EVALUATIONS OF AI APPLICATIONS IN HEALTHCARE STUDY GUIDE

MODULE 1: AI IN HEALTHCARE

LEARNING OBJECTIVES

- Recognize why we need AI in healthcare and understand what are the right questions we can ask AI to help solve
- Describe the breadth of AI in healthcare, from the molecular level to the population level
- Recognize the different categories of AI in medicine (biomedical research, translational research and medical practice)
- Explain why AI evaluations need to move beyond model accuracy

AI IN HEALTHCARE

Artificial Intelligence (AI): Involves the development of computer algorithms to perform tasks typically associated with human intelligence

AI spectrum of learning:

- Machine learning
- Representation learning
- Deep learning
- Natural language processing

Algorithms: A mathematical technique, generally developed by statisticians and mathematicians for a particular task, such as unsupervised learning or reinforcement learning

Models: Well-defined computations formed as a result of an algorithm that takes some value, or set of values, as input and produces some value, or set of values as output

AI solution: Refer to an evaluated and validated model

AI has the potential to provide high performance data-driven medicine, optimize care trajectories, suggest the right therapy for the right patient and improve the process of clinical assertions and decision making.

BIOMEDICAL RESEARCH

- Automated experiments
- Automated data collection
- Gene function annotation
- Literature mining

TRANSLATIONAL RESEARCH

- Biomarker discovery
- Drug-target prioritization
- Genetic variant annotation

MEDICAL PRACTICE

- Disease diagnosis
- Treatment selection
- Patient monitoring
- Risk stratification models

The best way to understand AIs potential use in healthcare is to think about its applications in three separate categories:

- Biomedical research: AI is assisting in automated experiments, automated data collection, gene function annotation, literature mining
- Translational research: AI is assisting in areas such as biomarker discovery, drug-target prioritization, and genetic variant annotation
- Medical practice: AI used for disease diagnosis, treatment selection, patient monitoring, and risk stratification models

Reasons why AI is needed in medicine:

- **AI and Data Synthesis-** AI can make sense of the overwhelming amount of data associated with a single disease or a single patient to highlight the relevant information needed to best guide the treatment and care for each individual patient.
- AI can improve **clinical reliability** and be used to help identify relevant information
- AI tools can tackle tedious, mundane tasks because they don't suffer from fatigue, distractions, or moods like their human counterparts. Therefore, AI tools can be useful for **reducing errors related to human fatigue**.
- **Patient-clinician engagement is another important area where we see AI solutions emerging.**
- AI can **improve patient outcomes** by prioritizing patients in more urgent need and by recommending individualized treatments that account for a patient's unique characteristics, which often fall outside of the selective clinical trial evidence.

- AI has the potential to improve resource utilization and efficiency, thereby **reducing overall healthcare costs**.
- AI is being used to **improve diagnostics.** AI can quickly and more accurately spot signs of disease in medical images such as MRIs, CT scans, ultrasounds and x-rays.
- AI can **help physicians understand patients' values and goals.** AI can assist by analyzing structured and unstructured patient data and present insights for physicians' consideration and to aid in shared decision making.

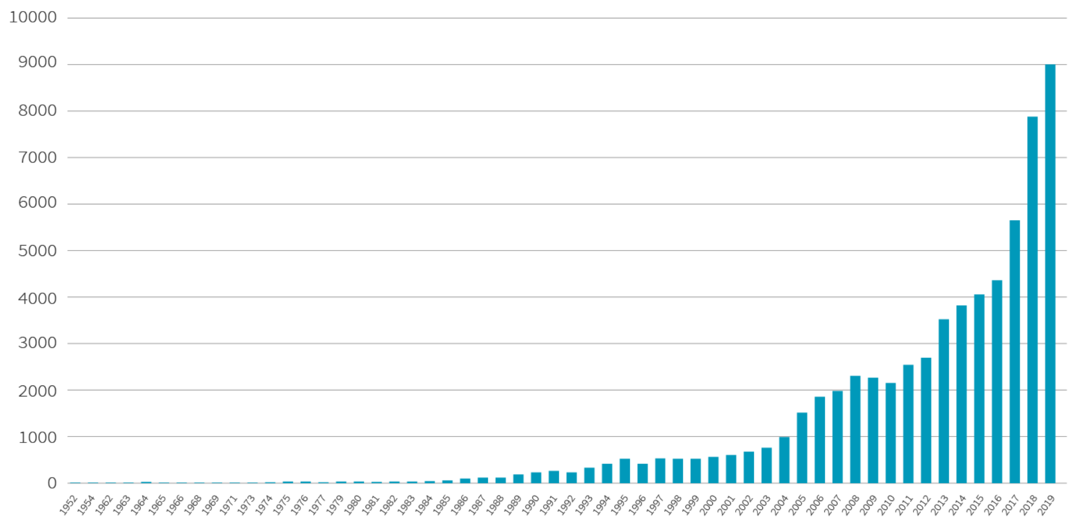
We can think about AI solutions in medicine across the domains of biomedical research, translational research and medical practice.

A non-exhaustive list of some recent AI applications in medicine that demonstrate how AI algorithms can benefit patients, doctors, and researchers through making diagnosis, treatment, discovery, and the practice of medicine faster, more accurate, and more efficient:

- Image analysis - AI is being used in clinical practice for disease diagnosis, AI solutions can reduce errors related to human fatigue and improve diagnostics**.**
- Drug discovery - AI development under the translational research branch. Key classification features are gene essentiality, mRNA expression, DNA copy number, mutation occurrence and protein–protein interaction network topology.
- Patient risk stratification (or Population level segmentation) - AI is used in clinical practice for risk stratification.
- Risk of 30-day readmissions – Another example of AI used in clinical practice for risk stratification.
- Basic biomedical research - AI is being used in basic research for automated experimentation.
- Home videos for autism diagnostics.

These examples provide a glimpse of the broad range of AI applications we see emerging in medicine and the opportunities to improve diagnostics, care delivery, access to care, and patient outcomes.

GROWTH OF AI IN HEALTHCARE



Both literature and media show that there is a hype around AI and this is coupled with an explosion of AI applications in healthcare. Recently, there has been a drastic increasing trend in medical literature related to “artificial intelligence” and “healthcare” or “medicine”.

Most research:

- Provides an overall description of the clinical problem
- Includes the architecture of the proposed model
- Provides the accuracy or precision of the model
- Provides thoughts about how their AI solution COULD be used in the healthcare setting

Think beyond simple accuracy or precision as a measure for AI models because these measures do not tell you anything about the impact of your model in the healthcare setting.

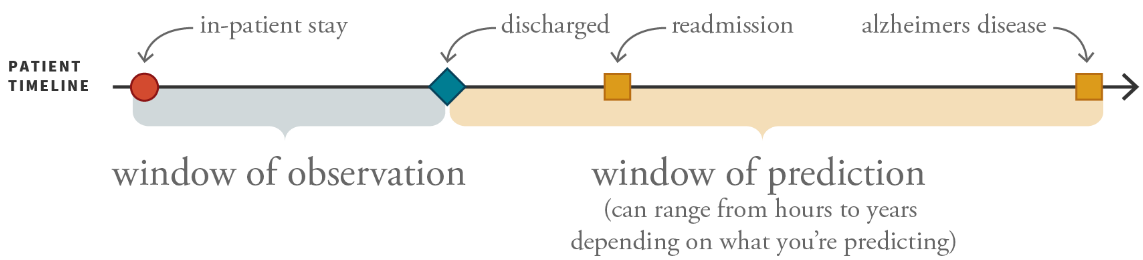
Often researchers focus on a model without considering if the model output can be mitigated or will have an impact on clinical care.

A systematic review identified articles predicting hospital readmission risk:

- Identified about 8,000 citations
- Included 30 studies that met the inclusion criteria
- Most models relied on retrospective administrative data
- Had poor discriminative ability (with a c-statistic less than 0.8)
- Only 1 study specifically addressed preventable readmissions

The ability to predict who is at high risk for hospital readmission may not be clinically useful if the model does not convey anything that the team does not already know. However, if the model contained data on discriminatory actions during the in-patient stay that were associated with low risk of readmission then this information could be taken into account to improve the utility of the AI solution. However, most algorithms are developed without considering the clinical utility, feasibility, or the overall impact of implementing the AI solution.

HOW TO KNOW IF AN AI MODEL IS GOOD



An example of a longitudinal data for a patient – There will be a window of observation, where you are collecting and aggregating data. This observation window could be the inpatient hospital stay. Then you set a time where you will start predicting an event. In our example, this could be time of hospital discharge. You make a prediction and that event can happen within the next hour or in the next N years, depending on what you are predicting. If you are predicting patient deterioration, it may happen within minutes or hours after the observation period. However, if you are predicting Alzheimer's Disease based on genotyping and family history, your event might not occur for years, maybe even decades. So, this gives you your action window. And depending on the prediction, this action window can be categorized as acute or long-term.

One must think about whether the action to my output has to happen immediately, or is it something that's going to happen over the next year, five years, or 10 years? Studies have shown that, regardless of acute or long-term actions, early warning lead times give more opportunities for action.

Interestingly, almost no effort goes into the action side of the equation. You make a prediction; you publish the paper and then you are done.

What we need to think about is how do we move beyond predictions and start thinking about the action side of the evaluation: How will the prediction be used and by whom - so that we can build better models and make sure they get integrated into clinical care.

Start thinking beyond predictions and to do this, we will start with a framework that allows us to conceptualize **outcome-action pairing (OAP)**. In OAP, the outcome is the purpose of the AI model: disease diagnosis, risk stratification, or event prediction. The action is the step that can be taken based on the outcome that will improve medical care. A different treatment pathway based on risk, a new treatment based on a diagnosis, or scheduled follow-up with a primary care physician to prevent a readmission.

To evaluate an AI solution beyond its predictive value, we need to think about a framework that can be systematically applied across the broad range of AI solutions emerging in healthcare. We need a framework that provides criteria to evaluate the utility, feasibility, and overall clinical impact of an AI solution.

- Identify stakeholders and beneficiaries
- Outcome action pairing
- Number needed to benefit from the model

- Data availability and quality
- Implementation costs
- Clinical Uptake and Evaluation
- Deployment challenges
- Maintenance over time

- **Utility**: Refers to the purpose of the model and it is what matters to the patient most
- **Feasibility**: Refers to what is needed to implement the AI solution in the healthcare setting
- **Clinical Impact**: The overall effect that the AI solution can have on clinical care, patient outcomes, and care standards

Artificial Intelligence spans the breadth of healthcare and there are so many opportunities to use AI to improve upon or augment healthcare delivery. AI can identify patterns in the expanding, heterogeneous data sets in healthcare to create models that accurately classify, predict or recommend actions. However, realizing the potential benefit of AI solutions for patients in the form of better care requires rethinking how we evaluate AI solutions. A framework for rigorously evaluating the performance of a model in the context of the subsequent actions it triggers is necessary to identify AI solutions that are clinically useful.

CITATIONS AND ADDITIONAL READINGS

Coquet, J., S. Bozkurt, K. M. Kan, M. K. Ferrari, D. W. Blayney, J. D. Brooks, and T. Hernandez-Boussard. 2019. "Comparison of orthogonal NLP methods for clinical phenotyping and assessment of bone scan utilization among prostate cancer patients." *J Biomed Inform* 94: 103184.

Hao, S., Y. Wang, B. Jin, A. Y. Shin, C. Zhu, M. Huang, L. Zheng, J. Luo, Z. Hu, C. Fu, D. Dai, D. S. Culver, S. T. Alfreds, T. Rogow, F. Stearns, K. G. Sylvester, E. Widen, and X. B. Ling. 2015. "Development, Validation and Deployment of a Real Time 30 Day Hospital Readmission Risk Assessment Tool in the Maine Healthcare Information Exchange." *PLoS One* 10(10): e0140271.

Jeon, J., S. Nim, J. Teyra, A. Datti, J. L. Wrana, S. S. Sidhu, J. Moffat, and P. M. Kim. 2014. "A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening." *Genome Med* 6(7): 57.

Kansagara, D., H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, and S. Kripalani. 2011. "Risk Prediction Models for Hospital Readmission: A Systematic Review."

Matheny, M. E., D. Whicher, and S. Thadaney Israni. 2019. "Artificial Intelligence in Health Care: A Report From the National Academy of Medicine." *JAMA*.

Noack, M. M., K. G. Yager, M. Fukuto, G. S. Doerk, R. Li, and J. A. Sethian. 2019. "A Kriging-Based Approach to Autonomous Experimentation with Applications to X-Ray Scattering." *Sci Rep* 9(1): 11809.

Rajpurkar, P., J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. P. Langlotz, B. N. Patel, K. W. Yeom, K. Shpanskaya, F. G. Blankenberg, J. Seekins, T. J. Amrhein, D. A. Mong, S. S. Halabi, E. J. Zucker, A. Y. Ng, and M. P. Lungren. 2018. "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists." *PLoS Med* 15(11): e1002686.

Tariq, Q., J. Daniels, J. N. Schwartz, P. Washington, H. Kalantarian, and D. P. Wall. 2018. "Mobile detection of autism through machine learning on home video: A development and prospective validation study." *PLoS Med* 15(11): e1002705.

Yu, K. H. and I. S. Kohane. 2019. "Framing the challenges of artificial intelligence in medicine." *BMJ Qual Saf* 28(3): 238-41.