AICare: An AI–Clinician Interaction System for Transparent and Actionable Clinical Decision Support

Junyi Gao[2,3], Zixiang Wang[1], Dehao Sui[1], Wen Tang[4], Yasha Wang[1], Ling Wang[5], Liantao Ma[1*], Yinghao Zhu[1]
1 Peking University, Beijing, China
2 Centre for Medical Informatics, University of Edinburgh, Edinburgh, United Kingdom
3 Health Data Research UK, United Kingdom
4 Peking University Third Hospital, Beijing, China
5 Affiliated Xuzhou Municipal Hospital of Xuzhou Medical University, Xuzhou, Jiangsu, China
* Corresponding author

Artificial Intelligence (AI) holds substantial promise for augmenting clinical decision-making, yet barriers to interpretability and real-world deployment remain[1]. In this work, we introduce AICare, an AI–clinician interaction system designed to enhance transparency, encourage expert feedback, and ultimately improve the deployment potential of deep learning models in healthcare. AICare integrates four key components (**Figure 1**): (1) visualization of complex patient electronic health record (EHR) data, (2) visualization of model predictions with both granular and longitudinal interpretability, as well as patient clustering and similarity assessments, (3) large language model (LLM)-based summarization for comprehensive clinical reports, and (4) a questionnaire module that captures expert feedback along with fine-grained behavioral data on how clinicians review and interpret patient information. By adopting emerging standards like OMOP for data structuring, AICare can be deployed offline in diverse healthcare environments.

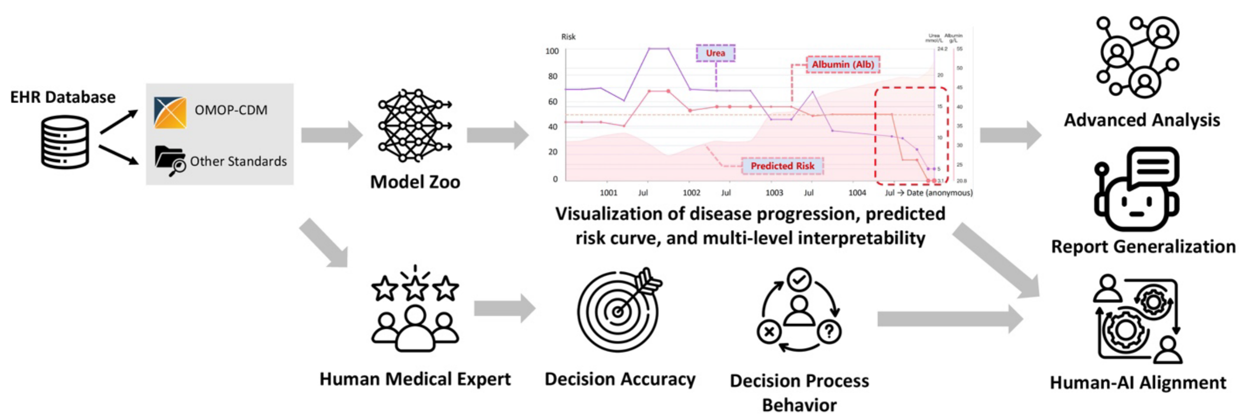

Figure 1 The structure of the proposed AICare pipeline.

One of the primary challenges in clinical AI is effectively visualizing heterogeneous, longitudinal EHR data that span laboratory results, procedures, and clinical events. AICare addresses this by consolidating these elements into an interactive interface that is both intelligible and actionable (**Figure 2**). This design aligns with clinical workflows requiring quick, real-time assessment of large volumes of patient historical data. Building on this visualization framework, we incorporate state-of-the-art deep learning models as well as conventional clinical prediction models that generate risk trajectories such as mortality or readmission risk, while also employing interpretable models[2-4] to highlight the most influential clinical features at both the feature (e.g., vital signs, lab trends) and visit (e.g., episodes of care) levels. By clarifying how individual model features contribute to predictions, we promote transparency and enable clinicians to more confidently calibrate their reliance on AI-generated recommendations.

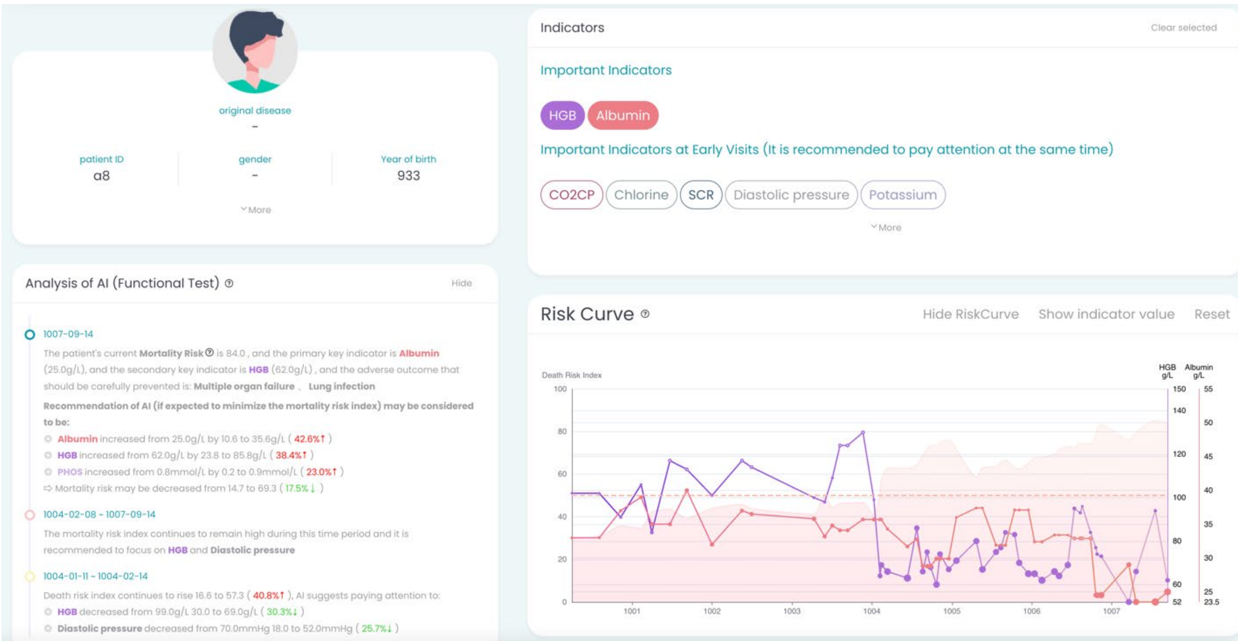

Figure 2 Screenshot of the AICare system. Clinicians can click on clinical features (right top) to show the progression curve. Predicted risk curve and feature importance (shown as different sizes of the dots in the right bottom figure) are visualized in the figure. Patient report is generated by the LLM (left bottom).

AICare goes beyond prediction to provide advanced analysis such as patient clustering based on latent model embeddings, helping clinicians contextualize typical disease progressions and validate a model's risk estimates (**Figure 3**). This capability is complemented by an LLM-based reporting mechanism that synthesizes complex patient data, model outputs, and cluster-level insights into clinically oriented summaries. By bridging the gap between raw machine learning outputs and actionable clinical intelligence, we strive to make AI-driven insights more accessible and meaningful in real-world settings.

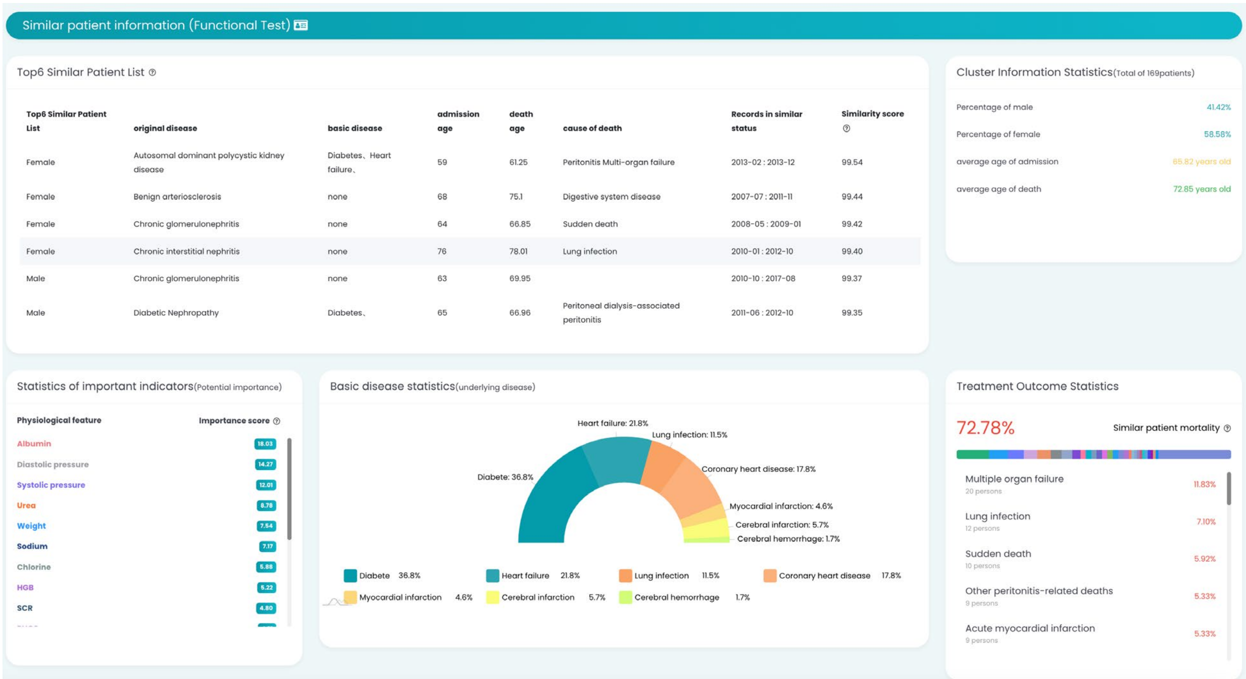

Figure 3 Advanced analysis results generated by AICare. Similar patients, cohort-level disease statistics, outcome statistics and other results are reported.

To further support the robust deployment of AI, AICare includes a questionnaire system that presents the same EHR-derived features the model considers, prompting clinicians to complete identical prediction tasks. By capturing the order and duration of clinicians' feature inspection, we gain insights into the reasoning processes involved in diagnosis. Comparing these human interpretability pathways with the model's rationales highlights areas of divergence or alignment. This dual interpretability framework can guide the refinement of machine learning models through reinforcement learning from human feedback (RLHF), ultimately contributing to the development of more trustworthy, generalizable "medical foundation models."

Overall, AICare aims to advance the deployment of AI models by prioritizing interpretability, feedback capture, and seamless integration into clinical practice. Through rigorous evaluation and iterative refinement, we envision this platform as a critical enabler for safe, transparent and constructive AI solutions in healthcare.

Reference

[1] Van Smeden, M., Heinze, G., Van Calster, B., Asselbergs, F. W., Vardas, P. E., Bruining, N., ... & Moons, K. G. (2022). Critical appraisal of artificial intelligence-based prediction models for cardiovascular disease. European heart journal, 43(31), 2921-2930.
[2] Gao, J., Xiao, C., Wang, Y., Tang, W., Glass, L. M., & Sun, J. (2020, April). Stagenet: Stage-aware neural networks for health risk prediction. In Proceedings of The Web Conference 2020 (pp. 530-540).

[3] Ma, L., Zhang, C., Wang, Y., Ruan, W., Wang, J., Tang, W., ... & Gao, J. (2020, April). Concare: Personalized clinical feature embedding via capturing the healthcare context. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 01, pp. 833-840).
[4] Ma, L., Gao, J., Wang, Y., Zhang, C., Wang, J., Ruan, W., ... & Ma, X. (2020, April). Adacare: Explainable clinical health status representation learning via scale-adaptive feature extraction and recalibration. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 01, pp. 825-832).