Assignment # 1: Paper Review and Summary

Course: CAP 6683 Artificial Intelligence in Medicine and Healthcare

Professor: Dr. Oge Marques

Student Number: Z23596812

Student Name: Renee Raven

Term: Fall 2022

High-performance medicine: the convergence of human and artificial intelligence

The current healthcare system is rife with errors, inefficiencies, inequities, and poor outcomes. Many look to AI to fix all that is broken. And given the apparent success with so many imaging classification tasks (Figure 1), there is a growing perception that AI can be trained to outperform doctors. This promotes an antagonistic relationship between doctors and AI. Ideally, AI should be just another tool in a doctor's bag and the doctors should be the final arbiter of the clinical usefulness of the AI's recommendation, not the AUC metric.

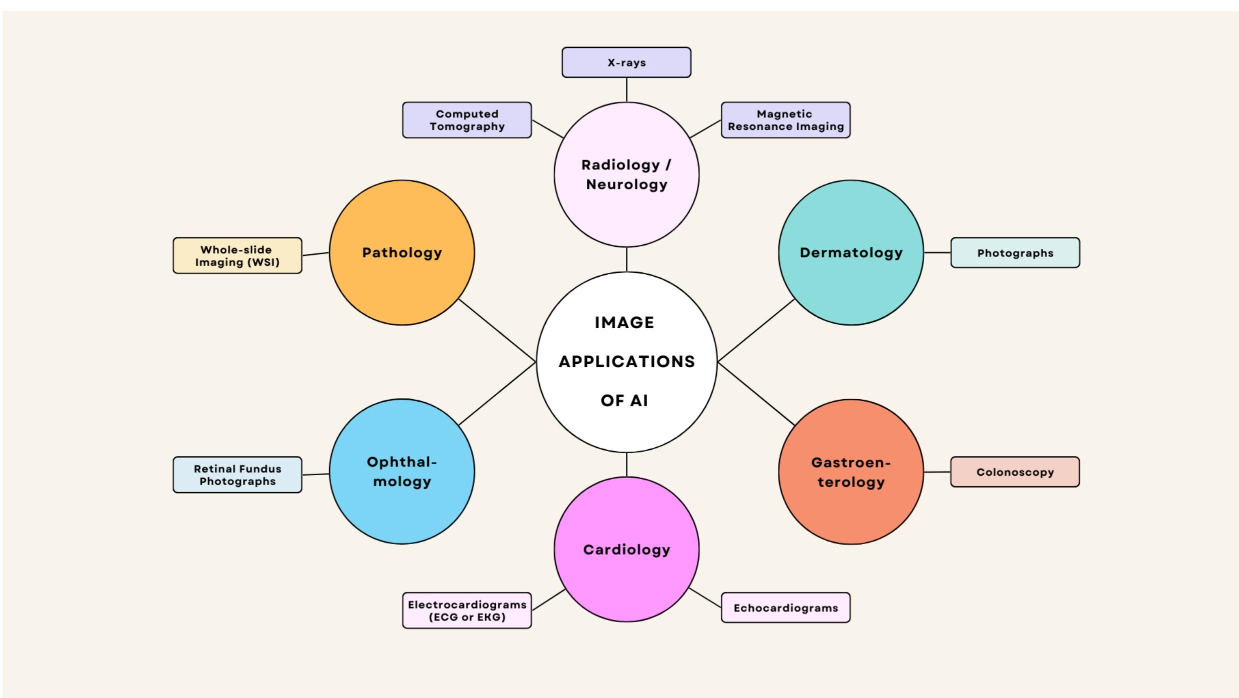

Figure 1. Image applications of artificial intelligence

Dr. Topol offers an overview of a dizzying number of artificial intelligence applications to healthcare and healthcare adjacent domains. Multimodal analysis of vast amounts of patient data may lead to personalized recommendations. Wearable sensors may become the norm and genetic counseling chat-bots could interpret our genomic data.

However, reliance on AI and algorithms without rigorous vetting and continuous validation could lead to systemic errors that cost lives and diminish quality of life, like medication errors. Additionally, bad actors and hackers may infiltrate and alter AI and smart devices to hurt people en masse. AI holds promise, but we must temper our optimism with caution and common sense.

AI in health and medicine

Based on two years of studying advances in medical AI, the authors suggest AI's success with image classification and random control trials (RCTs) has spurred growth and adoption. Areas of growth include drug discovery, medical signal interpretation, genomics, and techniques to extract information from biomedical literature and patient records.

Since much of medical data is unlabeled, multimodal, heterogenous, and/or noisy, it can be costly and cumbersome to prepare the data for conventional supervised learning. Different approaches such as self-supervised learning and clustering algorithms are being utilized to produce actionable results. In fact, sometimes these approaches generate novel insights that might otherwise not have been discovered.

How these insights can be deployed in the clinical setting and how people and AI can work collaboratively. It is important to discover effective and helpful ways to deliver AI insights/recommendations to people. Also, building trust is a formidable challenge. The AI must demonstrate accuracy, reliability, and possibly some level of explainability.

Finally, the authors discuss many ethical challenges such as bias in datasets and algorithms, data use, and security. And then there is the important question of who will be culpable for mistakes? Will it be the software engineer, the doctor, or the hospital administration?

Identifying medical diagnoses and treatable diseases by image-based deep learning

Creating an image prediction model that can be trained quickly and adapted to different types of images in different contexts would provide clinical utility. Building on TensorFlow and transfer learning, Kermany et al. have developed a useful model with validated performance in multiple tests and with 2 types of images: optical coherence tomography (OCT) and pediatric X-rays.

They focused on screening of images for two treatable causes of blindness and then verified the model's performance with X-ray screens for pneumonia. The OCT images were labeled either choroidal neovascularization (CNV), diabetic macular edema (DME), drusen, or normal. CNV and DME mapped to "urgent referral," drusen mapped to "routine referral," and normal mapped to "observation." A false-negative, where the need for an urgent referral is missed, is especially dangerous for the patient. Therefore, a weighted scoring system was utilized to reflect real-world performance of the model.

The results for the full training (over 100,000) images are impressive (Table 1). However, when the model is trained on 1,000 randomly selected images, it still performs fairly well. Additionally, the performance of the transfer learning model when diagnosing pneumonia from X-rays demonstrated a reasonable ROC when trained on only 5232 images.

Table 1. Comparison of results for training with large datasets versus smaller datasets

	Training images	Testing images	Number of final labels	Accuracy	Sensitivity	Specificity	Weighted Error	ROC
OCT Full model	108,312	The same 1,000	3	96.6%	97.8%	97.4%	6.6%	99.9%

OCT Limited model	1,000	images used	3	93.4%	96.6%	94.0%	12.7%	98.8%
Pediatric X-rays	5,232	624	2	92.8%	93.2%	90.1%	NaN	96.8%

The authors also performed binary classification experiments that produced strikingly strong performance (Table 2). It appears when the model is tested with abnormal versus normal OCT images, it is able to identify the correct class with an ROC > 99%. It performs less well discerning the presence of pneumonia on X-rays.

Table 2. Binary classification results

	Image labels		Categories	Accuracy	Sensitivity	Specificity	ROC
OCT	CNV	Normal	“urgent referral” vs “observation”	100.0%	100.0%	100.0%	100.0%
OCT	DME	Normal	“urgent referral” vs “observation”	98.2%	96.8%	99.6%	99.87%
OCT	Drusen	Normal	“routine referral” vs “observation”	99.0%	98.0%	99.2%	99.96%
Pediatric X-rays	Bacterial or viral	Normal	“pneumonia” vs “normal”	92.8%	93.2%	90.1%	96.8%
Pediatric X-rays	Bacterial	Viral	“bacterial” vs “viral”	92.8%	88.6%	90.9%	94.0%

Potential avenues for future performance improvements and clinical applications

- Increasing the number of images used for training the pre-trained models
- Incorporating advances in deep-learning techniques and architectures
- Cross pollination of non-imaging convolutional neural network applications

The authors have provided the python codes and datasets used for this paper (Table 3). They hope the community will develop screening tools that could be used clinically, especially in underserved areas.

Table 3. Information and links for code and datasets

Reproducible code and datasets used in this paper can be found online at data.mendeley		
File	Contents	Size
Code2017.zip	Python Codes and ReadMe file	19 KB
OCT2017.tar.gz	Validated and labeled OCT images	5 GB
ChestXRay2017.zip	Validated and labeled Chest X-Ray images	1 GB

I selected this paper because it offered the source code and the datasets, plus it could have a real impact in the clinical setting. I plan to download the files and experiment to see if I can use transfer learning to get it to work with well with other datasets.

References:

Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C.S., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., et al. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell *172*, 1122-1131.e9. https://doi.org/10.1016/j.cell.2018.02.010.

Rajpurkar, P., Chen, E., Banerjee, O., and Topol, E.J. (2022). AI in health and medicine. Nat Med *28*, 31–38. https://doi.org/10.1038/s41591-021-01614-0.

Topol, E.J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nat Med *25*, 44–56. https://doi.org/10.1038/s41591-018-0300-7.