

## **PERSPECTIVE**

# Making "CASES" for Al in Medicine

# Ge Wang\*

Biomedical Imaging Center, Center for Computational Innovations, Center for Biotechnology and Interdisciplinary Studies, Department of Biomedical Engineering, School of Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA.

\*Address correspondence to: wangg6@rpi.edu

In this perspective, "CASES" are made for AI in medicine. The CASES mean Confidence, Adaptability, Stability, Explainability, and Security of AI systems. We underline that these CASES can be addressed not only individually but also synergistically on the large model platform and using cutting-edge diffusion-type models.

In recent years, artificial intelligence (AI) has been playing an important role in medicine and healthcare, revolutionizing many things from screening and diagnosis to therapy and prognosis [1]. This paradigm shift is largely due to deep artificial neural networks that analyze vast amounts of diverse data, recognize hidden patterns, and make predictions in a data-driven fashion. Despite its huge potential to go beyond human capabilities, the translation of AI into clinical practice still faces challenges. To address some key challenges conveniently, we propose the acronym "CASES" for Confidence, Adaptability, Stability, Explainability, and Security, discuss associated results, and envision future opportunities.

Confidence (or uncertainty): A key requirement for AI in clinical settings is to give confidence levels in its decisions. Medical professionals need to trust AI's recommendations, which is often hindered by the inherent uncertainty in AI predictions. Uncertainty (aleatoric, epistemic, and numeric) arises due to various factors like data quality, algorithmic limitation, and biological and pathological complexity. While there are rigorous theory and effective methods in classic statistics to build confidence intervals, the gap is well recognized between classic statistics and contemporary AI especially deep learning [2]. To solve this problem, uncertainty quantification of AI systems is being actively researched to provide a confidence score alongside each deep prediction. Dropout [3] (randomly discarding neurons or breaking neural links during training) is the most popular technique to assess the uncertainty of a deep network output, which has been recently extended by us to Flipover [4] (randomly inverting the sign of some scaled neural outputs for adversarial defense). More sophisticated methods include conformal prediction [5] and recently published prediction-powered inference [6].

Adaptability (closely related to generalizability/transfer learning/domain adaptation): Adaptability refers to the ability of an AI system to perform accurately across related but different datasets. Often times, AI models trained on specific datasets may not generalize well, leading to decreased performance in other contexts [7]. This challenge can be tackled through transfer learning and domain adaptation, allowing AI models to avoid short-cut learning and adapt to new, unseen datasets without substantially compromising the performance [8]. To enhance the adaptability of AI systems, they should be trained using well-designed strategies, validated,

**Citation:** Wang G. Making "CASES" for Al in Medicine. *BME Front*. 2024;5:Article 0036. https://doi.org/10.34133/bmef.0036

Submitted 28 December 2023 Accepted 1 January 2024 Published 29 January 2024

Copyright © 2024 Ge Wang Exclusive licensee Suzhou Institute of Biomedical Engineering and Technology, CAS. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY 4.0).

and tested on diverse datasets that include various patient demographics, disease spectrums, and environmental factors [9].

Stability (robustness or reliability): Stability in AI refers to the robustness of its predictions when the input to a deep neural network is slightly changed by natural or adversarial noise [10]. Interestingly, adversarial instability is observed in not only deep image classification but also deep tomographic reconstruction [11]. In medicine, where patients' lives are at stake, it is crucial that AI systems maintain reliable performance by consistently minimizing the first and second types of errors in presence of interferences. Along this direction, we developed an Analytic, Compressive, Iterative and Deep (ACID) approach to address the instability of deep image reconstruction networks [12]. Also, an independent empirical study reported that "the existence of adversarial examples in classification tasks does not always carry over to NN-based solvers for inverse problems. Such reconstruction schemes may achieve state-of-the-art accuracy and can also, in certain cases, exhibit a similar degree of robustness as classical methods." [13] Furthermore, a latest theoretical approach was established with proven convergence and clear interpretability in the case of sparse data tomographic reconstruction [14].

Explainability (also referred to as interpretability): Explainability or interpretability is crucial in healthcare. Medical professionals are more likely to trust and adopt AI solutions that provide transparent and understandable decision-making processes. This means that AI models should not only be accurate but also provide insights about how they arrive at their conclusions. Techniques like feature scores and visualization tools can aid in making AI decisions more interpretable, as surveyed in [15]. However, it was recently noticed that not all interpretation tools are reliable; for example, see [16]. Several interesting schemes were proposed for principled interpretation of deep neural networks such as [17], but it remains open how to completely open the black box of deep learning.

Security (that is, patient privacy and confidentiality of other sensitive information): Security, particularly patient data confidentiality, is of paramount importance in hospitals and clinics. The integration of AI in medicine often involves the handling of sensitive patient data, which must be protected

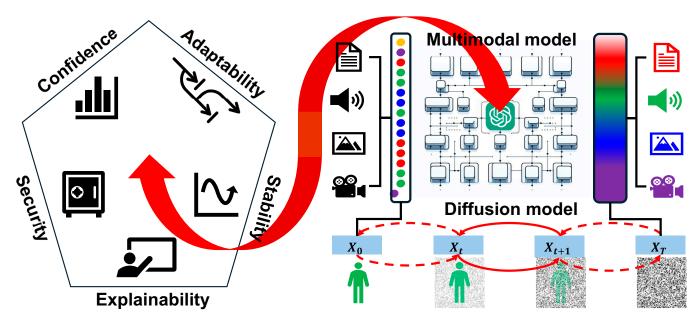


Fig. 1. Confidence, Adaptability, Stability, Explainability, and Security are important for Al in medicine and can be improved using emerging diffusion-type modes and multimodal large models.

from unauthorized access and breaches. This requirement appears contradicting to the need for big data for deep learning such as in a healthcare metaverse [18] and has thus promoted development of secure computing and federated learning [19].

In the above discussions, the CASES are made individually but each of the five issues is intrinsically entangled with the other issues. Hence, it would be optimal to make the CASES synergistically, as shown in Fig. 1. Fortunately, the emerging large language models and multimodal large models in the medical domain seem the ideal platform to address these problems together [20]. These large models, capable of processing and correlating information from diverse data types (like images, text, and genomic data), hold great potential for personalized medicine. In our own study on the first Large Image-Text model to handle volumetric chest computed tomography images, we found that the AI performance on multiple tasks as a whole is significantly better than the separated counterparts on these tasks [21]. Another exciting development in the deep learning field is generative AI via diffusion models that generally outperform the famous generative adversarial networks [22]. While federated learning only shares network parameters, we believe that sharing could be enabled at the raw data level using an enhanced diffusion model embedded with a security safeguard mechanism [23]. Since the diffusion models have a solid theoretical bases (stochastic differential equation, thermodynamics, and electrostatics), their elegant properties facilitate making the CASES effectively but at a more computational cost. To overcome this shortcoming, active research is going on to accelerate the sampling speed of the diffusion models such as using consistency model and Poisson flow generative

Typically, DL stands for deep learning. Now, it is also fine to consider D for diffusion-type models and L for large hybrid models. These and other frontiers are rapidly evolving toward a brighter future of medicine and healthcare. Specifically, the integration of AI into medicine after effectively addressing the CASES issues presents an unprecedented opportunity to

facilitate clinical throughput, improve patient outcomes, and eliminate healthcare disparities.

## **Acknowledgments**

**Funding:** This work was supported in part by the National Institutes of Health under R01CA237267, R01HL151561, R01EB031102, and R01EB032716.

**Competing interests:** The author declares that he has no competing interests.

#### References

- 1. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28(1):31–38.
- 2. Babu GJ, Banks D, Cho H, Han D, Sang H, Wang S. A statistician teaches deep learning. *J Stat Theory Pract*. 2021;15(2):47.
- Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *J Mach Learn Res.* 2014;15(56):1929–1958.
- 4. Liang YX, Niu C, Yan PK, Wang G. Flipover outperforms dropout in deep learning. *Vis Comput Ind Biomed Art*. In Press.
- Fontana M, Zeni G, Vantini S. Conformal prediction: A unified review of theory and new challenges. *Bernoulli*. 2023;29(1):1–23.
- Angelopoulos AN, Bates S, Fannjiang C, Jordan MI, Zrnic T. Prediction-powered inference. *Science*. 2023;382(6671):669–674.
- 7. Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: Limitations and recommendations from an analysis of FDA approvals. *Nat Med*. 2021;27(4):582–584.
- Kouw WM, Loog M. An introduction to domain adaptation and transfer learning. arXiv. 2019. https://doi.org/10.48550/ arXiv.1812.11806
- 9. Chao H, Shan H, Homayounieh F, Singh R, Khera RD, Guo H, Su T, Wang G, Kalra MK, Yan P. Deep learning predicts

- cardiovascular disease risks from lung cancer screening low dose computed tomography. *Nat Commun.* 2021;12(1):2963.
- 10. Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, Fergus R. Intriguing properties of neural networks. arXiv. 2014. https://doi.org/10.48550/arXiv.1312.6199
- 11. Antun V, Renna F, Poon C, Adcock B, Hansen AC. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proc Natl Acad Sci USA*. 2020;117(48):30088–30095.
- 12. Wu W, Hu D, Cong W, Shan H, Wang S, Niu C, Yan P, Yu H, Vardhanabhuti V, Wang G. Stabilizing deep tomographic reconstruction: Part A. Hybrid framework and experimental results. *Patterns*. 2022;3(5):100474.
- Genzel M, Macdonald J, März M. Solving inverse problems with deep neural networks – Robustness included? *IEEE Trans Pattern Anal Mach Intell*. 2023;45(1): 1119–1134.
- Ding C, Zhang Q, Wang G, Ye X, Chen Y. Learned alternating minimization algorithm for dual-domain sparse-view CT reconstruction. arXiv. 2023. https://doi.org/10.48550/ arXiv.2306.02644
- Fan F-L, Xiong J, Li M, Wang G. On interpretability of artificial neural networks: A survey. *IEEE Trans Radiat Plasma Med Sci.* 2021;5(6):741–760.
- 16. Zhang J, Chao H, Dasegowda G, Wang G, Kalra MK, Yan P. Revisiting the trustworthiness of saliency methods in radiology AI. *Radiol Artif Intell*. 2024;6(1):e220221.

- 17. Yu Y, Buchanan S, Pai D, Chu T, Wu Z, Tong S, Bai H, Zhai Y, Haeffele BD, Ma Y. White-box transformers via sparse rate reduction. arXiv. 2023. https://doi.org/10.48550/arXiv.2306.01129
- Wang G, Badal A, Jia X, Maltz JS, Mueller K, Myers KJ, Niu C, Vannier M, Yan P, Yu Z, et al. Development of metaverse for intelligent healthcare. *Nat Mach Intell*. 2022;4(11):922–929.
- 19. Dayan I, Roth HR, Zhong A, Harouni A, Gentili A, Abidin AZ, Liu A, Costa AB, Wood BJ, Tsai C-S, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat Med.* 2021;27(10):1735–1743.
- Liu F, Zhu T, Wu X, Yang B, You C, Wang C, Lu L, Liu Z, Zheng Y, Sun X, et al. A medical multimodal large language model for future pandemics. *Npj Digit Med.* 2023;6(1):226.
- Niu C, Wang G. CT multi-task learning with a large imagetext (LIT) model. arXiv. 2023. https://doi.org/10.48550/ arXiv.2304.02649
- Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis. arXiv. 2021. https://doi.org/10.48550/ arXiv.2105.05233
- Shi Y, Wang G. Conversion of the Mayo LDCT data to synthetic equivalent through the diffusion model for training Denoising networks with a theoretically perfect privacy. arXiv. 2023. https://doi.org/10.48550/arXiv.2301.06604
- 24 Hein D, Holmin S, Szczykutowicz T, Maltz JS, Danielsson M, Wang G, Persson. PPFM: Image denoising in photoncounting CT using single-step posterior sampling Poisson flow generative models. arXiv. 2023. https://doi.org/10.48550/ arXiv.2312.09754