Machine Learning in Medicine: Priorities for the Research to Product Gap

Alexander Lavin

Chief Scientist, Augustus Intelligence; Founder, Latent Sciences

AUGUSTUS

Cornell ML in Medicine | July 17, 2020

Latent Sciences

Outline

1. Intro
2. ML solutions in medical practice
3. AI Research ➟ Product
4. Examples
    a. CV in histopathology
    b. Probabilistic programmed disease modeling
5. Closing and discussion

Intro to Alexander Lavin

My work:

- Specialize in Bayesian ML and probabilistic computation
- Studied rocket science, then pivoted to the intersection of AI and neuroscience
- Scientist, engineer, founder, angel, advisor, consultant

Find me:

t: @theAlexLavin
w: lavin.io
e: lavin@latentsci.com

AUGUSTUS

AI solutions that drive human-machine synergies. We focus on data-efficient CV and NLP, and human-computer collaboration methods.

In medicines, Augustus works on machine vision for biopharma researchers and cancer diagnosis, and personalized medicine applications.

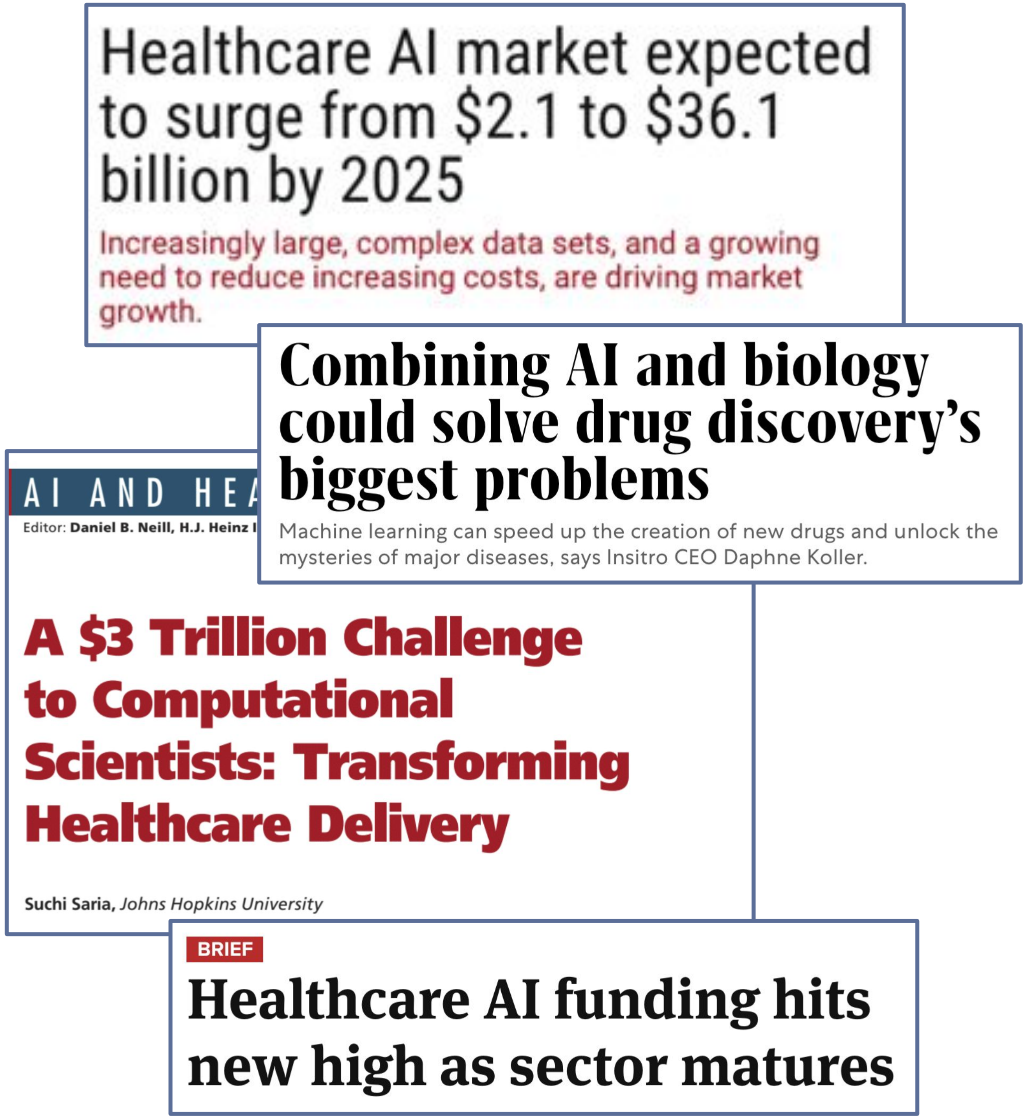
Latent Sciences

Productization of a patent-pending *Neurodegeneratives Prediction Engine* for state-of-the-art early prediction of Alzheimer’s (with blood-biomarkers).

Latent has worked with industry leading pharmacos and biotechs to boost neurodegen clinical trials, and is long-term developing a personalized medicines product.

Outline

1. Intro
2. **ML solutions in medical practice**
3. AI Research ➟ Product
4. Examples
    a. CV in histopathology
    b. Probabilistic programmed disease modeling
5. Closing and discussion



Healthcare AI market expected to surge from \$2.1 to \$36.1 billion by 2025

Increasingly large, complex data sets, and a growing need to reduce increasing costs, are driving market growth.

Combining AI and biology could solve drug discovery's biggest problems

Machine learning can speed up the creation of new drugs and unlock the mysteries of major diseases, says Insitro CEO Daphne Koller.

Editor: Daniel B. Neill, H.J. Heinz

A \$3 Trillion Challenge to Computational Scientists: Transforming Healthcare Delivery

Suchi Saria, *Johns Hopkins University*

BRIEF

Healthcare AI funding hits new high as sector matures

`Insitro` (Daphne Koller) : in-vitro experiments + AI and data generation, to predict what drug developers would see in-vivo

`Path AI` : state-of-the-art CV innovations for more accurate histopathology

`Tempus` : mining everything from genetic sequencing to image recognition

`Olive` : AI platform for automating the healthcare industry's most repetitive tasks

`Vicarious Surgical` : combines virtual reality with AI-enabled robots

`Cleveland Clinic` : distilling trillions of administrative and health record data points to personalize healthcare

`Johns Hopkins` : predictive AI techniques to improve the efficiency of patient operational flow

ML research progress	Real-world objectives
Counterfactual reasoning courses of treatment	Clinical risk prediction, for e.g. ICU optimization
NN-based survival models	ITE w/ complex, heterogeneous, longitudinal data
Auto-speech recognition of and topic extraction	Digital scribe: auto log notes, free-up doctors, track visits
Pneumonia detection in chest x-ray	Interpretable radiology assistant for lung inspection
Learning health knowledge graphs from EMR	Self-diagnostic symptom checkers

Misalignment in medical ML research and real-world solutions

Medical objectives != ML objective functions

Artificial intelligence := software systems that enable rational decision making under uncertainties

Thus we suggest that ML in medicine requires:

1. ML project lifecycle prioritizes interpretability, practical workflows, and uncertainty reasoning
2. Work with domain experts and medical professionals
3. Implement tight feedback loops externally *and* internally

“Kernel” of medical ML strategy

Diagnosis of the problem : Medical objectives != ML objective functions

Define high-level policy : AI is software systems that enable rational decision making under uncertainties

Coherent actions to execute the strategy:

1. ML project lifecycle prioritizes interpretability, practical workflows, and uncertainty reasoning
2. Work with domain experts and medical professionals
3. Implement tight feedback loops externally *and* internally

Rumelt, “Good Strategy Bad Strategy: The Difference and Why It Matters”

“Kernel” of medical ML strategy

Diagnosis of the problem : Medical objectives != ML objective functions

Define high-level policy : AI is software systems that enable rational decision making under uncertainties

Coherent actions to execute the strategy:

1. ML project lifecycle prioritizes interpretability, practical workflows, and uncertainty reasoning
2. Work with domain experts and medical professionals
3. Implement tight feedback loops externally *and* internally

Rumelt, “Good Strategy Bad Strategy: The Difference and Why It Matters”

Outline

1. Intro
2. ML solutions in medical practice
3. **AI Research ➞ Product**
4. Examples
    a. CV in histopathology
    b. Probabilistic programmed disease modeling
5. Closing and discussion

Phase 1: ML Research

Focus on unsupervised methods

- Labeled medical datasets can be prohibitively expensive
- Often the medical workflow bottleneck is data gathering/munging/labeling
- Expert labels may be incorrect!
    - error prone
    - qualitative and subjective in complex and heterogeneous diseases
- Inverse relationship between data-efficiency and usability

How?

- Generative models, namely VAE
- Methods that encode domain knowledge, i.e. PGM and PPL
- Also active learning / self-supervised learning (still super difficult!)

Phase 1: ML Research

Focus on aleatoric + epistemic uncertainties

- *Aleatoric uncertainty* measures the noise inherent in the observations.
- *Epistemic uncertainty* accounts for uncertainty in the model itself; i.e. capturing our ignorance about which model generated the data.

NNs are often miscalibrated -- i.e. over-confident because ignoring epistemic uncertainty

Probabilistic ML methods offer uncertainty reasoning for free: Gaussian Processes, Bayes NN, BayesOpt

Phase 1: ML Research

Focus on aleatoric + epistemic uncertainties

- *Aleatoric uncertainty* measures the noise inherent in the observations.
- *Epistemic uncertainty* accounts for uncertainty in the model itself; i.e. capturing our ignorance about which model generated the data.

NNs are often miscalibrated -- i.e. over-confident because ignoring epistemic uncertainty

Probabilistic ML methods offer uncertainty reasoning for free: Gaussian Processes, Bayes NN, BayesOpt

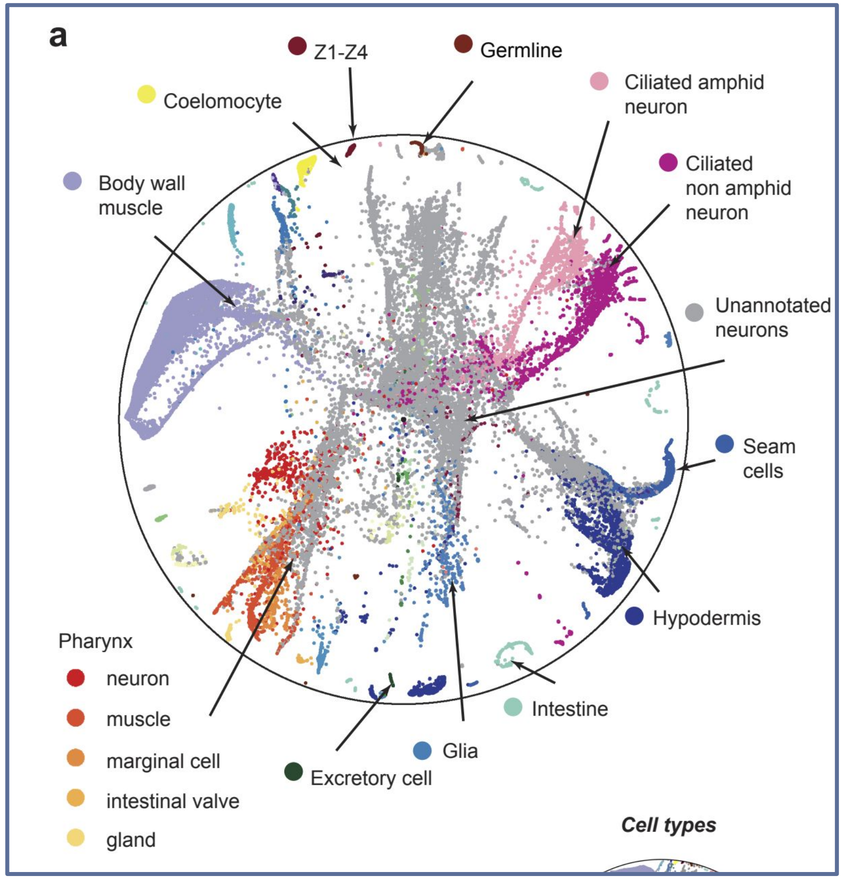
See Mihaela van der Schaar's **AutoProg**:
vanderschaar-lab.com/clinical-support

Phase 1: ML Research

Focus on model-based interpretability

- Methods that are by definition white-box can be very handy in research
- Probabilistic ML (namely probabilistic programming)
- Methods that learn a latent representation/embedding of data for visualization (e.g. VAE and 👉)

Neural Net methods struggle here, resort to post-hoc interpretation methods…



Poincaré map of C. elegans cell atlas (Klimovskaia et al. ‘20).

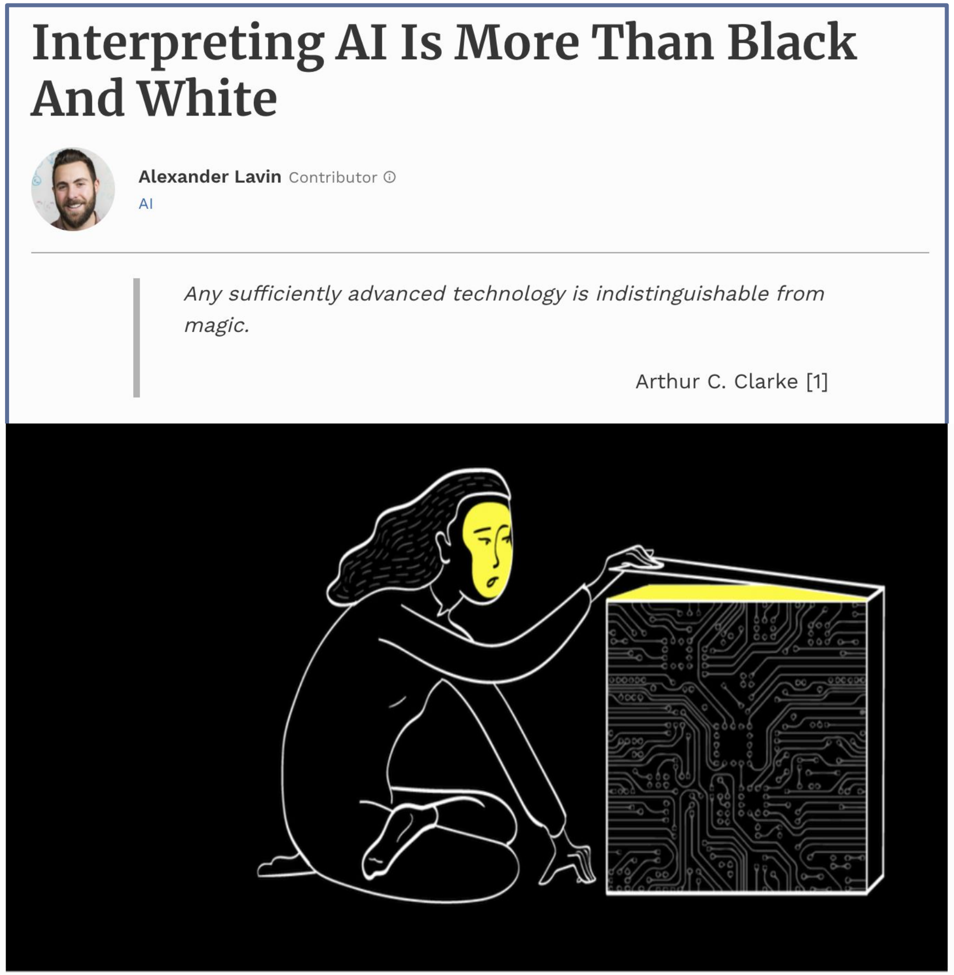
Several notions and abstraction levels of ML interpretability

Low-level: The ability to explain a model's behavior, answering to an ML engineer, "why did the model predict that?"

High-level: The ability to translate a model to business objectives, answering in natural language, "why did the model predict that?"

Post-hoc interpretation methods: applied after-the-fact, e.g. heatmaps and network activation viz.

Model-based interpretation: the model itself readily provides insights into the relationships and structures it learns from data



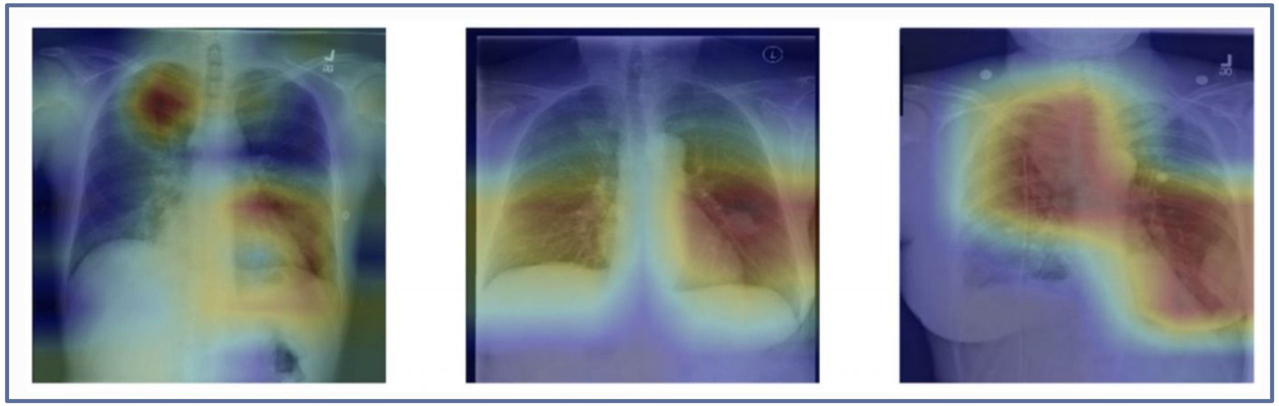
Phase 2: prototyping and development

Focus on **post-hoc interpretability**

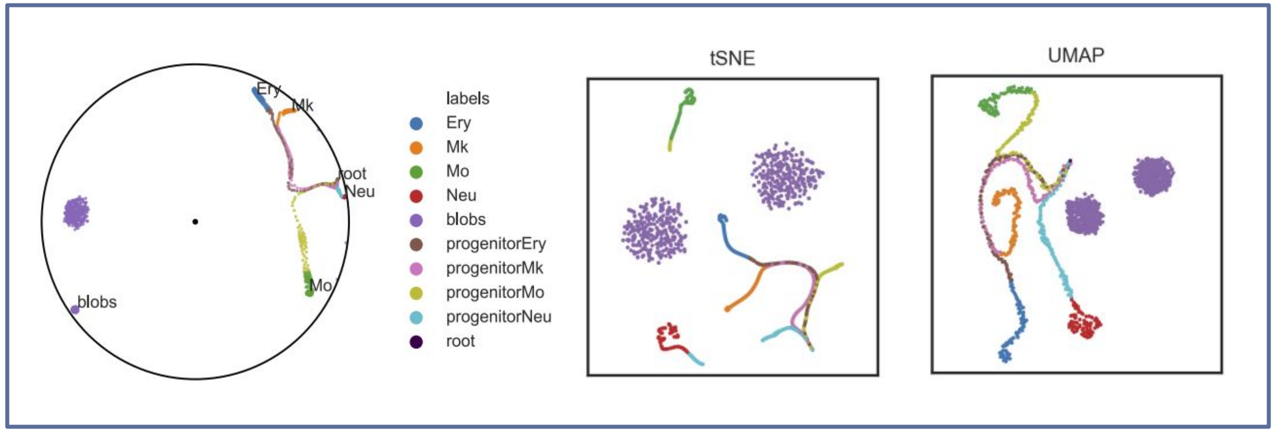
- Explaining models and predictions at a high-level
- Builds trust w/ non-ML experts (doctors and patients)

How?

- Modeling decisions that behoove these interp methods
- E.g. work with gradient-based and perturbation-based methods from captum.ai/docs/algorithms
- Dimensionality reduction → medical decision tool



CheXNet (Rajpurkar et al. ‘17) localizes pathologies it identifies using Class Activation Maps (Zhou et al. '16), which highlight regions that are most important for making a particular pathology classification.



Comparison of various embeddings for a synthetic model of myeloid progenitors differentiation -- Poincaré on the left, then two state-of-art visualization methods. (Klimovskaia et al. ‘20)

Phase 3: productizing and deployment

Focus on testing and feedback loops

- Testing is critical throughout the ML project lifecycle, but here it is near 100% of the efforts
- Monitoring for regressions and data distribution shifts
- ML in general and paramount in healthcare, need to identify and CI test the critical scenarios and data slices
- Testing and deployment flexibility for on-prem, hosted, distributed

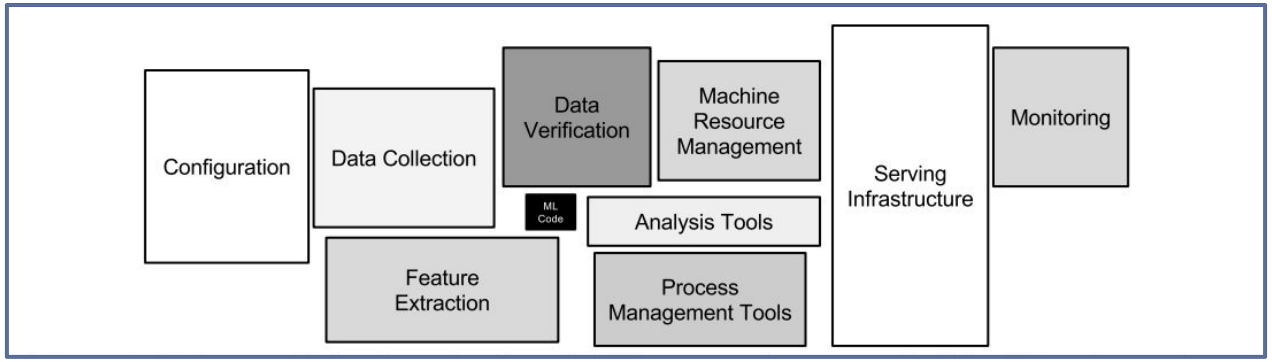


Figure from D. Sculley et al. '19: Only a small fraction of real-world ML systems are composed of the ML code (small black box in the middle). The required surrounding infrastructure is vast and complex.

Note that product features like confidence measures and model explanations are easy because we prioritized them in research.

Approaching ML research for eventual deployment

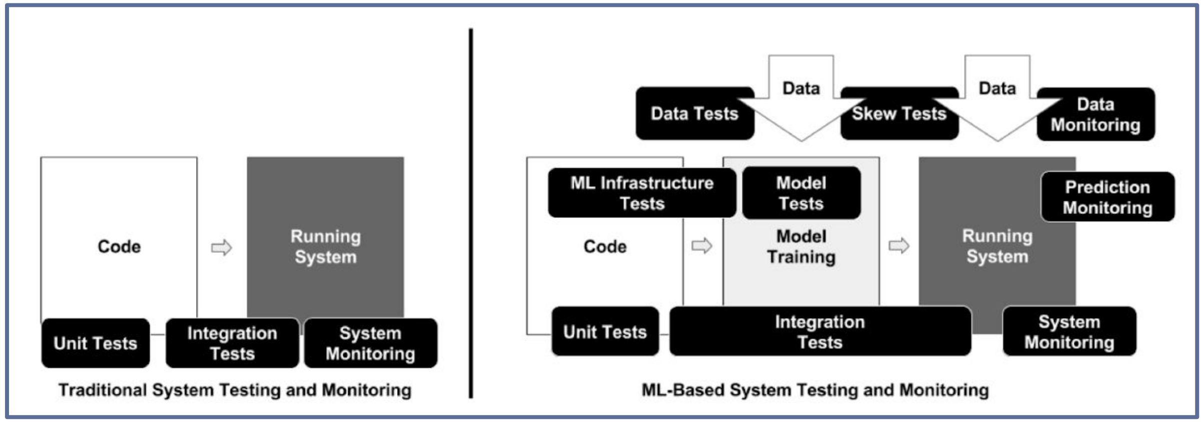
A lot we did not touch on…

ML is one component of a much larger integrated system.

Machine learning != Software engineering

Need to consider deployment scenarios and constraints earlier in the pipeline:

- Data and other deployment constraints
- AI ethics



`ML systems impose significant testing requirements on top of existing software testing.`

For more: Lavin & Renard (2020). Technology Readiness Levels for ML Systems. [arXiv: 2006.12497.](https://arxiv.org/abs/2006.12497)

Outline

1. Intro
2. ML solutions in medical practice
3. AI Research ➟ Product
4. **Examples**
   a. CV in histopathology
   b. Probabilistic programmed disease modeling
5. Closing and discussion

Computer vision and histopathology

Current CV histopathology methods learn end-to-end with labeled data.
Doesn’t translate to real-world:

- needs massive labeled data
- specific to a few cancer types (yet there are 100+ of brain cancer subtypes)
- not always justification for the classification
- typically uninterpretable

Misalignment:

Research objective is classification performance on clean benchmark datasets.

Real-world objective is to identify cancerous tissues to best inform medical professionals.

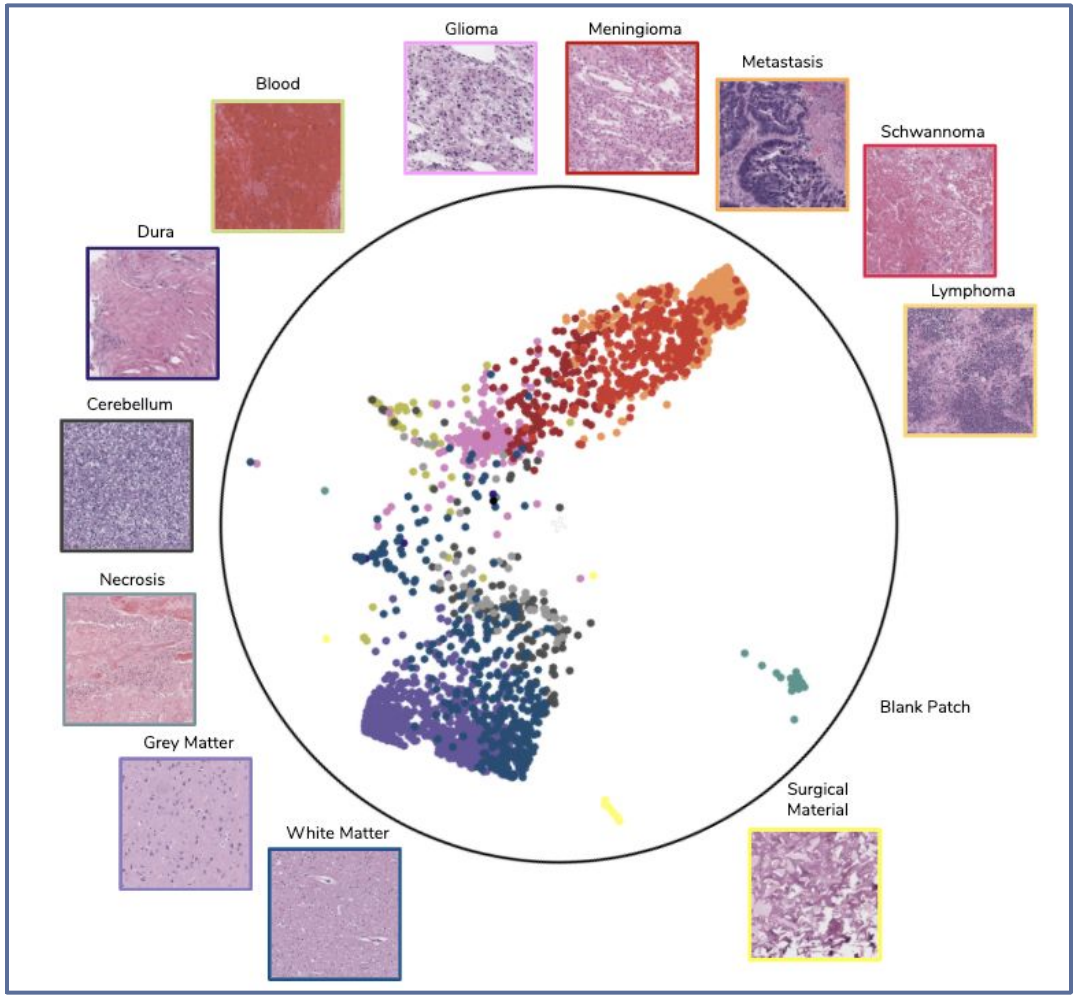
Unsupervised visual anomaly detection in neuropathology

We instead pursue an unsupervised, interpretable method:

- *Stereographic Projection Variational Auto-Encoder*
- Poincaré ball illustrates learned semantics and hierarchy
- samples from latent manifold yield reliable tissue images

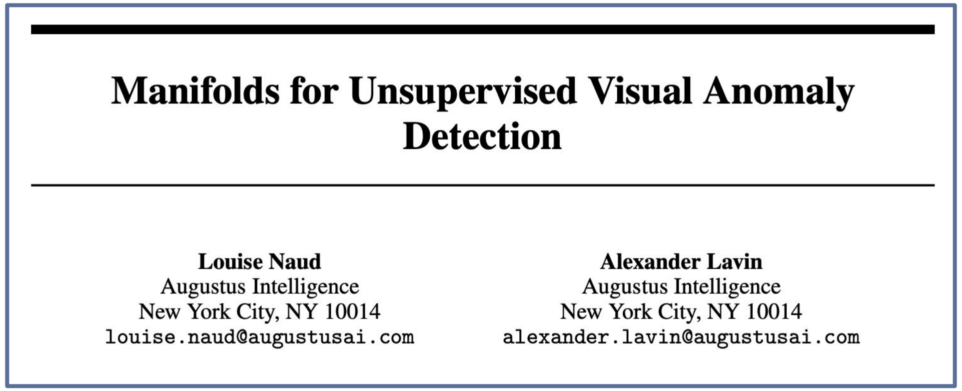
Why anomaly detection?

- methods generalize, and handle rare, unseen classes
- surface the most valuable information for medical professionals to make decisions



Naud & Lavin (2020). Manifolds for Unsupervised Visual Anomaly Detection. arxiv.org/abs/2006.11364.

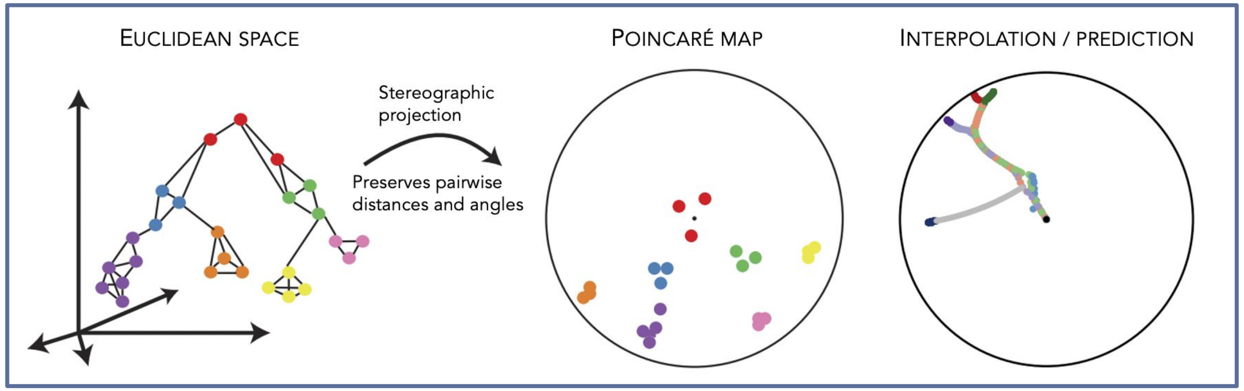
Unsupervised visual anomaly detection in neuropathology



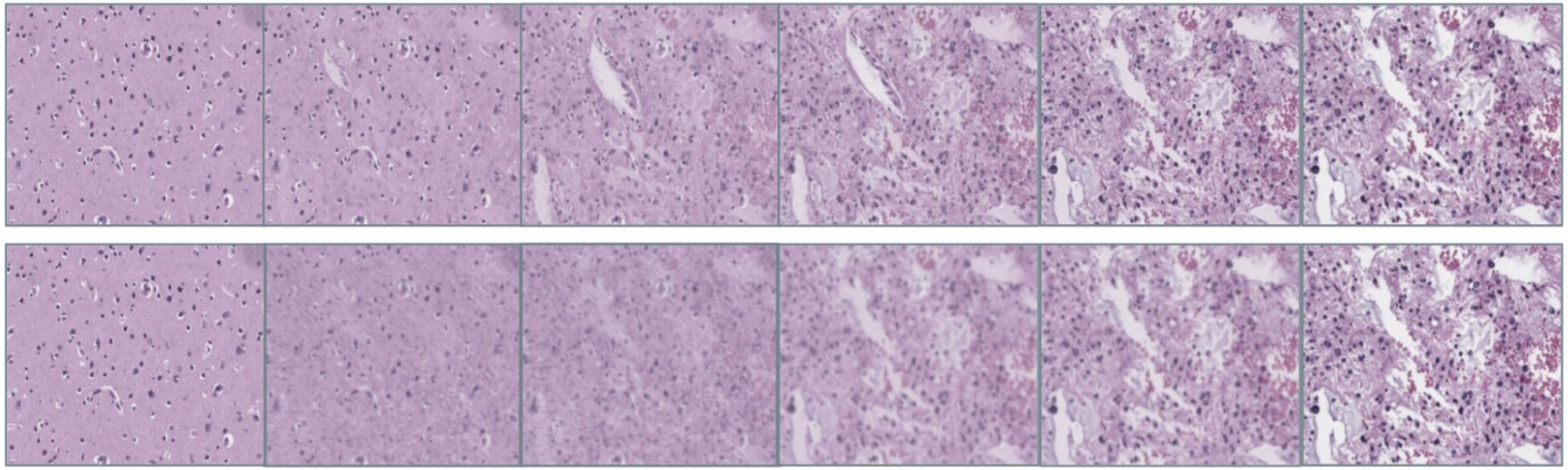
Manifolds for Unsupervised Visual Anomaly Detection

Louise Naud
Augustus Intelligence
New York City, NY 10014
louise.naud@augustusai.com

Alexander Lavin
Augustus Intelligence
New York City, NY 10014
alexander.lavin@augustusai.com



Geodesic interpolation along hyperbolic manifold



Linear interpolation approximating the hyperbolic manifold

Longitudinal modeling of Alzheimer’s Disease

AD has myriad and unique complexities:

- heterogeneous biological pathways and latent-time processes
- complex temporal patterns; survival is non-linear, features interact, non-stationary states
- onset of disease pathology != onset of symptoms, nor is dementia an absorbing state
- subjective diagnoses and infrequent clinical measures

Typical methods don’t suffice

- Deep learning approaches are too data hungry and black-box → need white-box and data-efficiency
- Need dynamic, active learning
- Need models that more faithfully represent disease states: i.e. continuous time

Misalignment:

Research objective is predicting survival, or diagnosing in broad buckets.

Real-world objective is individual-specific pre-symptomatic prediction.

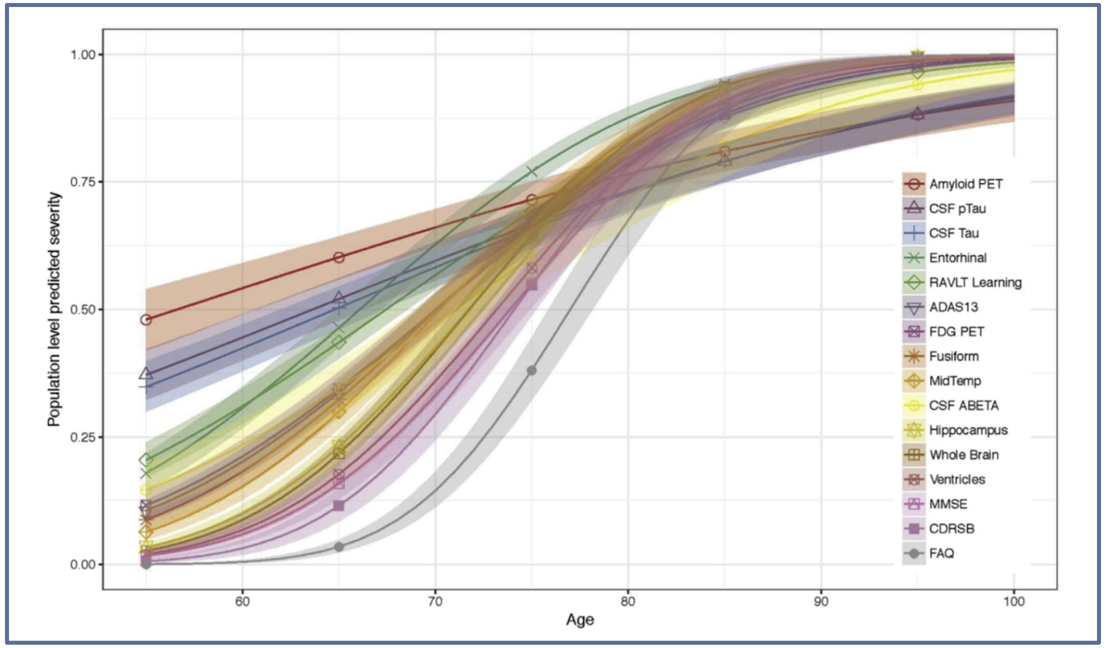
Probabilistic-programmed Gaussian Process models of neurodegeneration

We instead pursue methods and representations that enable,

- principled uncertainty reasoning
- unsupervised, data-efficient learning
- flexible modeling: individual-specific, encode domain priors
- interpretable system

Probabilistic programming

- generative models to describe biomarker progressions: monotonic GPs
- easily encode domain expertise
- uncertainty quantification for free
- PPL are by definition white-box



Top: Alzheimer's pathological cascade of biomarker trajectories, derived from ADNI.

Right: Example linear Gaussian model in the PPL Turing (Ge et al. '18). Models expressed as probabilistic programs are fundamentally white-box.

```
@model gdemo(x, y) = begin
  # Assumptions
  σ ~ InverseGamma(2,3)
  μ ~ Normal(0,sqrt(σ))
  # Observations
  x ~ Normal(μ, sqrt(σ))
  y ~ Normal(μ, sqrt(σ))
end
```

Outline

1. Intro
2. ML solutions in medical practice
3. AI Research ➟ Product
4. Examples
    a. CV in histopathology
    b. Probabilistic programmed disease modeling
5. **Closing and discussion**

Take-home messages

Research is the tip of the iceberg.

Misalignment problem: ML objectives != medical objectives

Working definition of AI: software systems that enable rational decision making under uncertainties.

Action steps: prioritize interpretability, practical workflows, and uncertainty reasoning at different phases of ML project lifecycle.

This is non-trivial; assuming ML methods will work out of the box will always fail in medical applications.

Explainability isn’t for understanding, it’s for trust.

Thank You

e: lavin@latentsci.com
t: @theAlexLavin
w: lavin.io

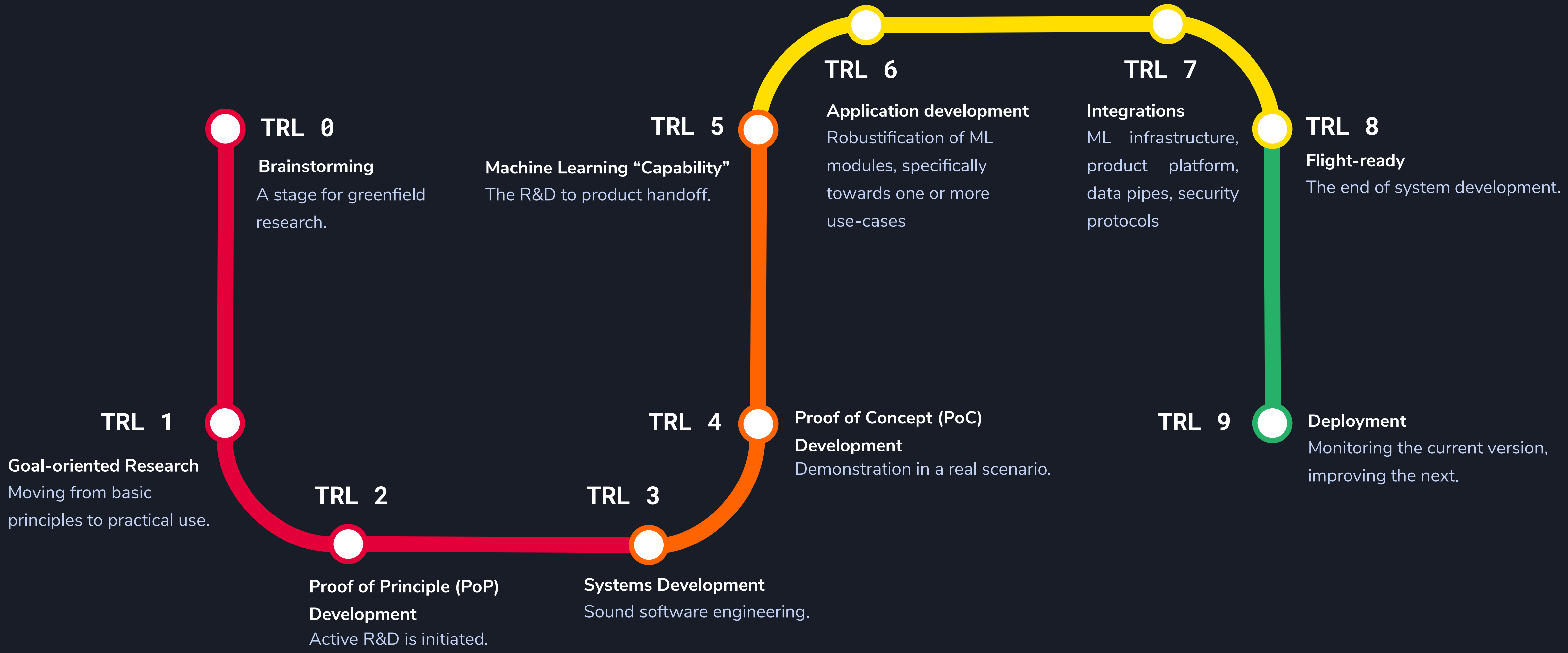
Appendix

References

TODO

- ...

Research ↔ Development ↔ Productization ↔ Deployment



Lavin & Renard. Technology Readiness Levels for Machine Learning Systems. *ICML 2020 Workshop on Challenges in Deploying ML Systems.* arXiv: 2006.12497.

TRL Cards

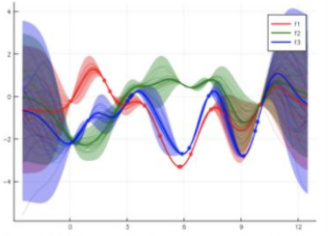
Tool for communicating ML technology readiness across all internal stakeholders.

Enables inter-team and cross-functional communication.

Lower-level and more process-oriented than other “ML cards” -- e.g. Google (Mitchell et al. ‘19) and Hugging Face.

Standardized “report cards” for TRL4ML stage reviews.

TECHNOLOGY NAME	Solar Array Optimization v1.0
TRL	7 <link to previous cards>
R&D OWNER / REVIEWER	A. Lavin / G. Renard
PROD OWNER / REVIEWER	S. Wozniak / S. Jobs
COMPONENT CODES	1.1, 4.2, 4.3
TL;DR	Applying our multivariate BayesOpt (MVBO) algorithm to the problem of solar panel configuration optimization, specifically towards client SolarUS.
Data considerations	Two datasets have been used to train and validate the system: 1. Pilot dataset provided by SolarUS 2. Simulated datasets (which we derived from SolarUS data, w/ Gaussian noise); explores add’l geographic regions and climates
Ethics	The datasets do not represent any biases. The algorithms have a very low carbon footprint. Augustus Ethics Checklist has been completed.

Model / alg details	MVBO runs iterative optimization over several surrogate GP models f1…n, each representing an independently modulated portion of the array field.  *Example with n=3 (radians and hours on the x and y axes)*
Metrics, results	MVBO algorithm converges to solution on opt. benchmark problems in ~1.0s on 4-core CPU. Full quantitative reports: < link to experiments wiki >
Caveats, known edge cases, recommendations	For the solar array problem we require multi-objective optimization: maximize energy-gain objective while minimizing panel-movement, accomplished via Pareto front optimization. This was stable on 98.8% of simulated scenarios (full range of solar exposures).
Key assumptions	We model solar radiance w/ simple Gaussian noise, and assume near-perfect actuation of solar panels.
Intended use	Optimize up to 5 continuous or discrete parameters of a given device, and a system of up to 40 devices.

The maturity of each model or algorithm is tracked via TRL cards. This card subset reflects our example BO algorithm at TRL 7.

Lavin & Renard. Technology Readiness Levels for Machine Learning Systems. *ICML 2020 Workshop on Challenges in Deploying ML Systems.* arXiv: 2006.12497.