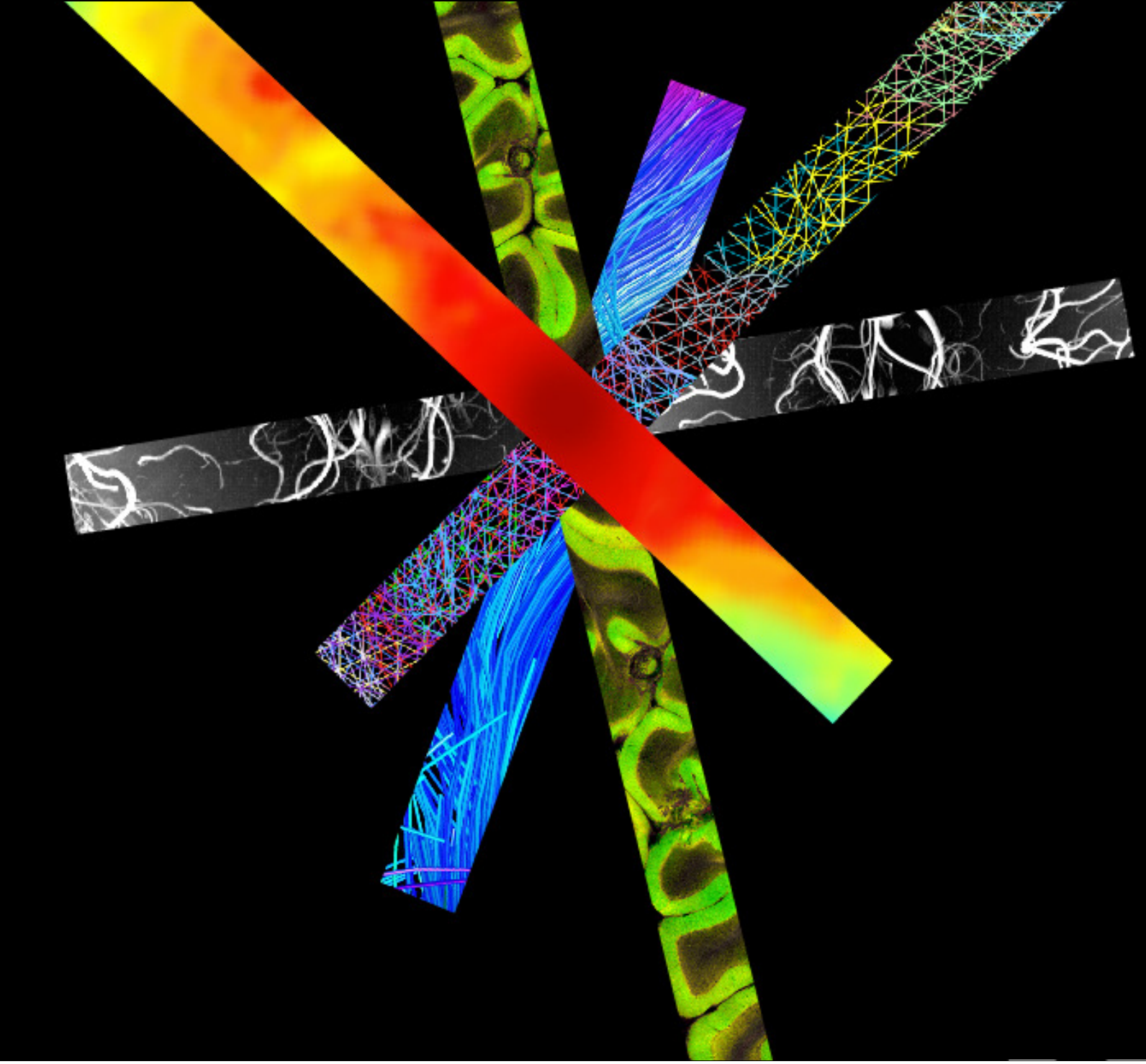


Deploying Clinical Machine Learning? Consider the Following...

Charles Lu, Ken Chang, Praveer Singh, Stuart Pomergrantz, Sean Doyle, Sujay Kakarmath, Christopher Bridge & Jayashree Kalpathy-Cramer
clu@mgh.harvard.edu



Motivation

- Despite high expectations for medical AI, few applications have been deployed
- Translation gap between machine learning (CML) research community and clinical users
- Healthcare domain has unique challenges such as acquiring large amounts of medical data, recruiting clinical expertise, and integration into existing hospital infrastructure

1. Clinical Context

1. **Identify pain points** – focus on developing tools for a particular stage or aspect of the clinical workflow such as triage, screening, or diagnosis
2. **Assess feasibility** – patient privacy, cohort selection, inter-annotator and intra-annotator variability, clinical performance metrics, clinical evaluation, user behavior modification
3. **Promote clinical champions** – can reaffirm clinical utility, guide development, and help with deployment by educating other clinical users and promoting project within healthcare bureaucracy

2. Clinical Validation

1. **Multi-site evaluation** – generalization gap can arise from differences in patient population, disease characteristics, or data acquisition settings
2. **Choose meaningful metrics** – E.g. goal is not merely to detect sepsis but to identify those cases of sepsis in which the clinician missed it and to detect early enough for intervention
3. **Perform user studies** – identify and alleviate points of friction; uncover small changes that result in substantial improvements to the user experience e.g. CM vs MM
4. **Embed situated workflows** – Should be a way for clinician to curate algorithmic results; e.g. What if there are multiple outputs which are not consistent with each other?

3. Deployment

1. **Development To Deploy** – pain points in the model deployment process can be mitigated by developing the model with deployment in mind from the outset.
2. **Choose an integration layer** – E.g., an application to detect motion artifacts might be deployed into the MRI scanner; while another application to triage stroke patients may be added directly into the clinical work-list software; while a critical findings notification application for stroke may be integrated into a mobile device environment
3. **Embrace healthcare standards** – Lack of support for standards increases integration costs and hospitals often fail to specify newer standards in their Request for Proposals (RFPs)

4. Monitoring

1. **Design for traceability** – Gradual accumulation of incident reports should alert manufacturers to failure modes, which might be more prone in certain risky subgroups or edge cases;
2. **Beware of bias** – prevent encoding and exacerbating existing disparities in data (during collection, selection, annotation, etc.), algorithm development (design choices, training parameters, testing evaluation), and framing (purpose, reification, intended vs. actual usage)
3. **Expect distribution shift** – Interesting dynamics will emerge as systems begin to affect clinical decision-making, which in turn affect patient outcomes, which in turn feedback into population data that will be used in future CML systems
4. **Missing outputs** – Once CML has become an established part of a clinical workflow, then a clinician may begin to rely on its presence and expect some default behaviors.

Survey

- We conduct interviews with clinical and technical practitioners from the Center of Clinical Data Science at the Massachusetts General Hospital and Brigham and Women's Hospital
- Participants had prior experience in developing medical AI with industry partners (e.g. GE Healthcare and Nvidia) for applications in radiology

Role	Years of experience	Considerations
Machine learning scientist	5	2.1, 3.2, 4.1, 4.2
Machine learning scientist	10	1.2, 2.2, 3.1, 3.3
Software engineer	30	2.4, 3.2, 3.3, 4.1
Clinical project manager	5	1.1, 1.3, 2.1, 4.2
Radiologist	25	1.3, 2.3, 4.3
Neurologist	10	1.3, 2.3
Clinical researcher	5	2.1, 4.2

Conclusion

These considerations gathered from our survey can help focus more targeted clinical machine learning to better situate into clinical context and facilitate the translation of promising research into real clinical benefit

<https://arxiv.org/pdf/2109.06919.pdf>