The Nightingale Project

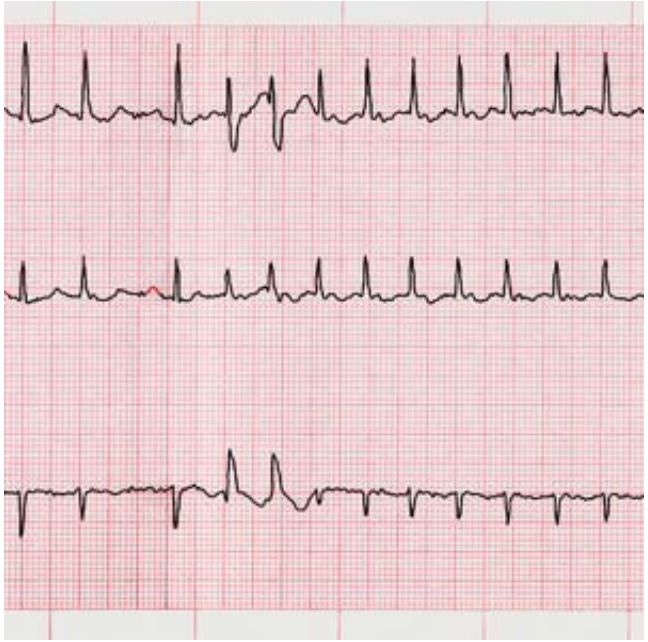
Bridging computer science & medicine to transform healthcare

The Potential

Detecting Afib using AI enabled ECG

"Patients could see this technology being used on a very routine basis very soon, probably within a few years, once we have **more robust data** and higher accuracy in the larger population."

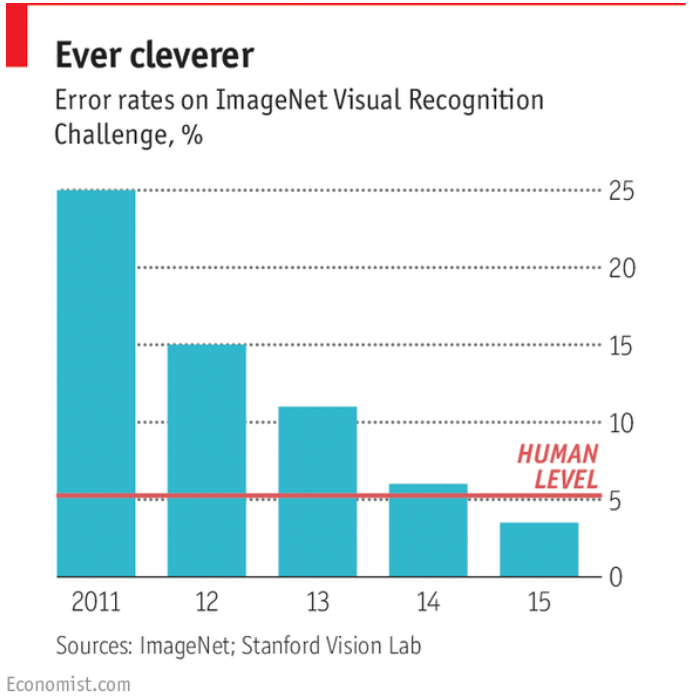
-Dr. Anuj Shah



Why doesn't medicine have more of these discoveries?

...Because the data is so hard to get

- We've seen huge performance gains over the last 10 years in visual object recognition bc of public data -- and we think medicine should follow their example.
- This is the goal of Nightingale: to make interesting, clinical data available to the public.



Nightingale

CHICAGO BOOTH
The University of Chicago Booth School of Business

SCHMIDT FUTURES

- An open data platform for research, bringing together
  - Leading health organizations
  - World class researcher communities

- Around interesting datasets
  - High-dimensional medical data (x-rays, ECGs, pathology)
  - Linked to real outcomes

Why We're Doing This

If we succeed, there's high return:

- World class researchers have access to a level of clinical data never made available before, fueling medical breakthroughs

- Increased research capacity and data infrastructure at leading health institutions –> more data-driven care and diagnostic improvements

If we fail, clinical data will remain a privately held, and we run the risk of setting back other efforts to make this kind of data a public good.

Challenges to executing our vision

- We don't just want single sets of data (i.e., easily identifiable set of x-rays)
- We also want ground truths, involving more complex outcomes data from various sources.

In starting to execute this goal, we faced two main barriers:

1) Why would hospitals give us data? We needed the right incentives
2) Privacy and data sharing logistics

Some of the things we heard

“Hospitals will charge a lot”

“Health systems won’t be interested unless they can make money from this”

“You won’t be able to grant data access in a way that hospitals are comfortable with”

What We Thought

Our hypothesis going in: We thought we would make many outbound sales calls in order to get just a few partners to commit.

The Reality

We actually had a high pickup rate -- out of 7 or 8 pitches, we secured commitments from 5 partners.

Lessons Learned: Partnerships

Among health systems, reasons for being interested varied:

- Some partners were very interested in doing research jointly before release
  - Defining research questions = good motivator
- Others were interested in building research capacity
- Others truly wanted to donate data for the public good

Where are we now?

- Nightingale now has commitments from **five partners**
    - Domestic and international

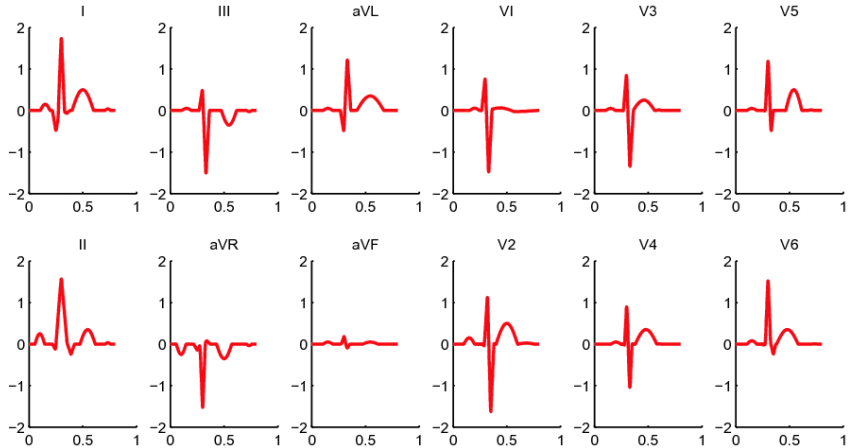
- Even if half of these partners fall out by the time we initially launch, there's still a big return given the richness and quantity of data that will be warehoused on the platform.

Example of partner research project/dataset

ECG waveforms linked to:

- Data on risk factors
- The results of prior diagnostic testing (such as catheterization)
- Other long term outcomes

Our main question: Is it possible to build a machine learning algorithm that would take in ECG waveforms, and output the probability that the patient has had a *prior* heart attack?



Example of partner research project/dataset

X-rays linked to:

- labels including demographics, fractures and outcomes related to frailty
- cancer surveillance data from national registries.
- Other long term outcomes on health, education

Our main question: Can we produce an algorithm that will predict patients who are likely to have future fractures, in order to identify patients who would benefit from treatment and physical therapy?

Use Case: A partner health center in the Bay Area

- Currently building a collaborative of researchers in this area who will work with the health system to identify key research questions,
- Then jointly work on datasets that stem from those research projects
- By helping this system build research capacity and open up the data, we hope to transform them into a model for data driven healthcare

Lessons Learned: Legal

- Legal part of building this: just as difficult as technical part!

- Key to have our own in-house legal team to advocate for our use case; strike balance between researcher needs and health system concerns

Lessons Learned: Data Architecture

- For each partner site, architecting the data so it's clean and linked with meaningful outcomes will certainly be a challenge, but one we know we can accomplish

- Some the biggest challenges haven't been technical: they are people and process challenges
  - IT, research, and information deficits at these health centers.
  - Institutional knowledge churn

Lessons Learned: Data Architecture

Computing Costs

	Year	Quarter	Q-Q Gowth	Registered Users	Avg Daily Active Users	Daily Active %	Computing Cost
GA - Year 1	2020	4	~	20	10	50%	$73,000
	2021	1	100%	40	13	31%	$91,250
	2021	2	200%	120	43	36%	$314,813
	2021	3	100%	240	71	29%	$516,703
Year 2	2021	4	90%	456	126	28%	$917,576
	2022	1	80%	821	214	26%	$1,560,914
	2022	2	65%	1,354	320	24%	$2,337,576
	2022	3	50%	2,031	419	21%	$3,056,028

If we're successful and have ~500 active users….we're talking about over 3,000 GPU hours/day & a little over $3,000,000 annually

How can this room help enable Nightingale?

- We would love your help!
- Research community: input on data and problems to focus on
- Funding community: ideas on how to fundraise for...
  - Seed grants to help fund researcher access
  - Seed/pilot grants to fund challenge grants that can help researchers
- Everyone: Feedback on anything you heard today

Thank you.

katy.haynes@nightingaleproject.a