

Bridging expertise with machine learning and automated machine learning in clinical medicine

Chien-Chang Lee^{1,2,3} MD, James Yeongjun Park⁴ MSc, Wan-Ting Hsu⁵ MSc

In this issue of the *Annals*, Thirunavukarasu et al.'s systematic review on the clinical performance of automated machine learning (AutoML) highlights its extensive applicability across 22 clinical specialties, showcasing its potential to redefine healthcare by making artificial intelligence (AI) technologies accessible to those without advanced computational skills.¹ This enables the development of effective AI models that could rival or exceed the accuracy of traditional machine learning (ML) approaches and human diagnostic methods.

This editorial discusses the critical role of ML as the foundation of AutoML, and addresses the complexities involved in model development that AutoML aims to simplify. It also examines the challenges associated with integrating AutoML into healthcare, such as performance variability and ethical concerns, emphasising the necessity of a principled approach to its deployment. We also advocate ongoing dialogue among healthcare stakeholders, increased investment in education, and a steadfast commitment to ethical standards, to fully leverage AutoML's transformative potential in patient care.

In recent years, the expansive potential of ML to transform medical care has gained widespread recognition. ML's capabilities, ranging from diagnosing diseases to predicting treatment outcomes, present an unprecedented opportunity to revolutionise healthcare delivery.² The advent of ML in medicine promises a seismic shift in patient care, offering more accurate diagnostics, personalised treatment plans and predictive health outcomes. However, the gap between the expertise required and the available resources in clinical settings has slowed down the adoption of ML in healthcare. This gap is largely caused by intricate programming skills required to develop and implement ML models, which poses a considerable barrier to its broad application in clinical practice. Many healthcare professionals, despite

acknowledging the immense benefits of ML, are deterred by the daunting technical complexities involved in programming and deploying ML algorithms. Watson et al. conducted interviews with academic medical leaders to understand these challenges and recommended solutions, such as building partnership with vendors to overcome them.³

Recognising this challenge, researchers and developers have been working tirelessly to develop solutions that bridge this gap. One such solution gaining traction is AutoML, which is designed to streamline the ML analysis process, allowing users with limited programming skills to harness the power of ML for various applications in healthcare. AutoML software packages offer intuitive interfaces that guide users through the entire ML workflow, from data preprocessing to model selection and evaluation. By automating complex tasks, such as feature engineering, hyperparameter tuning and model selection, AutoML empowers clinicians and researchers to focus on interpreting results and making informed decisions rather than grappling with technical complexities. The beauty of AutoML lies in its democratising effect on ML, making it accessible to a broader audience within the healthcare community. Clinicians, who may not have a background in computer science or statistics, can now leverage ML techniques to derive insights from medical data with relative ease. This democratisation of ML holds the potential to accelerate innovation and improve patient outcomes across various medical domains.

The review by Thirunavukarasu et al., conducted according to a PROSPERO-registered protocol (CRD42022344427), searched the Cochrane Library, Embase, MEDLINE and Scopus databases up to 11 July 2022.¹ Screening of abstracts and full texts, data extraction, and quality assessment were performed by 2 researchers, with disagreements resolved through discussion or third-

The Annals is an open access journal, allowing non-commercial use under CC BY-NC-SA 4.0.

¹ Department of Information Management, Ministry of Health and Welfare, Taiwan

² Department of Emergency Medicine, National Taiwan University Hospital, Taiwan

³ Center of Intelligent Healthcare, National Taiwan University Hospital, Taiwan

⁴ Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, US

⁵ Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, US

Correspondence: Prof Chien-Chang Lee, Department of Information Management, Ministry of Health and Welfare, No.488, Sec. 6, Zhongxiao E. Rd., Nangang Dist., Taipei City 115, Taiwan.

Email: hit3transparency@gmail.com, cclee100@gmail.com

party arbitration. The review included 82 studies featuring 26 distinct autoML platforms, primarily focused on brain and lung diseases among 22 specialties. Variable performance was observed across autoML platforms, with area under the receiver operator characteristic curve (AUCROC) ranging from 0.35 to 1.00, F1-score from 0.16 to 0.99, and area under the precision-recall curve (AUPRC) from 0.51 to 1.00. AutoML demonstrated strong performance metrics in the majority of trials, with AutoPrognosis and Amazon Rekognition emerging as top performers for unstructured and structured data, respectively. However, the quality of reporting was generally poor, with a median Developmental and Exploratory Clinical Investigations of DEcision support systems driven by Artificial Intelligence (DECIDE-AI) score of 14 out of 27. Despite this, autoML shows promise in various clinical contexts, performing comparably to bespoke computational and clinical benchmarks. Future research should focus on enhancing the quality of validation studies, as the integration of autoML with large language models holds potential for advancing user-defined goals in AI-driven healthcare development. Despite the varied performance of these platforms, a significant number of instances highlighted AutoML's capability to either match or surpass the efficacy of bespoke ML models, with significant advantages over traditional ML in detection and diagnosis.⁴ This revelation is groundbreaking, positing AutoML as a potential equaliser in the realm of medical AI, thereby facilitating a more inclusive and efficient approach to patient care that is both personalised and scalable.

Acknowledging the transformative potential of ML in clinical settings, analysis of sepsis patient data from the US National Inpatient Sample by Park et al. serves as a testament to this.² Through the development of models employing advanced computational techniques and benchmarking their performance against the Super Learner model, they have illustrated the power of traditional ML approaches in medical research. This experience, while not directly involving AutoML, highlights the significant impact that ML can have on understanding and treating complex conditions such as sepsis. It also underscores the potential for AutoML to further simplify and expedite these analytical processes in future studies. By automating the more technical aspects of model development, AutoML could make it feasible for a wider range of healthcare professionals to

engage with ML, thereby broadening the scope of its application in clinical research and patient care. Despite not utilising AutoML,² the insights gained lay a foundational understanding of ML's impact, setting the stage for AutoML to potentially streamline similar research endeavours. The variability in AutoML's performance, as noted in the review by Thirunavukarasu et al.,¹ coupled with the need for rigorous validation protocols, presents challenges that must be navigated carefully. Ethical considerations, data privacy and algorithmic bias are additional factors that necessitate a principled approach to AI's⁵ and AutoML's deployment in clinical settings.

It is important to acknowledge that AutoML comes with several inherent limitations. First, it lacks the flexibility to adapt ML algorithms, which may limit its applicability to specific use cases or evolving data landscapes. Second, the rapid pace of ML algorithm development means that AutoML may struggle to incorporate the latest advancements in a timely manner, potentially lagging behind state-of-the-art approaches. Additionally, AutoML may not offer several advanced training methods, such as pre-training, which could restrict its effectiveness in certain complex scenarios. Last, the rapid evolution of large language models has significantly reduced the barrier to entry for ML programming, potentially overshadowing the utility of AutoML in certain contexts. These limitations underscore the importance of considering the specific needs and constraints of each project when deciding whether to employ AutoML or traditional ML approaches.

The journey towards fully integrating AutoML into healthcare is complex and multifaceted, fraught with challenges but brimming with potential. To bridge this gap, we recommend establishing interdisciplinary teams that combine clinical knowledge with ML expertise to oversee the implementation of AutoML solutions, ensuring they are clinically relevant and effectively integrated into patient care workflows. As we stand on the cusp of this new era in healthcare, the promise of AutoML to revolutionise patient care is palpable, contingent upon our collective commitment to harnessing this technology responsibly and effectively. The integration of AutoML into clinical practice, especially within specialties such as neurology and pulmonology, heralds a new epoch of medical treatment, where the confluence of human expertise and AI paves the way for unprecedented advancements in patient care and health outcomes.

Keywords: *AI, artificial intelligence, automated machine learning, autoML, machine learning, medical AI*

REFERENCES

1. Thirunavukarasu AJ, Elangovan K, Gutierrez L, et al. Clinical performance of automated machine learning: A systematic review. *Ann Acad Med Singap* 2024;53:187-207.
2. Park JY, Hsu TC, Hu JR, et al. Predicting sepsis mortality in a population-based national database: Machine learning approach. *J Med Internet Res* 2022;24:e29982.
3. Watson J, Hutyra CA, Clancy SM, et al. Overcoming barriers to the adoption and implementation of predictive modeling and machine learning in clinical care: what can we learn from US academic medical centers? *JAMIA Open* 2020; 10:167-72.
4. Bai A, Si M, Xue P, et al. Artificial intelligence performance in detecting lymphoma from medical imaging: a systematic review and meta-analysis. *BMC Med Inform Decis Mak* 2024;24:13.
5. Busch F, Adams LC, Bressemer KK. Biomedical Ethical Aspects Towards the Implementation of Artificial Intelligence in Medical Education. *Med Sci Educ* 2023;33:1007-12.