

Fig. 1: Accelerating Patient Care with Innovative Triage: Our novel approach allows medical facilities to rapidly triage a vast number of normal patients, streamlining the prioritization process and greatly enhancing throughput for a more efficient healthcare experience.

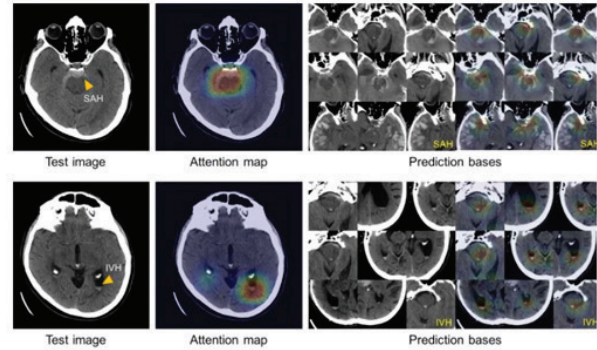


Fig. 2: An AI Solution to Repurpose SafeAI: Employing explainable AI for the detection of acute intracranial hemorrhage using a small dataset (Nature Biomedical Engineering, 2019)

## SafeAI: Live Error-free or Die



### Synho Do, PhD

Director, Lab of Medical Imaging and Computation, Mass General Brigham;  
Assistant Professor, HMS  
[sdo@mgh.harvard.edu](mailto:sdo@mgh.harvard.edu)

Healthcare professionals often care for numerous patients in high-pressure situations, even for brief moments. Even minor mistakes or delays can negatively impact a patient's life and lead to substantial economic losses. As a result, these skilled professionals frequently experience heavy workloads and burnout.

AI technology was expected to help mitigate this issue, but widespread implementation in real clinical environments has yet to be realized. The FDA has approved over 500 healthcare AI algorithms, most of which achieve 95% AUROC or higher or 80% sensitivity/specificity. However, this still leaves room for error, which raises the question: who will detect these errors and who is accountable for them? If AI is to be implemented systematically, its use should not make a doctor's life more difficult.

To address this concern, we have developed SafeAI, an algorithm that "eliminates mistakes while swiftly identifying only normal cases." This doctor-friendly solution can be used independently or as a complementary addition to existing AI applications, providing extensive coverage and excellent scalability.

Our SafeAI is specifically engineered to operate at 100% sensitivity and will not function if this threshold is not met. Therefore, SafeAI can swiftly confirm normal or non-urgent outcomes. In potentially positive or equivocal cases, SafeAI avoids making an inaccurate prediction. Instead, it communicates uncertainty to the user by stating, "I don't know."

This technology is a product of Explainable AI, Zero Error Tolerance AI, and Continuous Learning AI, which have been extensively researched in our laboratory. We published related content in three Nature portfolio journals in 2022: focusing on identifying the sources of prediction uncertainty [1], automatically labeling necessary data in open datasets [2], and repurposing existing AI for new applications [3].

With over 10 IPs secured, we are in the process of commercializing this technology. We are currently seeking funding to submit two algorithms for FDA approval. SafeAI is groundbreaking in that it enables AI to be leveraged to improve diagnostic throughput in healthcare settings for handling 60–80% of cases that are normal or non-urgent while minimizing false negatives.

[1], Chua, M., Kim, D., Choi, J., Lee, N. G., Deshpande, V., Schwab, J., ... & Do, S. (2022). Tackling prediction uncertainty in machine learning for healthcare. *Nature Biomedical Engineering*, 1-8.

[2], Kim, D., Chung, J., Choi, J., Succi, M. D., Conklin, J., Longo, M. G. F., ... & Do, S. (2022). Accurate auto-labeling of chest X-ray images based on quantitative similarity to an explainable AI model. *Nature communications*, 13(1), 1867.

[3], Chung, J., Kim, D., Choi, J., Yune, S., Song, K., Kim, S., ... & Do, S. (2022). Prediction of oxygen requirement in patients with COVID-19 using a pre-trained chest radiograph xAI model: efficient development of auditable risk prediction models via a fine-tuning approach. *Scientific Reports*, 12(1), 21164