



Ten guidelines for contributors to medical artificial intelligence research

Dohyoung Rim*

Rowan Co., Seoul, Korea

To all esteemed readers and contributors of the *Ewha Medical Journal*,

The number of medical papers incorporating deep learning technologies has increased significantly in recent years. As a result, medical journals now require specialized editorial expertise to properly evaluate these developments. Beyond validating the medical value, efficacy, and research outcomes, it has become necessary to verify the technical aspects of deep learning research methods and the analysis of results. *Ewha Medical Journal* (EMJ) is a highly progressive publication that actively supports and encourages medical research utilizing artificial intelligence (AI). Recognizing the importance of this specialized AI role, we have established the position of AI Editor, which is a role that remains rare globally. I joined EMJ this year as the AI article editor and have recently reviewed and guided several submissions. This letter aims to provide a general guide, from an engineering perspective, for preparing medical research papers that incorporate artificial intelligence. It is designed to be easily understandable, though it is neither exhaustive nor intended as a master-level resource, nor does it function as a checklist. For a reliable, comprehensive guideline, we recommend consulting the TRIPOD (transparent reporting of a multivariable model for individual prognosis or diagnosis)-AI statement [1,2].

Manuscript format: imbue the medical manuscript with engineering depth

Medical and engineering papers differ markedly in their formats, with notable distinctions in abstract structure (structured

versus unstructured) and the organization of chapters. From an engineering standpoint, the overall format of an AI-driven research paper in medicine is not the primary concern; adherence to the existing format of medical journals is generally sufficient. However, certain elements deemed essential from an engineering perspective must be thoroughly addressed in the manuscript.

Reproducibility: the lifeblood of experimentation, demonstrate it transparently

Medical research employing artificial intelligence typically involves experiments using AI models, from which insights are derived through the analysis of experimental results. It is essential that readers are able to reproduce both the experiments and their outcomes based solely on the information published in the paper, even when the code is not provided. To ensure this, unlike in conventional medical papers, the following details must be clearly specified, and a dedicated section may be included for this purpose: data, preprocessing methods, model details, training details, training results, and result analysis.

Data: the foundation of research, describe it accurately and in detail

All data used in the research must be described in detail, regardless of its origin or type. When utilizing publicly available datasets, the means and methods of acquisition must be explicitly stated. For self-collected data, an even more meticulous description is

*Corresponding email: dh-rim@hanmail.net

Received: July 30, 2025 Accepted: July 30, 2025

© 2025 Ewha Womans University College of Medicine and Ewha Medical Research Institute

© This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



necessary. Wherever feasible, self-collected datasets should be made publicly accessible; if this is not possible, they should at least be made available to reviewers for evaluation.

Preprocessing: a core determinant of outcomes, disclose it transparently

Data preprocessing methods have a profound impact on the results and the validity of the experiment. While general processes such as normalization and the handling of missing values are important, the way training, validation, and test datasets are composed is especially critical for the development and validation of AI models. Including any portion of the training data in the test dataset constitutes “data leakage” and renders the experimental results meaningless. Furthermore, preprocessing procedures for all 3 datasets, as well as for any required external test data, must be applied consistently to ensure the results are meaningful.

Model details: the core of the design, describe it for reproducibility

Ideally, the experimental code should be made publicly available. However, if this is not possible, the information provided in the paper should be detailed enough for readers to reconstruct the code, acquire and preprocess the data as described, and reproduce the research results. As with the data, if public disclosure of the code is not feasible, it should at least be accessible to reviewers.

Training details: the importance of environment and settings, record them meticulously

Training details encompass specifics about the environment and hyperparameters used during model training. This includes comprehensive information such as the operating system, central processing unit, graphics processing unit, Python and library versions, as well as detailed hyperparameter settings like learning rate, dropout ratio, optimizer, and its configurations. Even with identical data, preprocessing steps, and model code, the final training results and performance can vary considerably based on these settings.

Training results: visualize the process, present the results with clear metrics

Machine learning, and particularly deep learning, requires a structured training process. During training, loss and performance

metrics must be monitored to determine whether the model has been adequately trained or if overfitting has occurred. Typically, these metrics are presented graphically to illustrate their progression over time. Including loss and metric graphs will greatly aid readers in understanding the training process and the outcomes achieved.

Result analysis: task-appropriate metrics, interpret the findings in depth

AI tasks in medical research broadly encompass regression and classification, as well as segmentation and generation. For regression and classification tasks, essential metrics include mean square error, receiver operating characteristic (ROC) curves, and the area under the ROC curve. For classification tasks, the confusion matrix, precision, sensitivity (also referred to as recall), and F1-score should be presented and discussed.

Comparative models: prove research value and ensure persuasiveness

If the research centers on the model itself, its superiority should be demonstrated by comparison with other models. However, in AI-powered medical research, the primary concern is the utility of the model in a medical context, rather than innovation for its own sake. Consequently, direct quantitative comparisons with other studies may be omitted when standardized experimental conditions cannot be assured. Nevertheless, even in such cases, the experimental results should be presented with sufficient persuasiveness to establish the research's value.

Discussion: bridge with clinical practice, clearly convey value

The ultimate aim of research utilizing artificial intelligence is to harness its potential as a valuable tool in clinical practice. Although the research methods and results may be more aligned with engineering content, it is vital that clinicians and medical researchers who will apply these findings can fully comprehend their significance. Therefore, the Discussion section should articulate the findings and limitations in the context of clinical practice, clearly highlighting the improvements and significance this study brings to the field. By doing so, we bridge the gap between engineering and medicine, effectively conveying the practical value of the research to readers.

Indeed, these technical details may seem somewhat unfamiliar to readers accustomed to traditional medical journal articles. This

is because the writing of AI-based research papers demands far more detailed methodological explanations than traditional medical papers. However, by faithfully incorporating the elements emphasized above, researchers can substantially enhance the reliability and academic value of their work. Furthermore, such transparent reporting increases the likelihood that research findings will be applied in clinical practice, leading to meaningful advancements.

EMJ is at the forefront of this wave of change and actively encourages the participation of medical students and early-career researchers in particular. We will continue to strive to lower the barriers to entry for AI research and to support innovative studies. We hope your valuable research achievements will be widely shared through this journal, ultimately contributing to the advancement of medicine.

Sincerely,

ORCID

Dohyoung Rim: <https://orcid.org/0000-0003-2022-6333>

Authors' contributions

Dohyoung Rim did all the work.

Conflict of interest

Dohyoung Rim has edited *Ewha Medical Journal* since May 2025 as an AI article editor. However, he was not involved in the peer review process or decision-making. Otherwise, no potential conflict of interest relevant to this article was reported.

Funding

None.

Data availability

None.

Acknowledgments

None.

Supplementary materials

None.

References

1. Collins GS, Moons KG, Dhiman P, Riley RD, Beam AL, Van Calster B, Ghassemi M, Liu X, Reitsma JB, van Smeden M, Boulesteix AL, Camaradou JC, Celi LA, Denaxas S, Denniston AK, Glocker B, Golub RM, Harvey H, Heinze G, Hoffman MM, Kengne AP, Lam E, Lee N, Loder EW, Maier-Hein L, Mateen BA, McCradden MD, Oakden-Rayner L, Ordish J, Parnell R, Rose S, Singh K, Wynants L, Logullo P. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385:e078378. <https://doi.org/10.1136/bmj-2023-078378>
2. Collins GS, Moons KG, Dhiman P, Riley RD, Beam AL, Calster BV, Ghassemi M, Liu X, Reitsma JB, Smeden MV, Boulesteix AL, Camaradou JC, Celi LA, Denaxas S, Denniston AK, Glocker B, Golub RM, Harvey H, Heinze G, Hoffman MM, Kengne AP, Lam E, Lee N, Loder EW, Maier-Hein L, Mateen BA, McCradden MD, Oakden-Rayner L, Ordish J, Parnell R, Rose S, Singh K, Wynants L, Logullo P. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods: a Korean translation. *Ewha Med J* 2025;48:e48. <https://doi.org/10.12771/emj.2025.00668>