that their function is limited to a single task. (Imagine a radiologist who could only diagnose rib fractures.) Therefore, their accuracy, AUROC, etc., are not relevant for our purposes.
- We are not concerned about what features they used. Note: Deep Learning, when it uses the images themselves as input, is less dependent on feature selection by experts- increasingly complex features are discovered progressively by each successive network layer - but accurate labeling of a sufficiently large number of images is necessary, including identification of subclasses, as discussed in our response to Reviewer 1.
- In any case, readers desirous of the details of a particular study can look up its bibliographic citation.

4. Author should show the differences and similarities of the current review with other ones with a focus on strengths and weaknesses.

We believe that our paper is the first of its kind in the field of Medical Imaging in advocating the viewpoint of a holistic approach to AI that combines multiple methodologies.

As stated above, we are not reviewing individual AI applications for imaging-based diagnosis, and therefore a comparison with review papers addressing these would not serve a useful purpose. Obviously, they are intended to be comprehensive, while our list is not: please refer above.

5. Future work should be extended about each group of the understudied method.

While there are no guarantees about what the future holds, the latter half of the paper, particularly sections 2 and 5, addresses these issues. While innovation in symbolic AI appears to have plateaued, statistical and DL methods continue to be actively developed.

6. The literature review is too poor. It is highly recommended to improve it. In this respect, the applications of AI in different fields of study, such as engineering, economy, and medicine need to be explained via the following references in each section.

1.1. Neural Networks (NNs): Deep Learning

a. Artificial Intelligence and COVID-19: Deep Learning Approaches for Diagnosis and Treatment DOI: 10.1109/ACCESS.2020.3001973

b. Deep Learning Techniques and COVID-19 Drug Discovery: Fundamentals, State-of-the-Art and Future Directions: DOI: 10.1007/978-3-030-67716-9_2

c. A Conceptual Deep Learning Framework for COVID-19 Drug Discovery DOI: 10.1109/UEMCON53757.2021.9666715

d. Deep Learning Techniques for Model Reference Adaptive Control and Identification of Complex Systems DOI: 10.1109/ME49197.2020.9286698

1.2. Training in Machine Learning:

a. Hybrid Machine Learning Techniques and Computational Mechanics: Estimating the Dynamic Behavior of Oxide Precipitation Hardened Steel DOI: 10.1109/ACCESS.2021.3129454

b.Using an ANN Approach to Estimate Output Power and PAE of A Modified Class-F Power Amplifier DOI: 10.23919/AE49394.2020.9232787

c. A novel neural-based approach for design of microstrip filters, DOI: 10.1016/j.aeue.2019.152847

d. Design and modeling of a compact power divider with squared resonators using artificial intelligence DOI: 10.1007/s11277-020-07960-5

e. Neuro-Fuzzy Approaches to Estimating Thermal Overstress Behavior of IGBTs, DOI: 10.1109/PEMC48073.2021.9432584

f. An ANFIS Approach to Modeling a Small Satellite Power Source of NASA, DOI: 10.1109/ICNSC.2019.8743333

NOTE: We raised the concern to the Editor-in-Chief that all the above papers have the same first author, M.B. Jamshidi, earlier.

With due respect, we disagree with this suggestion on the grounds of practicality: there are so many applications of AI that, even if we limited ourselves to Deep Learning, this review would be the length of a book.

Also, since the present journal happens to be “Artificial Intelligence in Medical Imaging”, we presume that the journal’s target reader is fully aware that AI has been applied in numerous and diverse domains, if only via newspapers and other mass-media content.
However, such a reader is unlikely to understand even the basics of the problem domains of the second set of references, which deal with applications in electronics, materials science, physics, and satellite technology. Such knowledge of the basics is necessary to appreciate what problem/s these papers are tackling.

While COVID-19 is topical, two of the three citations are related to Drug Discovery (cheminformatics), which is out of scope for this journal. Of note, there appears to be some redundant publication: one citation is a book section, the other a paper, but there is significant content overlap.

The only citation that is relevant to this paper is 1.1.a, and we have cited it.

7. Trustworthy of AI in medical applications should be discussed

Reviewer 1 made a similar point, additionally suggesting some references, including proposed regulatory frameworks. We have included and discussed these in Sections 3 (unchanged from the previous version) and Section 4, which has been greatly expanded: see the response to Reviewer 1.

8. Title is not correct and it needs to change from “AI mini-review – Enhancing Medical-Imaging Artificial-Intelligence (MIAI) through Holistic use of time-tested key imaging and clinical parameters & MIAI Future Insights” to “Artificial Intelligence Medical Imaging Empowered by time-tested key imaging and clinical parameters”

The Editorial staff also had concerns that the original title was too long and unwieldy. However, Reviewer 2’s suggested title is not grammatical “Artificial Intelligence Medical Imaging” is run-on. (Interpolating the word “-assisted” or “-enhanced” as word 3 would make it so.)

Our modified title is:

“Enhancing Medical-Imaging Artificial Intelligence through Holistic Use of Time-Tested Key Imaging and Clinical Parameters: Future Insights”

9. Abstract should be clearly rewritten.

We have followed Reviewer 1’s suggestions in the matter: we have additionally referred to content suggested by Reviewer 1 in the revised Abstract.

Reviewer #3:

Scientific Quality: Grade B (Very good)

Language Quality: Grade B (Minor language polishing)

Conclusion: Accept (General priority)

Specific Comments to Authors: The mini-review manuscript is very well written and meet all the criteria checklist for publication in this journal. Only few minor modifications are needed:

- **On page1, Abstract, define DL.**
  We have removed the abbreviation since it is not reused in the Abstract.
- **Introduction section, part 1a, give reference for BRS**

  We have supplied a reference - Ronald Ross's authoritative book on the same. (Ross also has a BRS Newsletter to which one can subscribe freely.)
- **Properly format "Justify" the text of the manuscript.**
  Done.