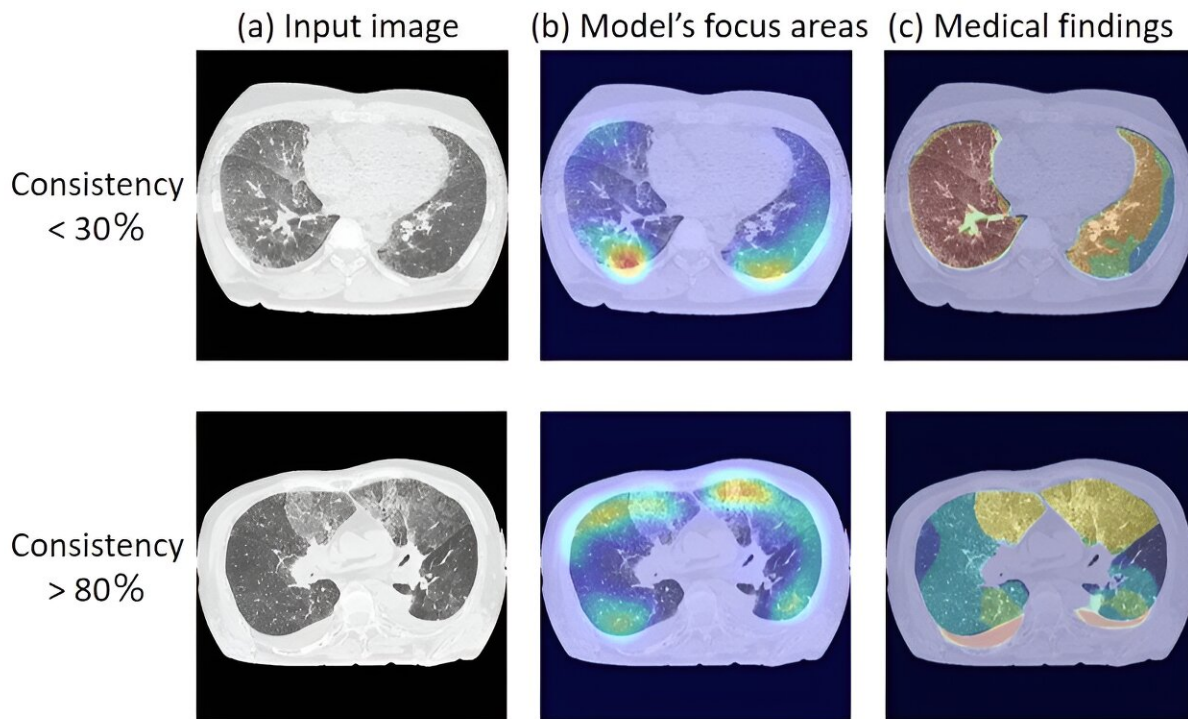


# Team studies medical validity of deep learning models in diagnosing drowning

February 27 2024



Examples showing (a) the input image, with (b) the model's "focus areas" and (c) the "medical findings" projected onto the image. The intensity of the red color indicates a higher contribution to the diagnostic outcome. Among two correctly classified cases, the top case shows less than 30% consistency, while the bottom case shows over 80% consistency; however, significant differences are observed in the medical findings (colored in red). This highlights an unexpectedly large discrepancy between the model's focus and human expertise. Credit: Tohoku University

A research team from Tohoku University Graduate School of Medicine undertook a detailed examination of the medical validity of deep learning models using post-mortem imaging for diagnosing drowning. The results reveal an inconsistency between the deep learning models' results and the medical professionals' observations.

Their [retrospective study](#) was [published](#) in the *Journal of Imaging Informatics in Medicine* on February 9, 2024.

During an autopsy, it is difficult for [medical professionals](#) to reach a diagnosis of drowning. There is no single sign or test to be used to diagnose drowning. Autopsy imaging, such as postmortem computed tomography, can aid in the drowning diagnosis.

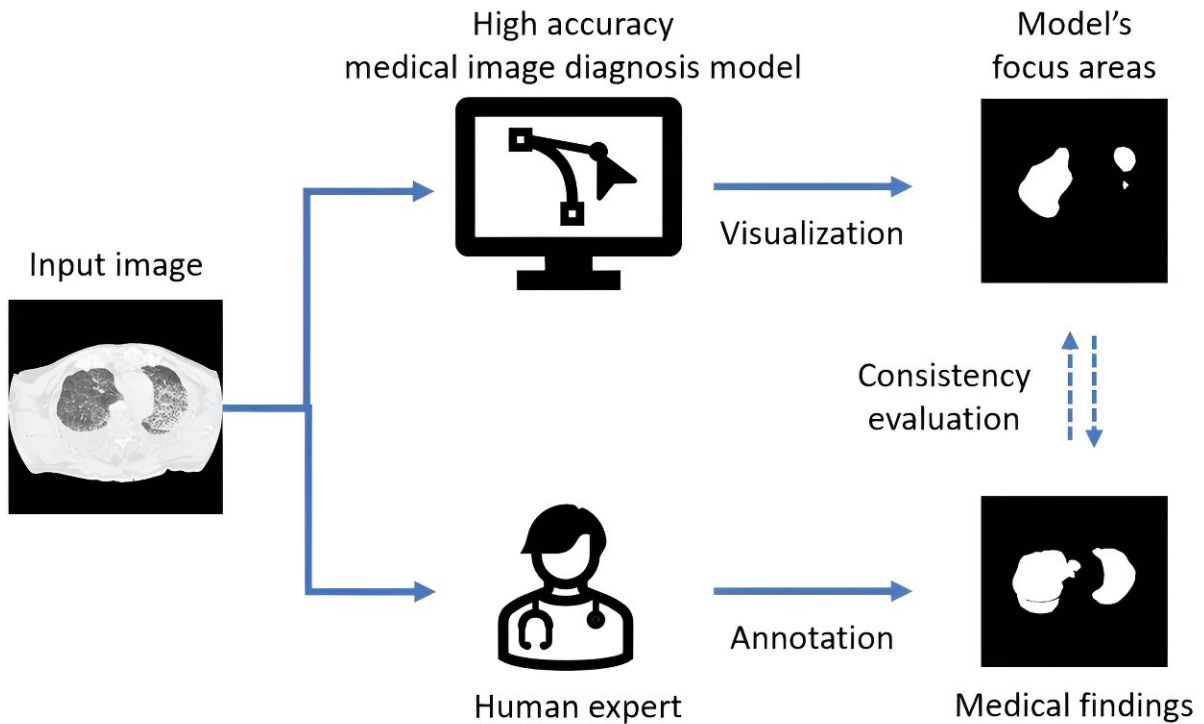
In earlier studies, [deep learning models](#) that perform well have been used for a drowning diagnosis. Some suggest that the deep learning technology performs as well as medical professionals. However, the medical validity of these deep learning models has not been fully tested to determine their accuracy in real-world clinical settings.

For example, no research has been conducted to ensure that the image featured in the deep learning models align with medical findings. This gap raises the potential for [discrepancies](#) between artificial intelligence diagnoses and physician assessments.

So the Tohoku University team undertook a study to assess the medical validity of deep learning models that had achieved high classification performance for drowning diagnosis. Their study included autopsy cases on people aged 8–91 years who underwent postmortem computed tomography between 2012 and 2021. Of these cases, 153 were drowning cases and 160 were non-drowning cases.

The researchers trained three deep learning models from a previous

work and generated saliency maps. The saliency maps highlight the prominent areas where people's eyes focus first. Researchers refer to these areas as the model's focus areas.



To validate the reliability of artificial intelligence in medical image diagnosis, a visualization technique was used to extract the model's "focus areas" and evaluate how well these align with the "medical findings" annotated based on a radiologist's assessments of the same images. Credit: Tohoku University

The team compared the model's focus areas with image regions that were "medical findings" annotated by the radiologists. Their comparisons revealed that in some cases, only 30% to 80% of the model's focus areas were consistent with the critical areas noted by the radiologists. The discrepancies the team uncovered in their study reveal

the importance of adopting different validation methods and demonstrate the challenges in evaluating the performance of artificial intelligence systems.

"The findings underscore the need for new training methods that align artificial intelligence model internals with the complex decision-making processes based on human expertise," said Yuwen Zeng from the Department of Radiological Imaging and Informatics, Tohoku University Graduate School of Medicine, Japan.

Most artificial intelligence systems used for medical diagnosis are based on classification models. These models have achieved high classification accuracy and can provide visual explanation of their predictions. But there is no [quantitative evaluation](#) on whether such visual explanation is valid for medical diagnosis or not.

"This study reported inconsistency between the decision-making basis of deep learning models and medical expertise, raising the concern of evaluating deep learning models to ensure their reliability in real-world medical scenarios. Such awareness is crucial as it may impact the development and deployment of artificial intelligence technologies in health care," said Zeng.

The team faced a unique challenge associated with evaluating the visualization results of deep learning models, particularly in the context of medical images. "Currently, there is no established gold standard for the quantitative assessment of these results, and visual evaluation remains the predominant approach. The existing evaluation methods are primarily designed for natural images, where objects are easily defined," said Zeng.

For medical images, this challenge is further amplified because of the inherent ambiguity in the targets' edges and features. The team's research

showed the need for advancements in evaluation methodologies designed specifically for medical images, taking into account the complexities of the data.

The team notes that because of the hierarchical and top-down architecture of classification models, using a single class label as the only supervision information may prevent the deep learning model from learning all the factors that contribute to the final [diagnosis](#) result.

"Our next step for this project is to introduce additional human expertise into the model to constrain the distribution of model parameters, which could alleviate such inconsistency between model and human expertise. This approach seeks to bridge the gap between the model's decision-making process and the complexity of human expertise," said Zeng.

**More information:** Yuwen Zeng et al, Inconsistency between Human Observation and Deep Learning Models: Assessing Validity of Postmortem Computed Tomography Diagnosis of Drowning, *Journal of Imaging Informatics in Medicine* (2024). [DOI: 10.1007/s10278-024-00974-6](#)

Provided by Tohoku University

Citation: Team studies medical validity of deep learning models in diagnosing drowning (2024, February 27) retrieved 28 April 2024 from <https://medicalxpress.com/news/2024-02-team-medical-validity-deep.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.