AI *for* Epidemiology

B. Aditya Prakash

College of Computing

Georgia Institute of Technology

Aditya Lab

Georgia Tech

AIxHIV, Harvard

March 2, 2026

Georgia Tech

Thanks to the organizers

Milind Tambe
Harvard University

Davin Choo
Harvard University

Lingkai Kong
Harvard University

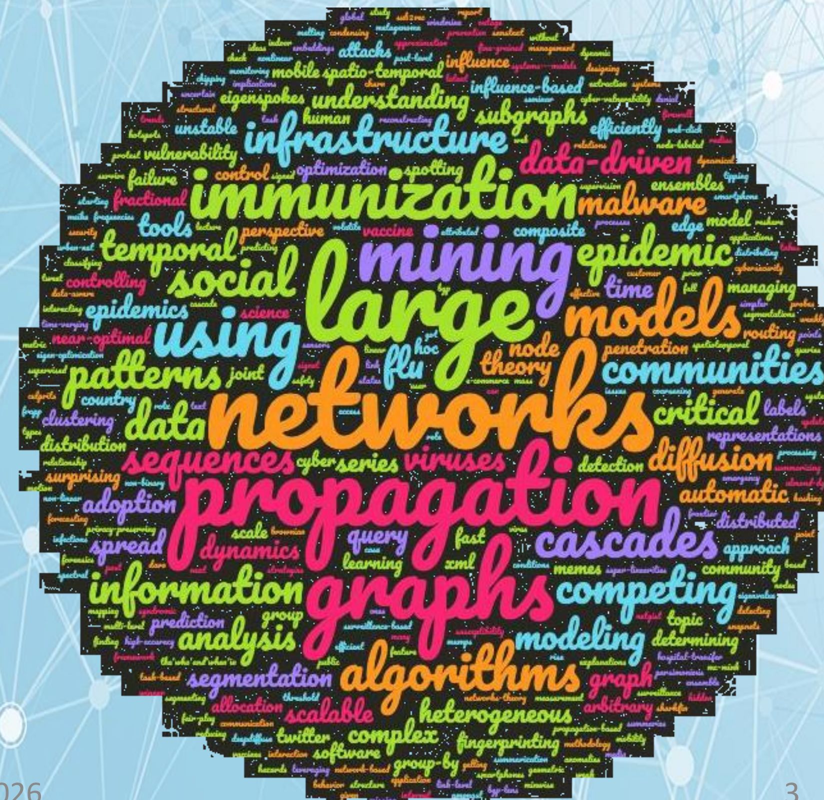
A bit about our lab…

♥ data

Emphasis on **solving big-data problems in networks and sequences**, motivated from high-impact applications such as *public health, urban computing, engineering, security and the Web*.

Background

- Data Science
- AI/Machine Learning
- Graphs, Sequences



This talk: AI for EPI

- Robust Forecasting
- Differentiable Epidemiology

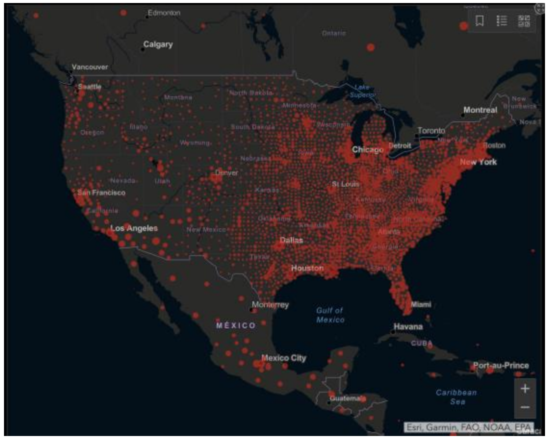
THREAD - 1

ROBUST AND CALIBRATED FORECASTING

Georgia Tech

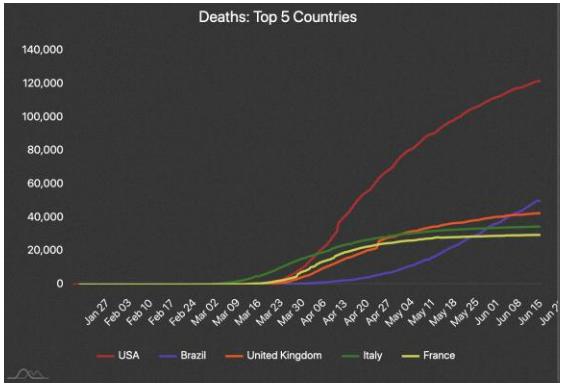
Covid-19 trajectories

Global spatial incidence distribution

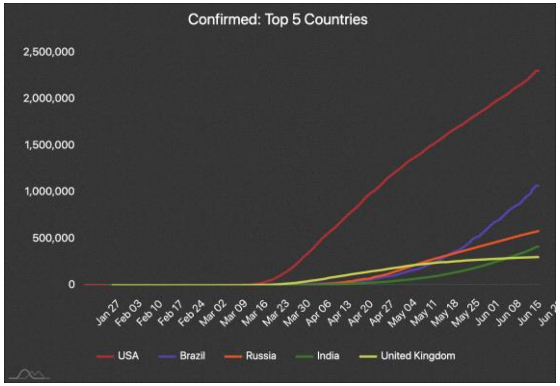


Spatial incidence distribution in USA

[source: https://gisanddata.maps.arcgis.com]



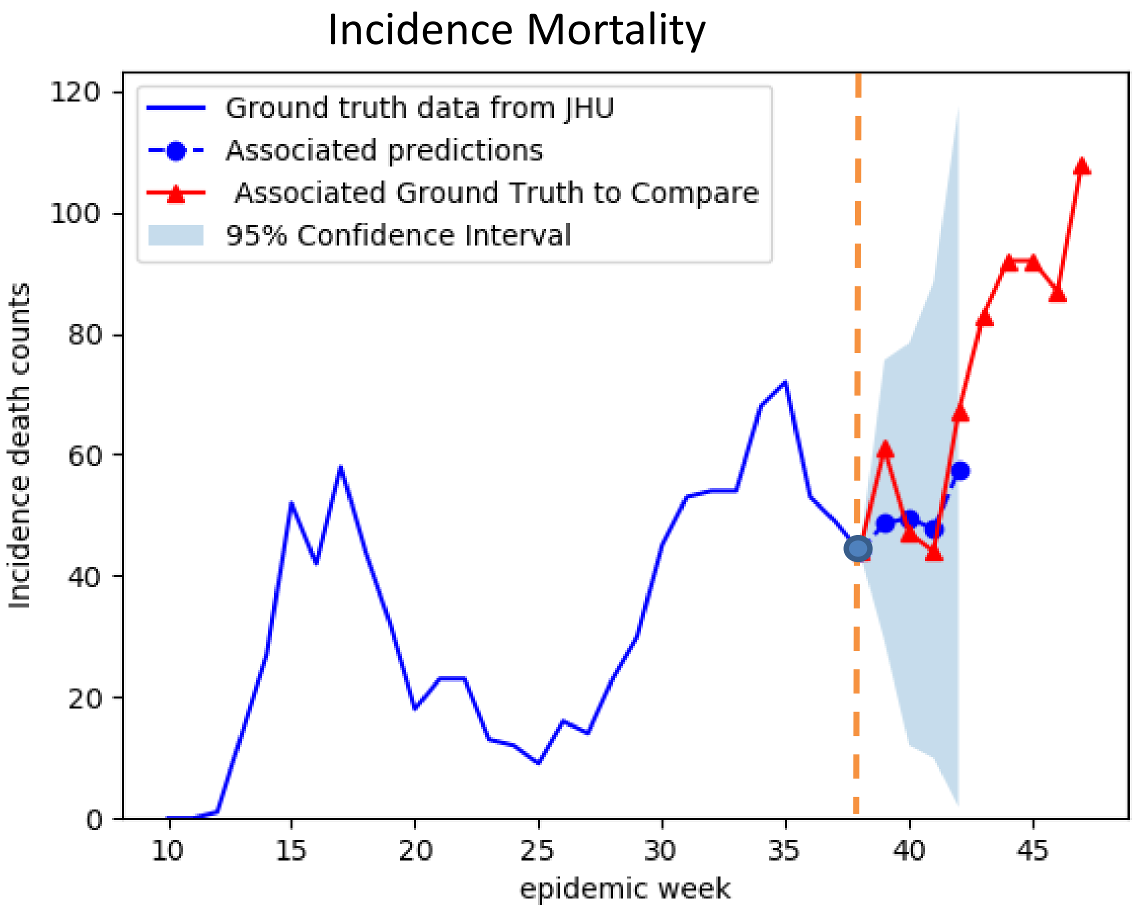
Cumulative Mortality



Confirmed cases Cumulative

[source: https://nssac.bii.virginia.edu/covid-19/dashboard/]

Real-time Forecasting



Possible near future:

- Goes down
- Stays still
- Goes up

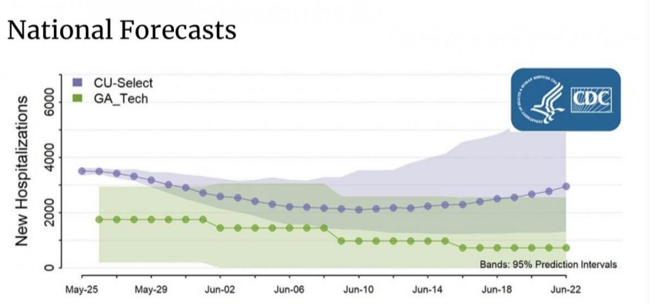
Depends on:

- Interventions in place
- Current number of infections
- Contact patterns
- Exposure to disease
- Etc

Why Forecasting?

An outlook to the future allow communities to

- Allocate resources/budget
  - Ventilators, enable more ICU beds
- Inform public policy
  - E.g., mandate shelter in place?
- Improve preparedness
- Public Communication
- …



https://www.cdc.gov/forecast-outbreak-analysis/index.html

CDC Centers for Disease Control and Prevention
CDC 24/7: Saving Lives, Protecting People™

Center for Forecasting and Outbreak Analytics

Home

Welcome to CFA

The Center for Forecasting and Outbreak Analytics (CFA) uses data, modeling, and analytics to respond to outbreaks in real-time to drive effective decision-making.

Learn more about CFA

Highlights

- Monkeypox Technical Report #3 NEW!
- COVID-19 Forecasts: Deaths | CDC
- COVID-19 Forecasts: Hospitalizations | CDC
- Early CFA Successes

CFA in Action

- About Us
- History of CFA
- News, Updates & Events
- Publications & Resources

Partnerships

- Engaging with STLT Partners Info
- Infectious Disease Transmission Models

Context

- One of our Lab’s focus: Explore performance and utility of **data-driven** methods in epidemiology/public health (surveillance, interventions, vaccination …)
  - Data from multiple sources is often more sensitive to what is happening ‘on the ground’
  - Complementary helpful perspective to other traditional models

- Example: AI models for forecasting

Why Deep Learning?

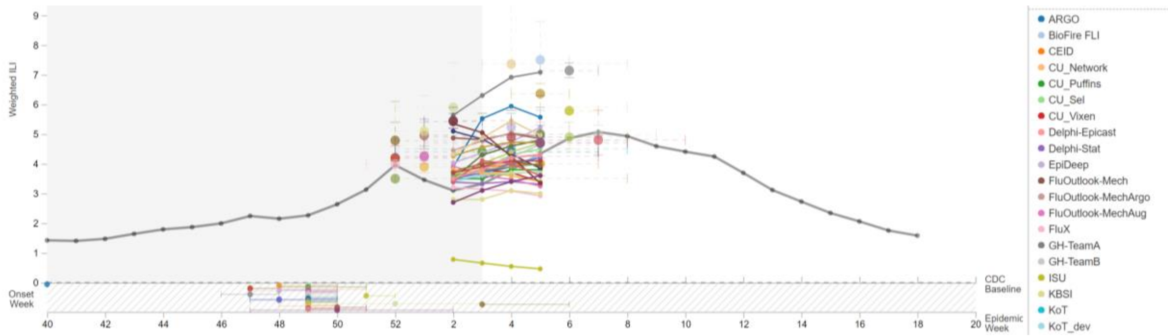
- Flexible, scalable, efficient technology
- Excellent choice to model non-linearities
- Able to incorporate different knowledge representations
- Very active research area

Our participation in CDC Forecasting efforts

Bijaya Adhikari (-> Iowa)

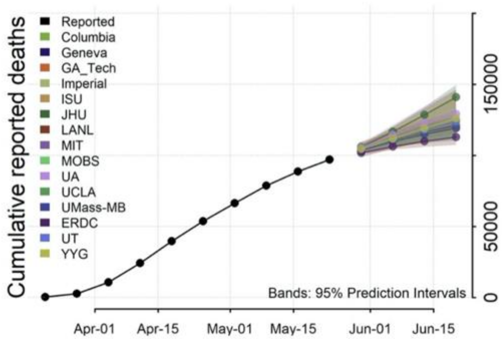
Alex Rodriguez → Michigan

Target 1: Weighted influenza like illness count per week

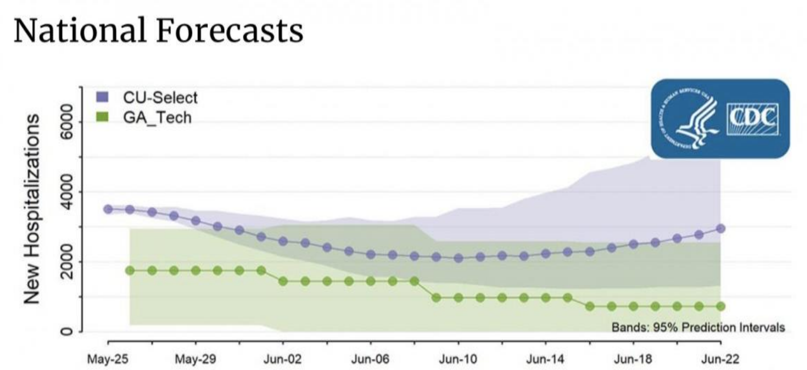


More focused on Short-term predictions

Target 2: COVID related Mortality per week



Target 3: Daily COVID induced Hospitalizations



referred to by the CDC director, media etc…

Our Impact

[PNAS 2022]

Only individual Deep Learning driven model in **top-5 for accuracy** in the CDC-led evaluation for more than 2 years

CDC

nature biotechnology

News | Published: 04 September 2020

Massive data initiatives and AI provide testbed for pandemic forecasting

Cormac Sheridan

Nature Biotechnology 38, 1010–1013 (2020)

3189 Accesses | 47 Altmetric | Metrics

IEEE Spectrum

FiveThirtyEight

1 of 11 shown on their page since summer 2020

Covered Widely in Media/Press (examples)

1st Prize

facebook

Carnegie Mellon University

Out of 115 global participants

2nd Prize

C3.ai COVID-19 Grand Challenge

43 Countries

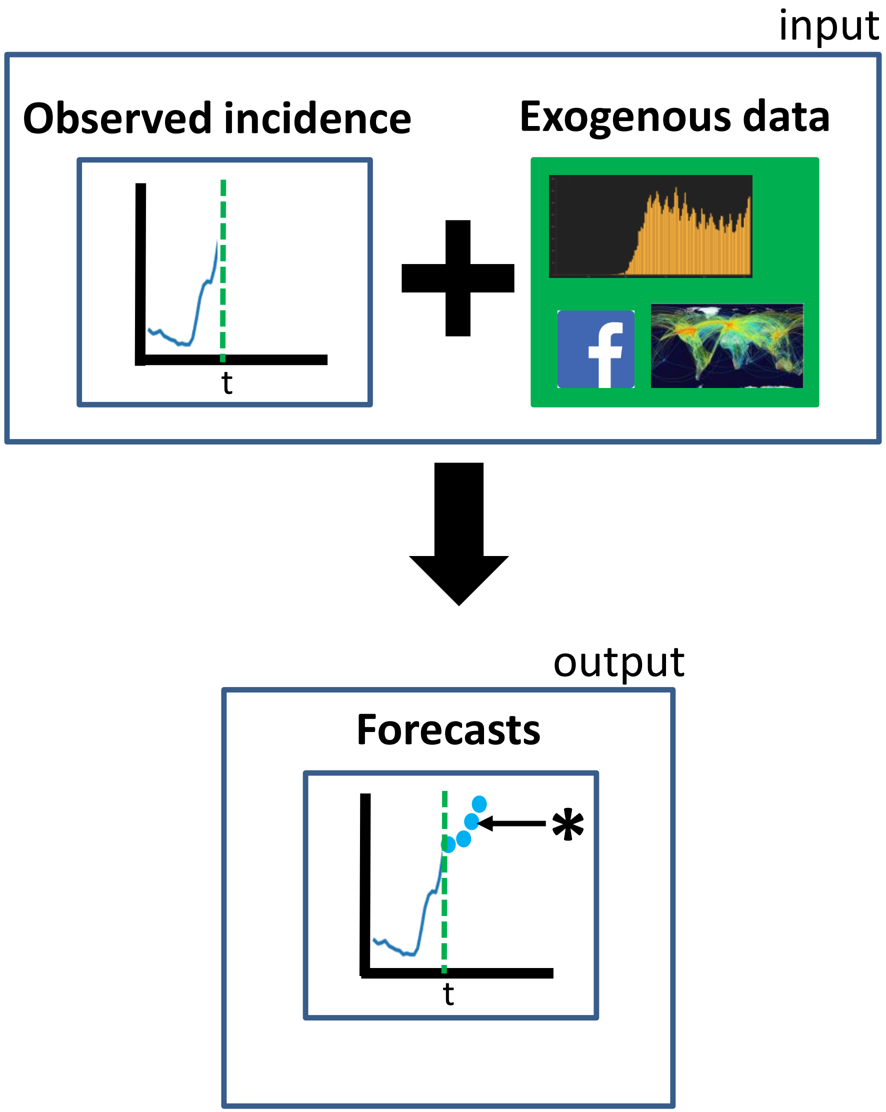
777 Participants

Used on campus, hospitals (UIHC)…..

No Historical data!

Formulation

- **Given**
  - Partially observed mortality and hospitalization incidence curve till day t.
  - Exogenous data sources
- **Predict**
  - Future mortality incidence for next four weeks on weekly basis
  - Future hospitalization incidence for next four weeks on daily basis



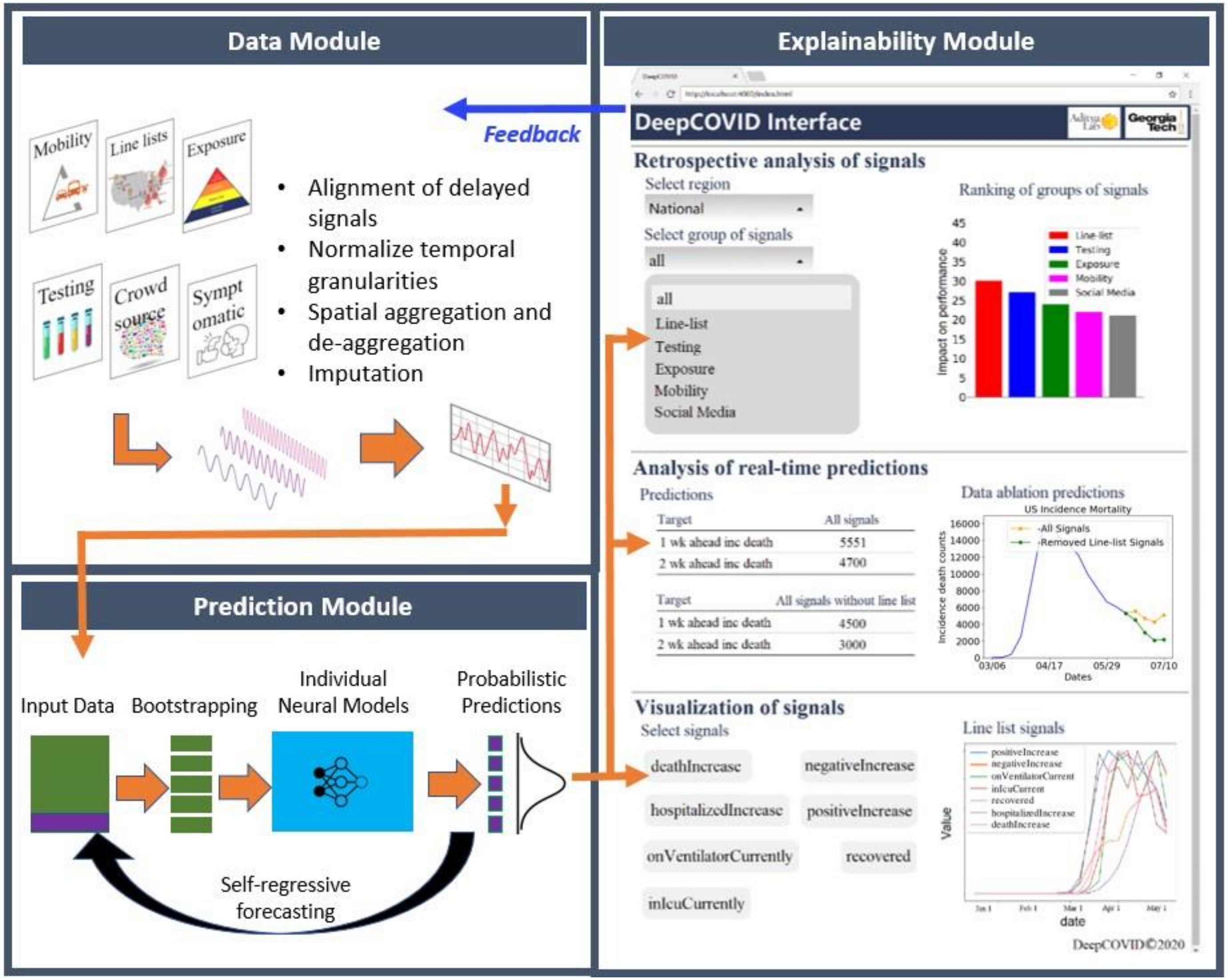
Georgia Tech

CoVDS Dataset

- Covid Surveillance Dataset: 21 signals across 50 states
- Target: mortality upto 4 weeks ahead
- Multiple sources
  - Patient line-list: like hospitalization from CDC, ICU admits
  - Testing: From CDC and Covid-Tracking
  - Mobility: from Google and Apple
  - Exposure: Proximity with POIs
  - Social Survey: FB Symptom survey
- Real-time forecasting from April 2020

CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION
The COVID Tracking Project
DELPHI

Our Operational Framework



[Rodriguez+ AAAI 2021]

Several lessons learnt

- Uncertainty is important!
- Explainability is important!
- Human input and feedback is important!
- Data is extremely chaotic and dirty!
  - Not to be underestimated
  - Evaluation itself may not be reliable

  ….

  ….

Challenges in Probabilistic Forecasting

- Learning well-calibrated forecast distribution is challenging
- Real-time data is noisy and unreliable
- Most methods don't provide good uncertainty estimates

OUR FOCUS:

Robust (<u>accurate</u> and <u>well-calibrated</u>) Forecasting

(1) Modeling Backfill

- Back2Future: Leveraging Backfill Dynamics for Improving Real-time Predictions in Future

Harshavardhan Kamarthi, Alexander Rodriguez,
B. Aditya Prakash

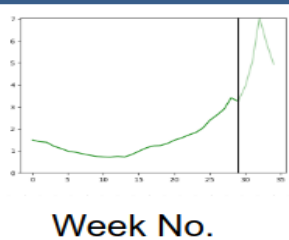
[In ICLR 2022]

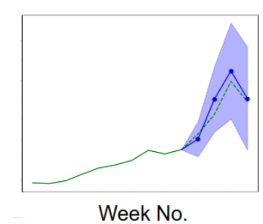
Real-time forecasting and evaluation setup

- Use signals from surveillance data at current week to train for future weeks
- Get ground truth for current week to evaluate scores for all models

Get ground truth available at current week → Evaluate models → Take decison

Get data available at current week → Update the model → Generate the forecasts

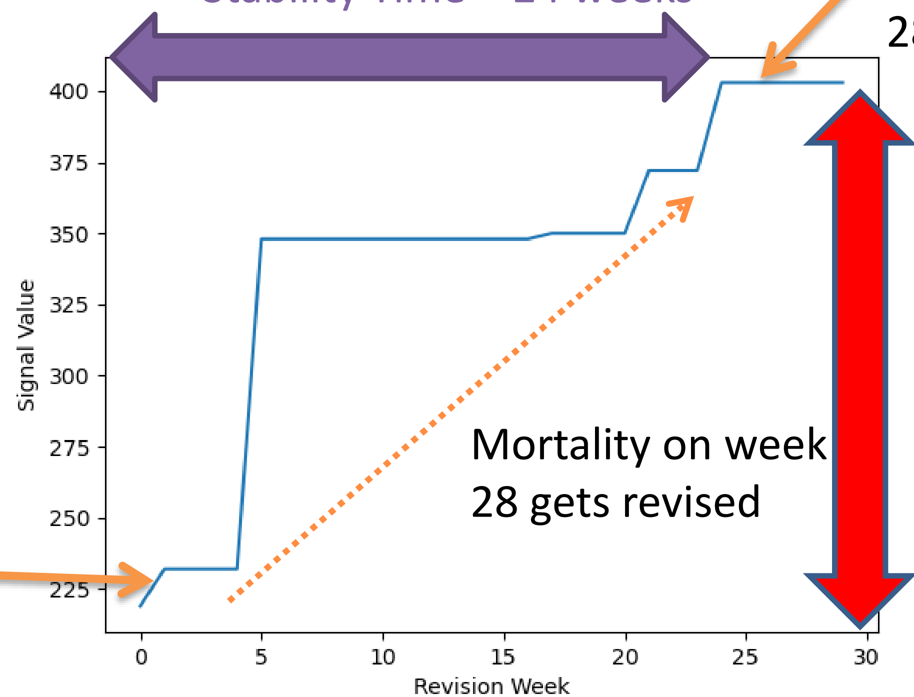




But targets *AND* features get revised!

Mortality finally stabilizes to around 400 at week 28+24 = week 52 !

Reasons?
Human error, data instability, delays, disasters



Mortality as repoted on week 28

Backfill Error = |404-223|/|404|= 44.8%

Revision of mortality for TX released on week 28

- Called Backfill phenomenon
- Observed for targets in context of flu [Reich et.al PNAS 2019, Chakraborty et.al PLoS 2018]

Common in Time-Series Forecasting where data is revised

- Important Machine-learning problem
- Wide range of applications
  - Security
  - Sales and Retail
  - Supply Chain Optimization 
  - Weather forecasting 
  - Epidemiology

Machine Learning Can Predict Wind Energy Efficiency

Download PDF Copy

By Surbhi Jain — Mar 25 2022
Reviewed by Susha Cheriyedath, M.Sc.

How to Predict Traffic on Google Maps for Android

written by Kaushal | January 30, 2019

Is the worst of the Covid-19 pandemic behind us? We asked 8 experts.

The case for cautious optimism for the US.

By Keren Landman, Umair Irfan, and Brian Resnick | Mar 2, 2022, 7:00am EST

Texas Economy

Texas Employment Forecast

March 25, 2022

The Texas Employment Forecast estimates that jobs will increase 3.4 percent in 2022, with an 80 percent confidence band of 2.6 to 4.2 percent.

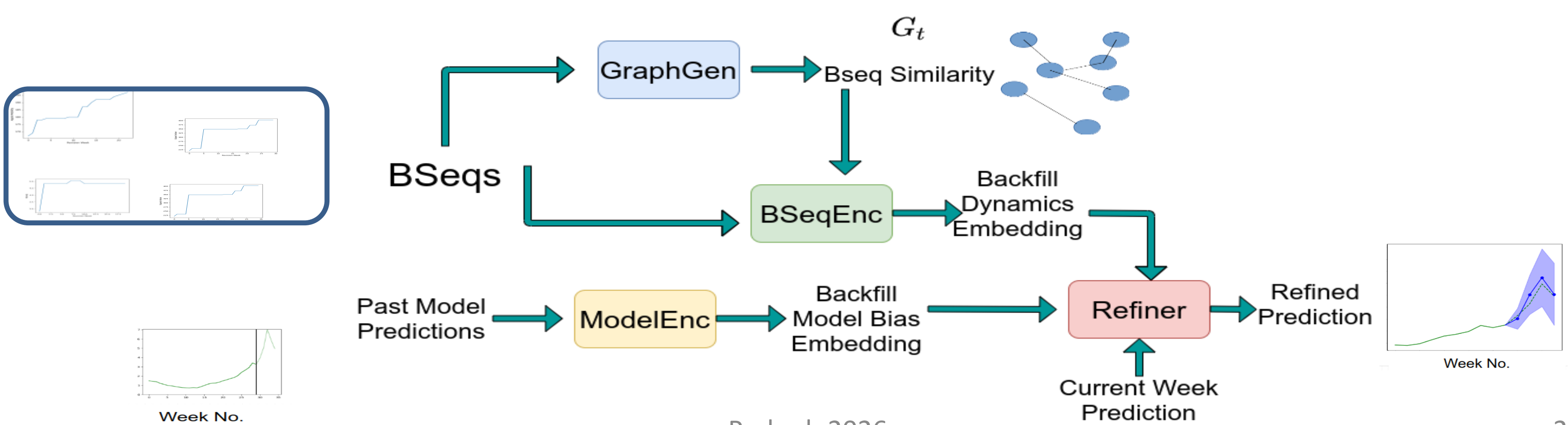
Hard to solve!

- Mechanistically extremely hard to model
- Standard statistical methods not expressive enough

- Need a flexible scalable tool like AI

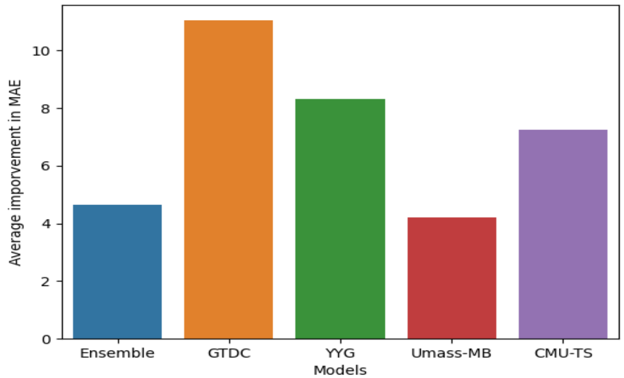
Back2Future = GNNs + Pre-training

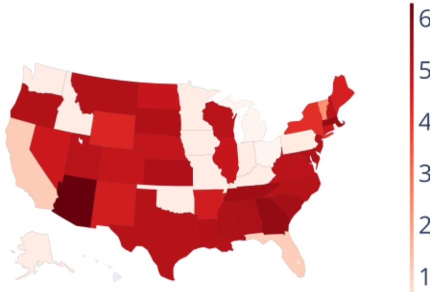
- Deep Learning based Neural Network to solve BRP and ERP
- Learns similarity across signals over all regions - **Graphgen**
- Learns latent representations of Bseqs leveraging similarity and GNNs - **BseqEnc**
- Learns latent representation of Backfill Model bias based on its history of predictions - **ModelEnc**
- Refines current week prediction of model - **Refiner**



B2F improves performance of top models

- Consider some best performing models of **Covidhub** (YYG, Umass-MB, CMU-TS, GTDC) [Cramer et.al 2021]
- Average 6.65% improvement in MAE. Over 10% in 25 states, over 15% in 5 states
- Refined MAE scores of YYG, Umass-MB, CMU-TS, GTDC on par with best individual models.
- Ensemble: Refines scores by 5.18% (2 week ahead) – 3.6% (4 week ahead) with some states over 15%





Average % improvement in MAE scores

* More results in paper

(2) Multimodal Irregular Time-Series Analysis

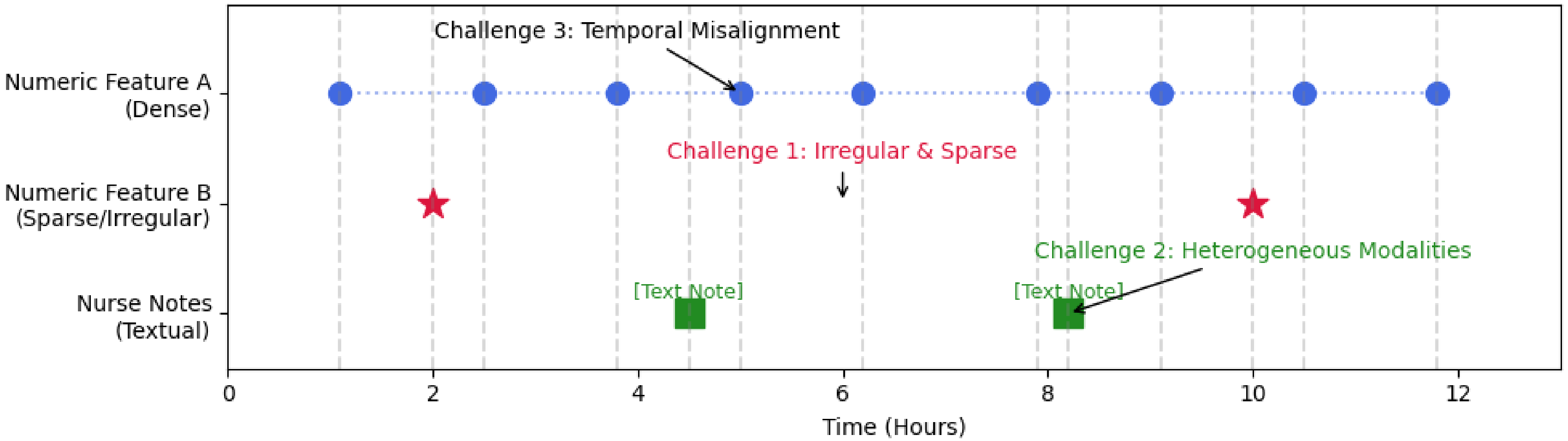
- Background: Children's Healthcare of Atlanta (CHoA) Database
  - > 50k electric health records (EHRs) of admitted children per month
  - Task: predict future vital signs using past vital signs and exogenous features (lab measurements, nurse notes, etc.)

Children's℠ Healthcare of Atlanta

[Shangqing Xu et al, under submission]

Challenges inside CHoA Database

- Using exogenous features is non-trivial
  - Irregular and sparse features
    - Many exogenous signs are presented only once / twice per day
  - Heterogenous modalities
    - Numeric measurements; Textual medical notes



Why is the problem hard to address?

- Unsure distribution of modalities
  - 90% text, 10% num; or 1% text, 99% num
- Useful information can be either between or within modalities
  - “Note: chest pain”, “heart rate: 120”
- Temporal relationship under irregularity matters
  - “10:00 chest pain” ←→ ”10:01 blood pressure: 140” should be captured
  - “10:00 chest pain” ←→ “12:00 blood pressure: 100” should be ignored

Our solution: Multimodal Irregular Seamless Transformer (MIST)

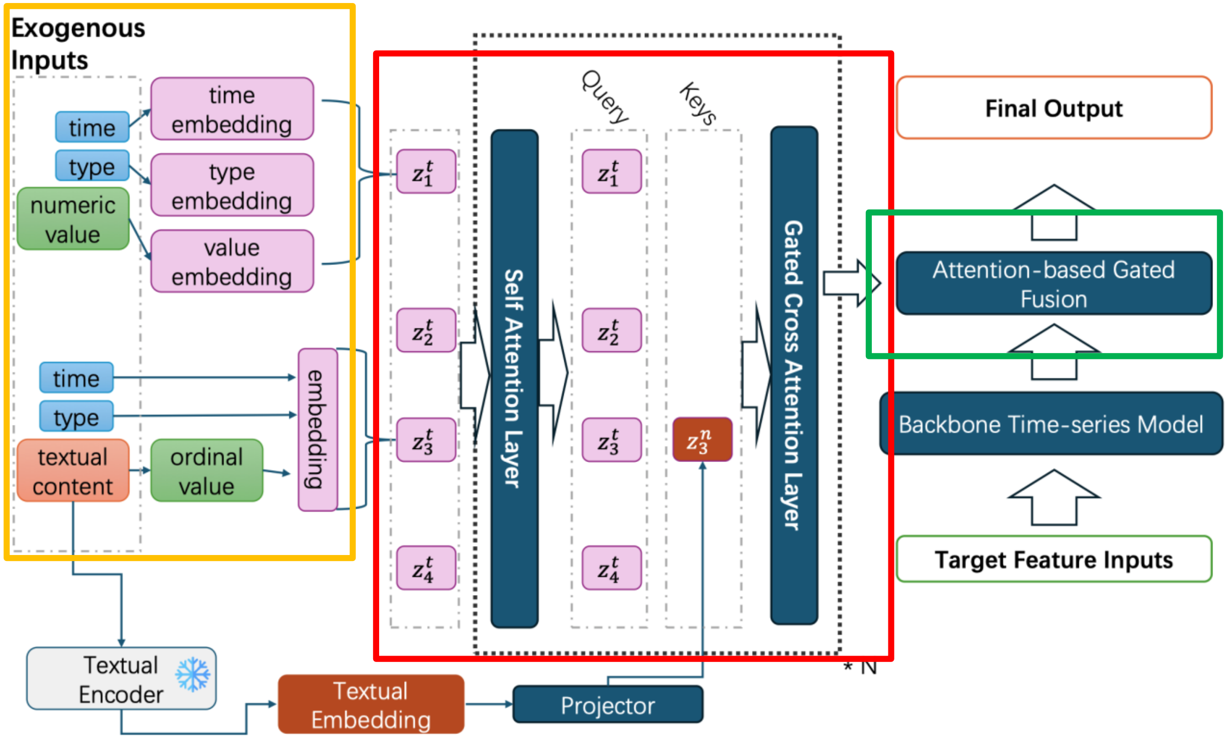
- Core idea: Seamless cross-modality attention
  - "Seamless": no prior assumptions on modality distribution

  - In each block:

1. Self-Attn inside numeric sequence
2. Cross-Attn between numeric (as query) and text (as key)
3. Fuse numeric attn with cross-attn output

Overview of MIST

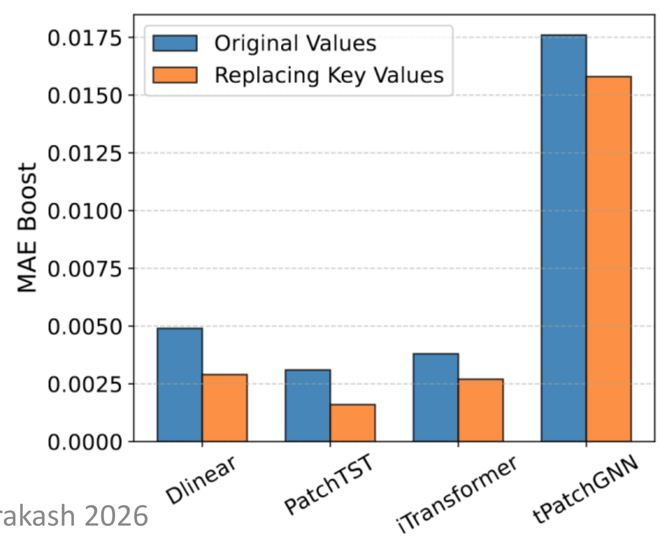
- Core design: **seamless** cross-modality attention
- Dual-type triplet encoding handles irregularity
- Gated-fusion module corrects backbone outputs



Empirical Results

- Dataset: CHoA, MIMIC, and 5 more general purpose multimodal datasets
- MIST improved 4 state-of-the-art unimodal TSF models, more than baseline multimodal framework
- Case study: improvement on CHoA is sensitive to important features verified by doctors

Backbone TSF Model	Exogenous Encoding	ChoA		MIMIC	
		MAE	MSE	MAE	MSE
DLinear	-	.5040	.5286	.4820	.4822
	IMM-TSF	.5035	.5283	.4832	.4605
	MIST(Ours)	**.4991**	**.5270**	**.4722**	**.4563**
PatchTST	-	.4881	.5085	.4838	.4692
	IMM-TSF	.4883	.5095	.4829	.4677
	MIST(Ours)	**.4850**	**.5084**	**.4804**	**.4631**
iTransformer	-	.5003	.5362	.4904	.4767
	IMM-TSF	.4987	.5325	.4910	.4779
	MIST(Ours)	**.4965**	**.5287**	**.4876**	**.4645**
tPatchGNN	-	.5035	.5205	.4937	.4899
	IMM-TSF	.4919	.5203	.4923	.4750
	MIST(Ours)	**.4859**	**.5077**	**.4901**	**.4682**



(3) Severe Heatwaves

- Severe heatwaves are one of the major causes for unnatural deaths in urban area
- The 2003 European heatwave: over 40,000 deaths[1] , including approximately 2,000 in the Netherlands and 3,000 in Spain

[1] García-Herrera, Ricardo, et al. "A review of the European summer heat wave of 2003." Critical Reviews in Environmental Science and Technology 40.4 (2010): 267-306.

Impact of Severe Heatwaves

- Exposure to deadly heatwaves has great threat to organ damage [2]
- This calls for early warning of incoming deadly heatwaves

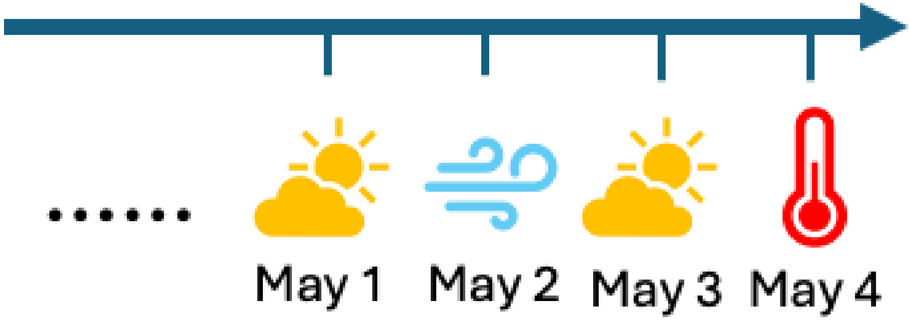
Table. Organs Damaged by Physiological Mechanisms Triggered by Heat Exposure

	Mechanisms				
Organs	Ischemia	Heat Cytotoxicity	Inflammatory Response	Disseminated Intravascular Coagulation	Rhabdomyolysis
Brain	①	⑦	⑬	⑳	
Heart	②	⑧	⑭		
Intestines	③	⑨	⑮	㉑	
Kidneys	④	⑩	⑯	㉒	㉕
Liver	⑤	⑪	⑰	㉓	㉖
Lungs		⑫	⑱	㉔	㉗
Pancreas	⑥		⑲		

[2] Mora, Camilo, et al. "Twenty-seven ways a heat wave can kill you: deadly heat in the era of climate change." Circulation: Cardiovascular Quality and Outcomes 10.11 (2017): e004233.

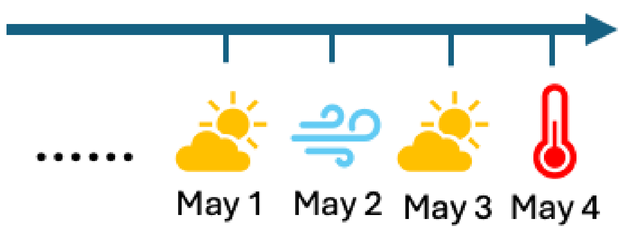
Predicting Heatwaves is Easy

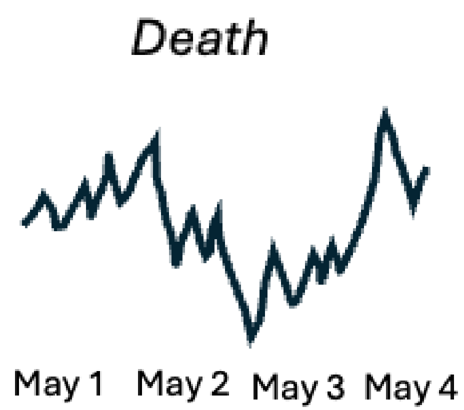
- Heatwaves: days with high temperature
- Predicting heatwaves → temperature / weather prediction
  - Plenty of data and good methods



Predicting Deadly Heatwaves is Hard

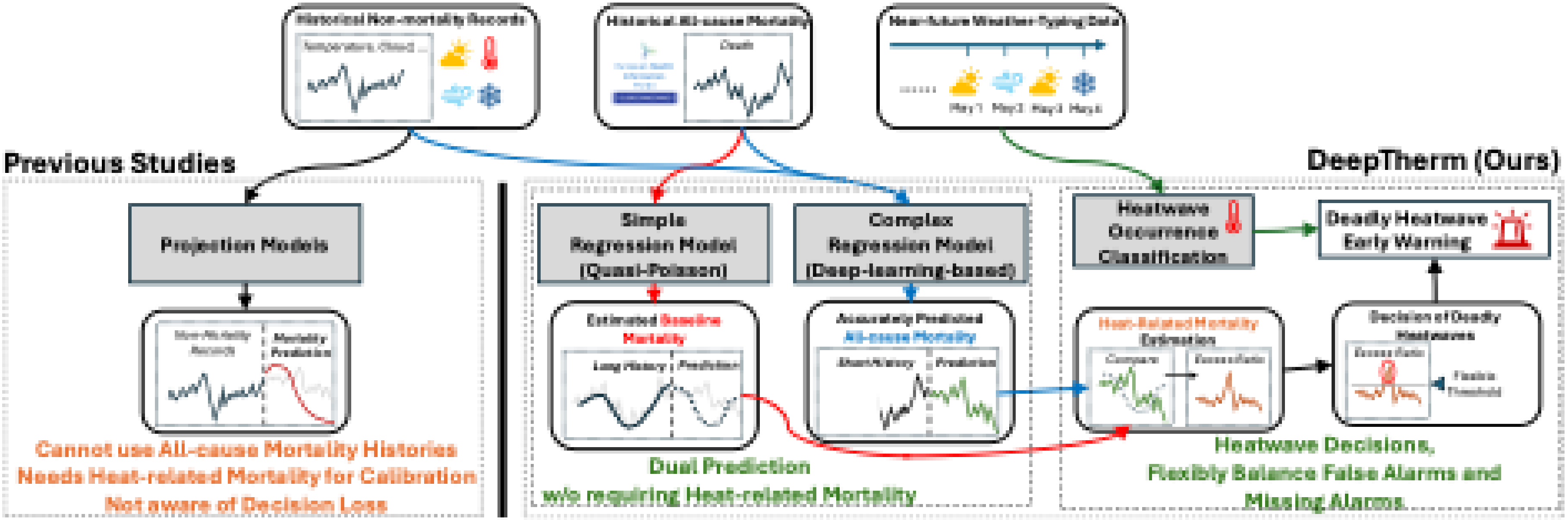
- Deadly heatwaves: heatwaves that causes **high excess mortality**
  - If no such heatwaves, many people can be saved
- This is hard!





Heat related Mortality

[Shangqing Xu et al, in review]



(3) LPTM: Large Pre-trained Time-series Model

[Harshavardhan Kamarthi and B. Aditya Prakash, Neurips 2024]

Diverse Multi-domain time-series datasets

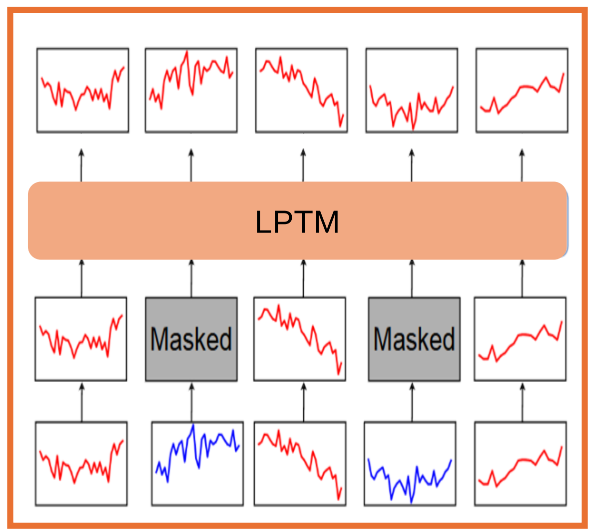
Finance

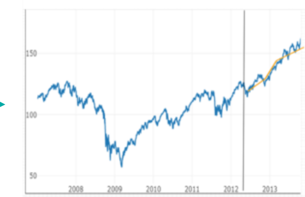
Retail

Economics

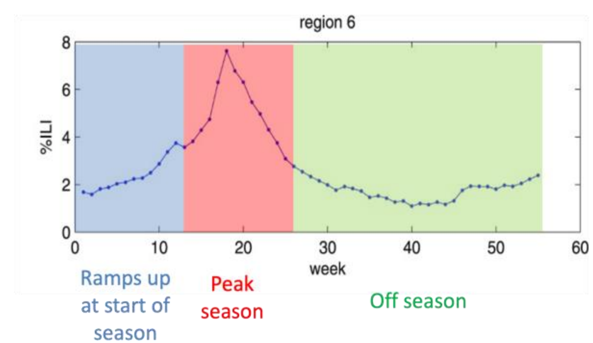
Epidemiology

Foundational Model





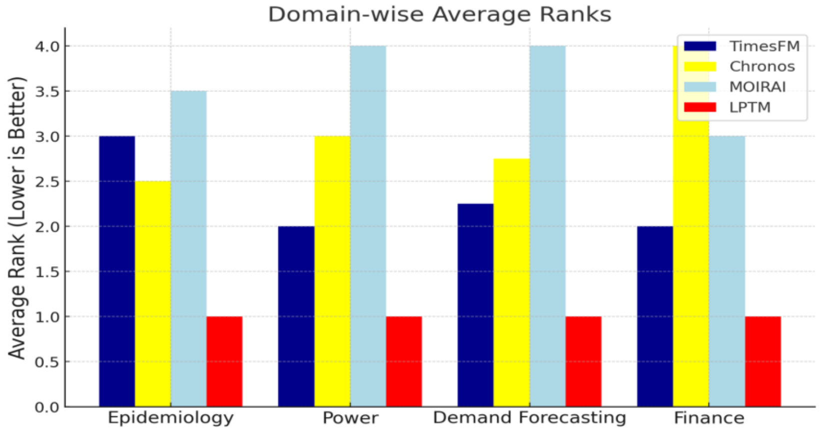
Secret Sauce: Adaptive Segmentation

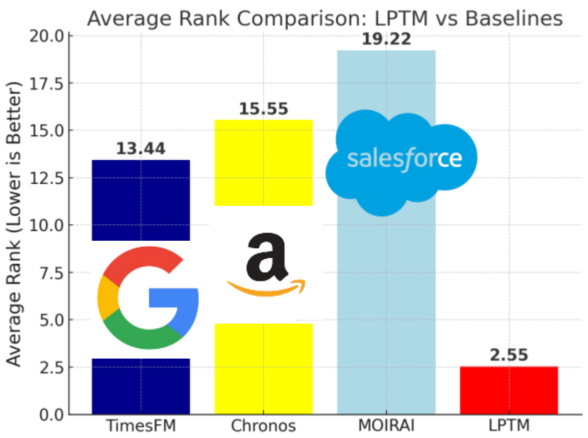


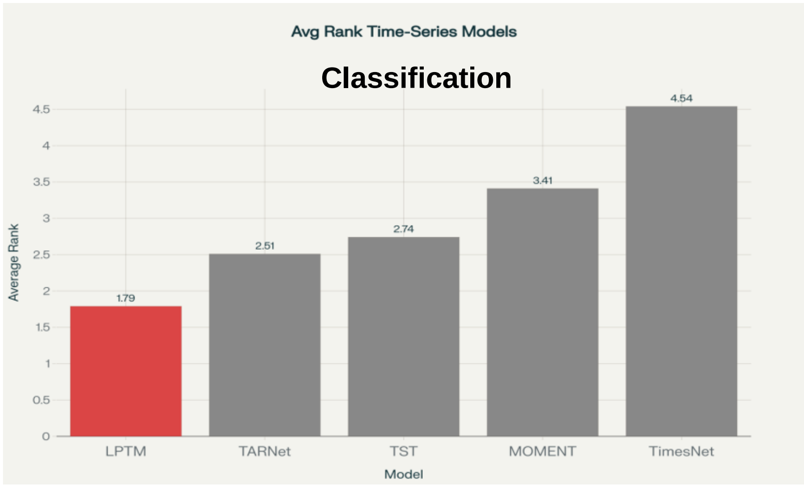
State-of-the-art accuracy

State-of-art Accuracy

Across domains like demand, retail, finance, health





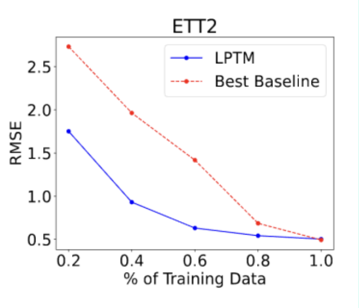


Lower the better!

Data and Compute Efficiency

High Data Efficiency

Model requires little data (few-shot) to achieve state-of-art performance



Needs 2/3rds data to be SOTA

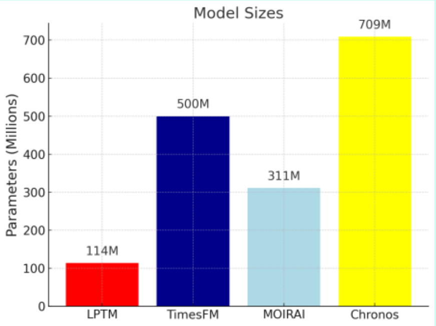
High Compute Efficiency

Fine-tine and inference on laptop-grade GPUs

2-6x smaller than competitors

Model Card

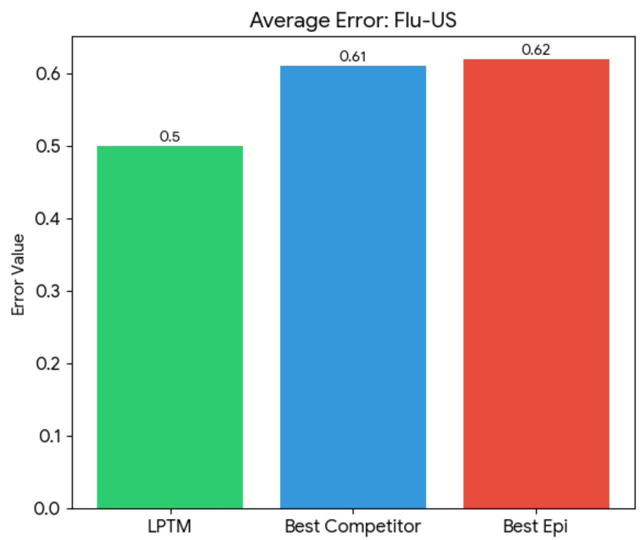
- 114M Parameters
- Inference memory: 1.4 GB
- Fine-tune memory: 1.6 GB
- Pre-train memory: 3.3 GB

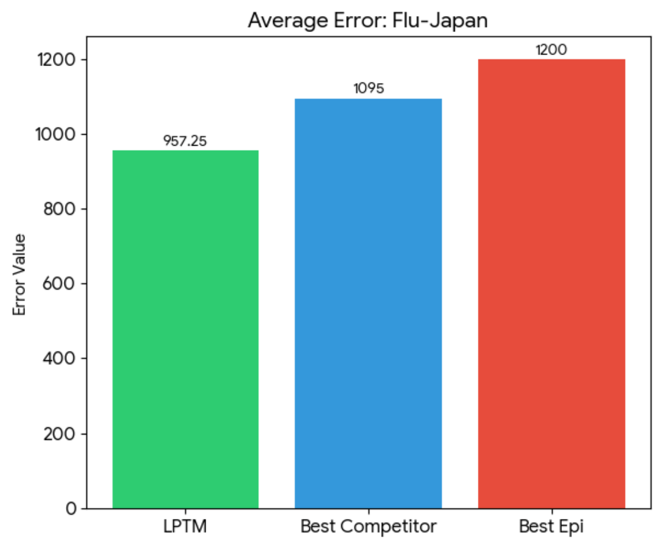


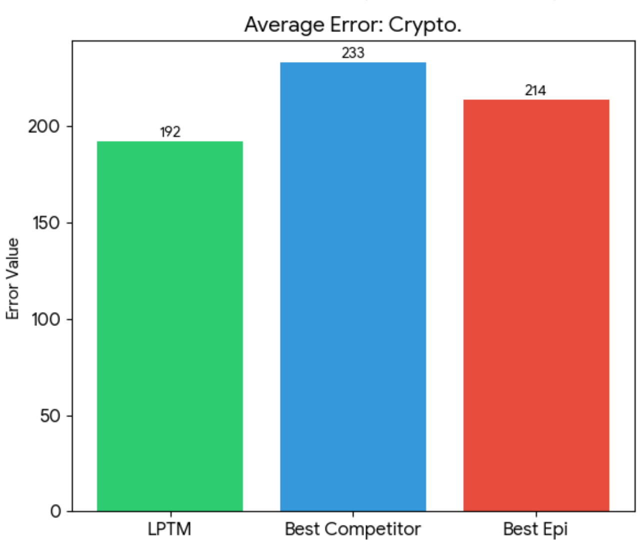
Georgia Tech

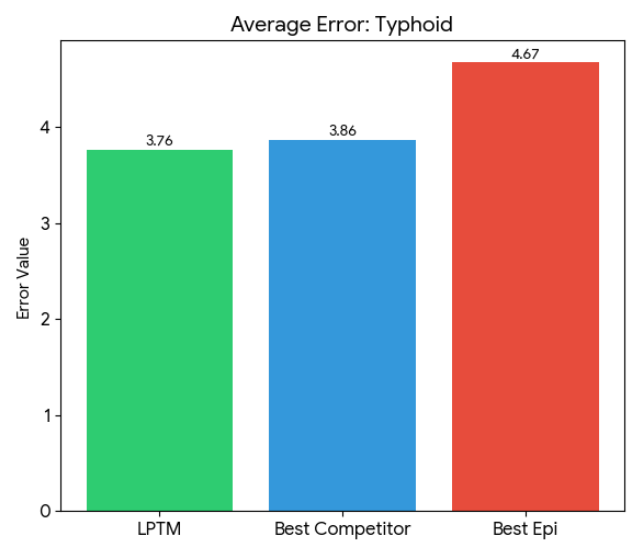
Performance on Epi tasks

Used Past data from Project Tycho as part of pre-training dataset









Samay: Model Library

Zero-shot. Multi-domain.

Pre-training Foundational Time-series Models (FTSM)

📊 Diverse Multi-domain time-series datasets

🔬 Self-supervised Pre-training Foundational Model

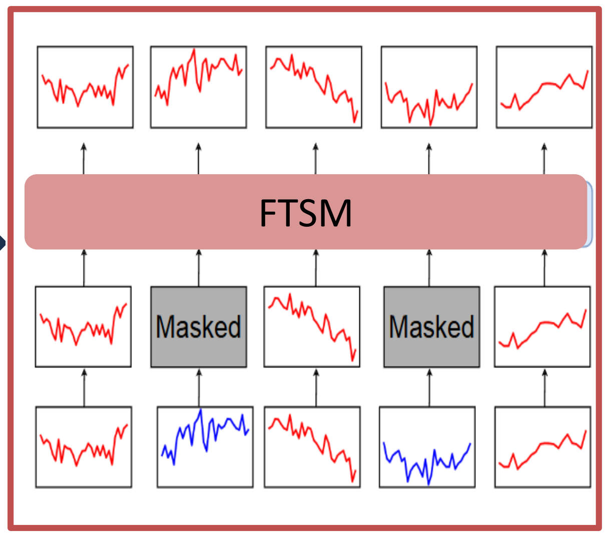
Nasdaq

Finance

Retail

Economics

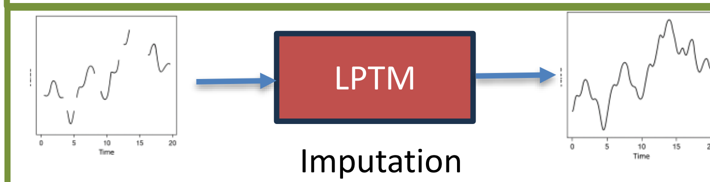
Epidemiology



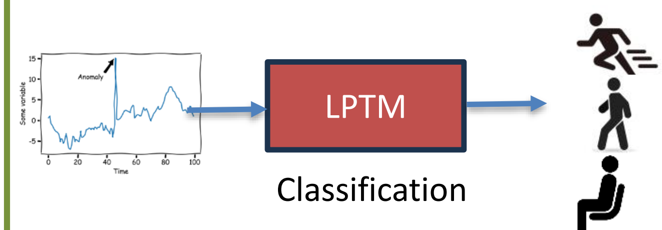
Applications

Zero-shot Forecasting

- No training Data required
- Adaptable to multiple domains



Imputation



Classification

Downstream time-series analysis

- Low data requirement
- Efficient compute time for fine-tuning

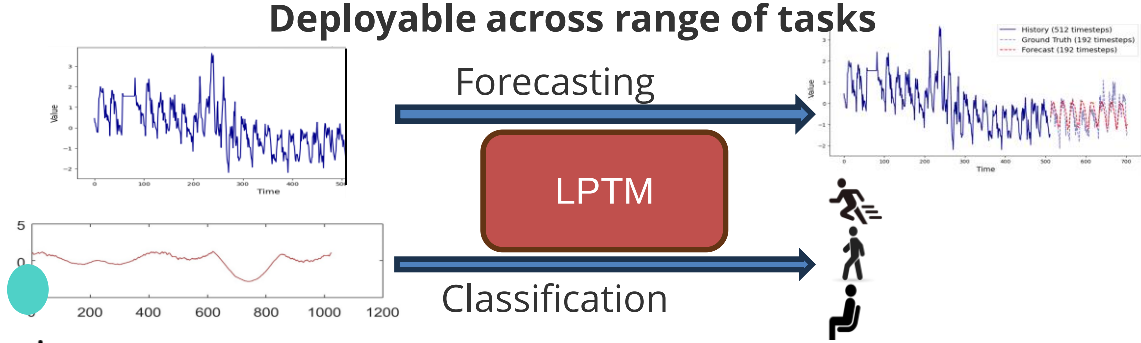
https://github.com/AdityaLab/Samay

Samay: Easy Usage across tasks

Turnkey deployment

- Fine-tune to diverse applications, tasks
- Minimal data engineering
- No Hyperparameter tuning

Deployable across range of tasks



Raw time-series to reliable predictions in 5 lines of code!

1 **Initialize Model**

```
config = {
    "task_name": "forecasting2",
    "forecast_horizon": 192,
    "seq_len": 1112,
    "head_dropout": 0,
    "weight_decay": 0,
    "max_patch": 16,
    "freeze_encoder": False,  # F
    "freeze_embedder": True,  # F
    "freeze_head": True,  # The l
    "freeze_segment": True,  # Fr
}
model = LPTMModel(config)
```

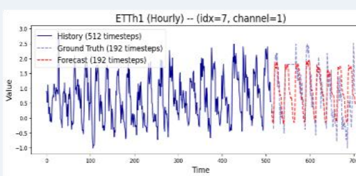
2 **Quick Fine-tune**

```
val_dataset = LPTMDataset(
    name="ett",
    datetime_col="date",
    path="../data/data/ETTh1.csv",
    horizon=96,
    task_name="forecasting2",
    mode="test",
)
```

```
finetuned_model = model.finetune(train_dataset)
```

3 **Generate Forecast**

```
avg_loss, trues, preds, histories = model.evaluate(
    val_dataset, task_name="forecasting2"
)
```



THREAD 2

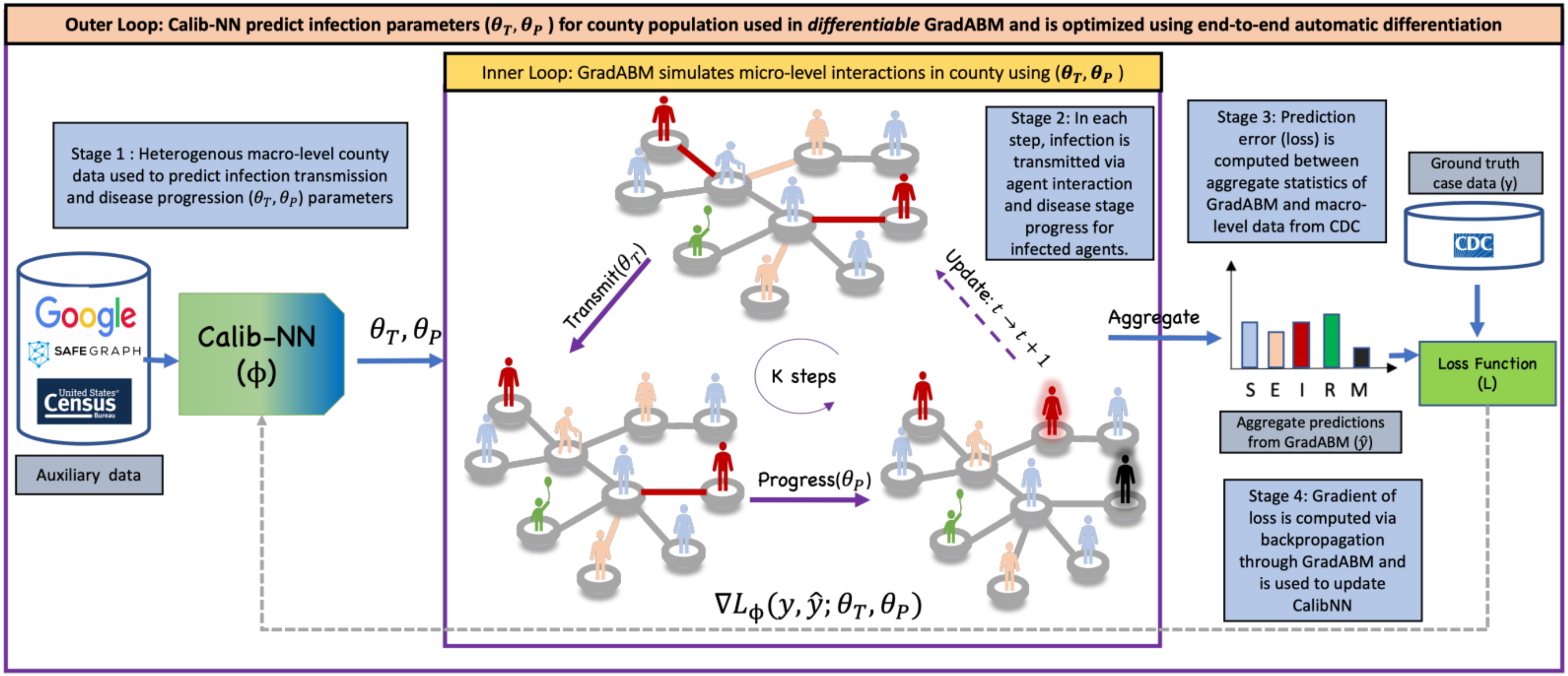
DIFFERENTIABLE EPIDEMIOLOGY

End-to-end Differentiable Learning w/ ABMs

[Chopra and Rodríguez+, AI4ABM @ ICML 2022, AAMAS 2023] *Best paper award*

Stage 1: Param. prediction
Encoder-decoder GRU (Calib-NN)

Stage 2: Disease transmission + progression
Message passing in graph neural network (GNN) + reparametrization trick



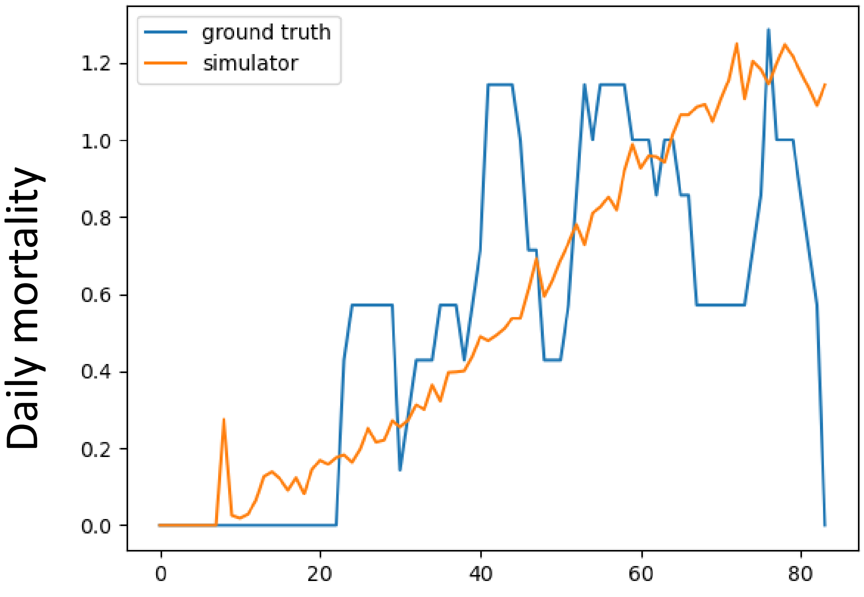
Vaccine Strategy and Policy

[Chopra et al, *Vaccine Jnl.* 2023]

- Incorporating voting data to inform preventive measures
- COVID-19 mortality
- Olmsted county, MN

MAYO CLINIC

Fitting ground truth data



*Approx. Oct 2020 to Jan 2021

Georgia Tech

[Nature Digital Medicine 2025, ICHE 2024, SIAM DM 2024 (best poster award)]

Jiaming Cui
GT -> VT

HAI importation cases

- Patients who have already been infected when admitted into hospital
- An important source of HAIs: Contributes to 13% of cases in Germany, 18.9% in Spain

Source: ecdc EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

JOHNS HOPKINS MEDICINE

UNIVERSITY of VIRGINIA

HAI importation cases identification

- Non-trivial: Many importation cases may show no symptoms (i.e., asymptomatic)

- Surveillance tests
  - PCR tests
  - Culture tests

PCR tests

Culture tests

Georgia Tech

HAI importation cases identification

- Non-trivial: Many importation cases may show no symptoms (i.e., asymptomatic)

- Surveillance tests: PCR & culture tests

- Still problematic
  - Expensive to test everyone: ~$20 per PCR test [1]
    - ~8k patients in UVA hospital every year: 160k per year
    - Does not include the price of more devices, technicians…
  - Unavoidable delays

[1] Roth et al. Cost Analysis of Universal Screening vs. Risk Factor-Based Screening for MRSA. PLOS One 2016.

Nosocomial infections

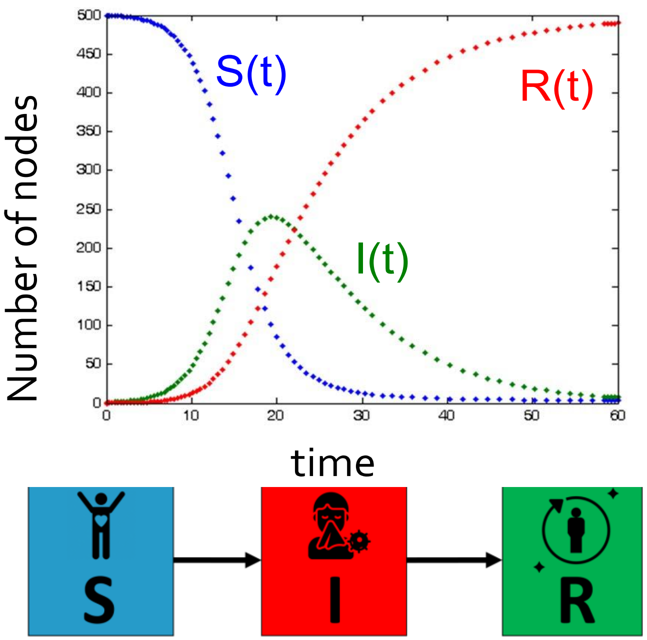
- The HAI importation cases also lead to many nosocomial infections
- Nosocomial infections: Patients get infected in hospital.
- Why we also want to identify them: To prevent further HAI spread

Current methods

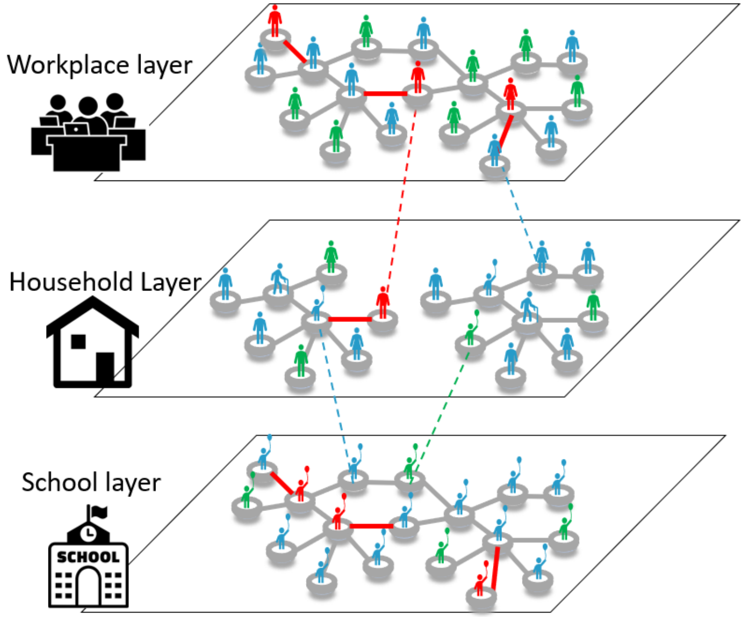
- Current methods can be classified into two classes
  - Modelling based methods
  - Machine learning (ML) based methods

Modelling based methods

- Type:
  - ODE models
  - Metapopulation models
  - Agent-based models



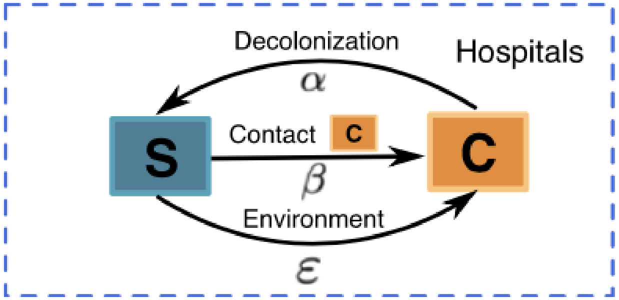
ODE models



Agent-based models

Modelling based methods: Steps

1. Build epidemiological models to describe HAI spread.

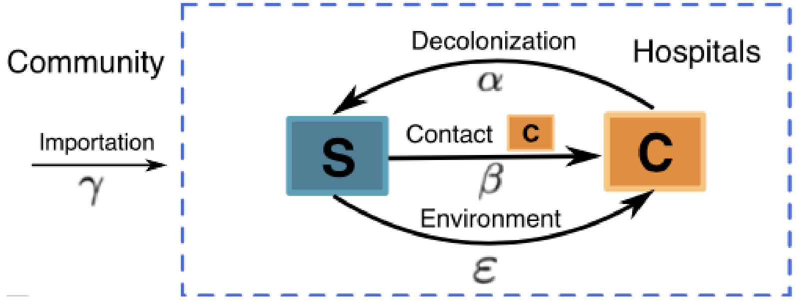


[1] Sen et al. Identifying asymptomatic spreaders of antimicrobial-resistant pathogens in hospital settings. PNAS 2021.

Modelling based methods: Steps

1. Build epidemiological models to describe HAI spread.

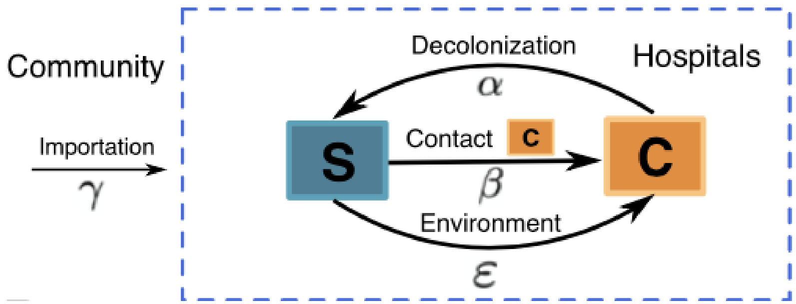
2. Add parameters for importation probability.

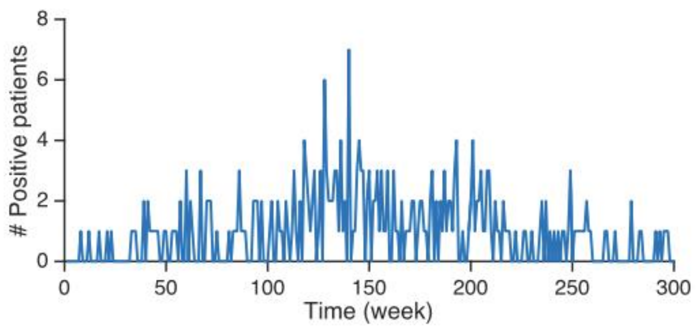


[1] Sen et al. Identifying asymptomatic spreaders of antimicrobial-resistant pathogens in hospital settings. PNAS 2021.

Modelling based methods: Steps

1. Build epidemiological models to describe HAI spread.
2. Add parameters for importation probability.
3. Calibrate models to observed data (e.g., daily number of cases) to estimate the parameter values.



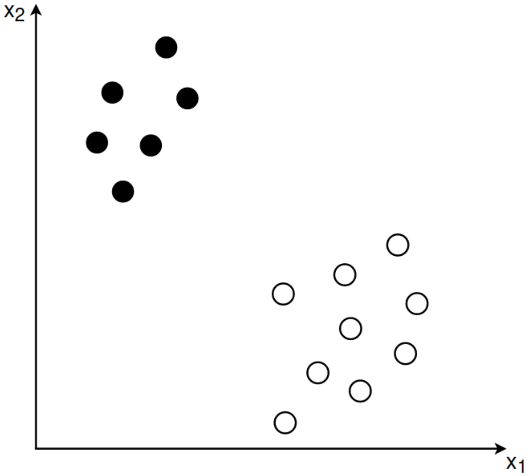


[1] Sen et al. Identifying asymptomatic spreaders of antimicrobial-resistant pathogens in hospital settings. PNAS 2021.

Georgia Tech

ML based methods: Steps

1. Formulate as a supervised classification problem.

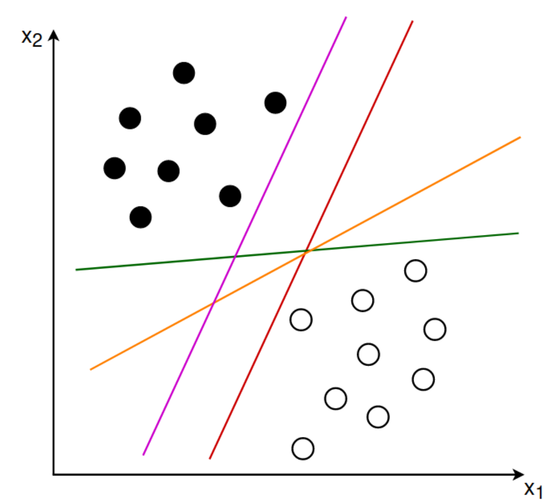


Georgia Tech

ML based methods: Steps

1. Formulate as a supervised classification problem.

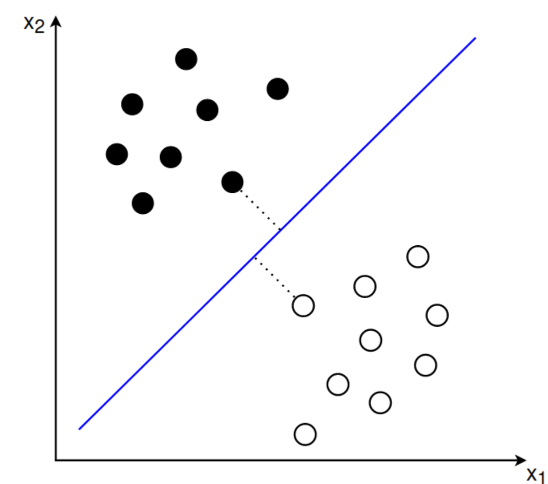
2. Use ground-truth labels (i.e., PCR/culture test results) to train the ML method until converge.



Georgia Tech

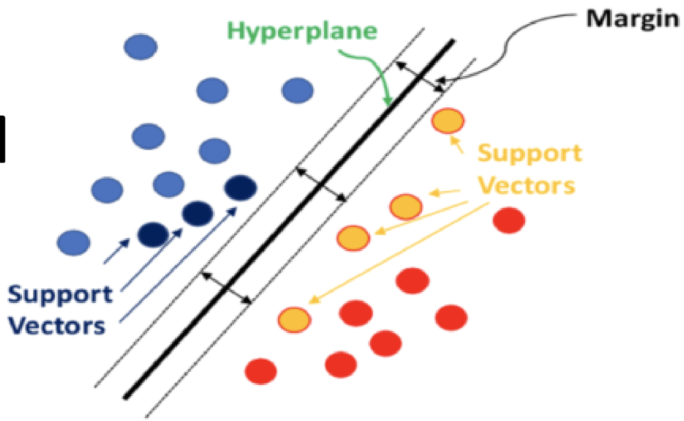
ML based methods: Steps

1. Formulate as a supervised classification problem.
2. Use ground-truth labels (i.e., PCR/culture test results) to train the ML method until converge.
3. Use the trained ML method to predict importation cases and evaluate.

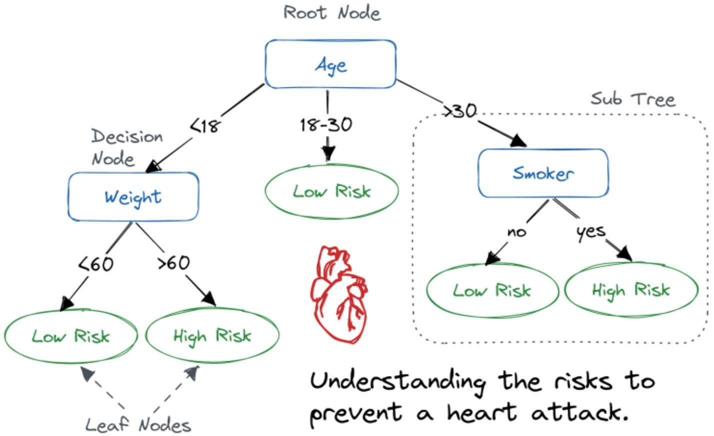


ML based methods

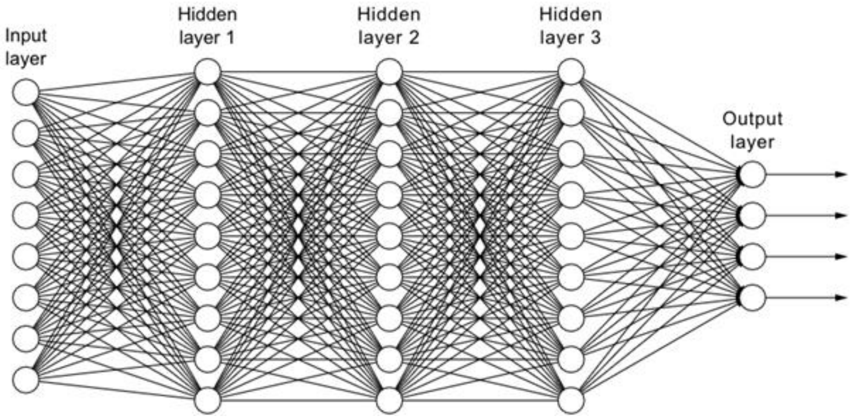
- Examples:
  - Support vector machine (SVM
  - Decision trees
  - Deep neural networks



SVM



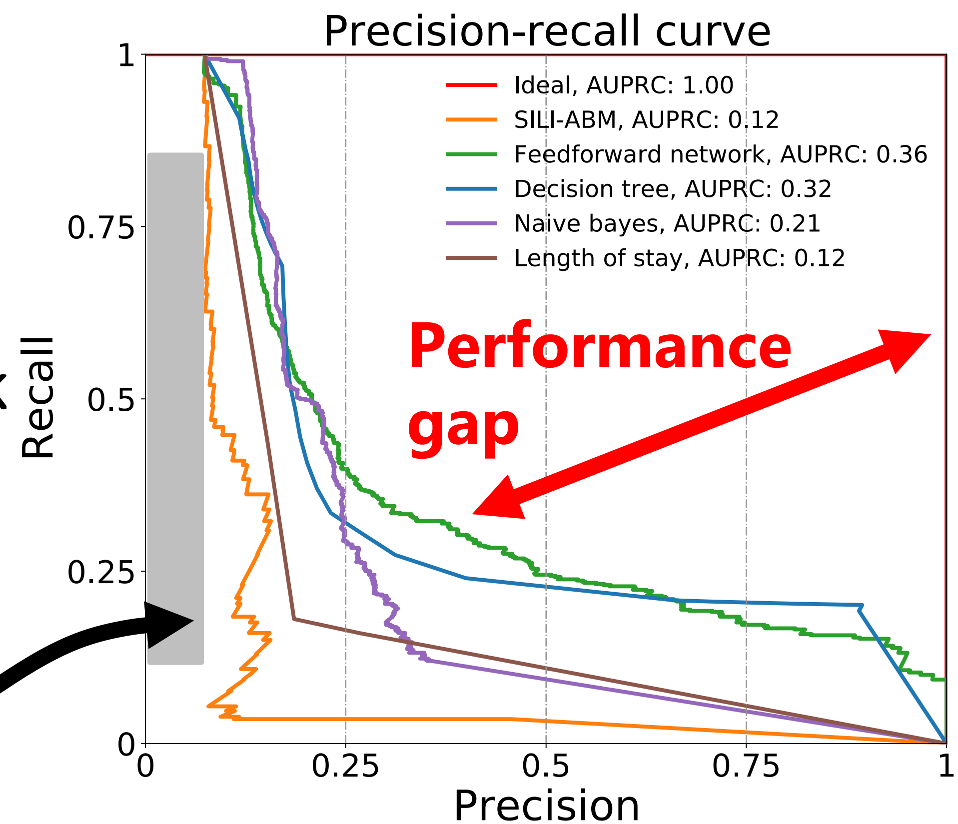
Decision tree



Deep neural networks

BUT: Current methods are not performing well!

- Modelling based methods:
  - SILI-ABM [1]
- ML based methods:
  - Feedforward network
  - Decision tree
  - Naïve Bayes



Clinically inapplicable: Too low precision (i.e., wasting too much testing resources)

[1] Sen et al. Identifying asymptomatic spreaders of antimicrobial-resistant pathogens in hospital settings. PNAS 2021.

Current methods not performing well

- Modelling based methods:
  - SILI-ABM [1]

- ML based methods:
  - Feedforward network
  - Decision tree
  - Naïve Bayes

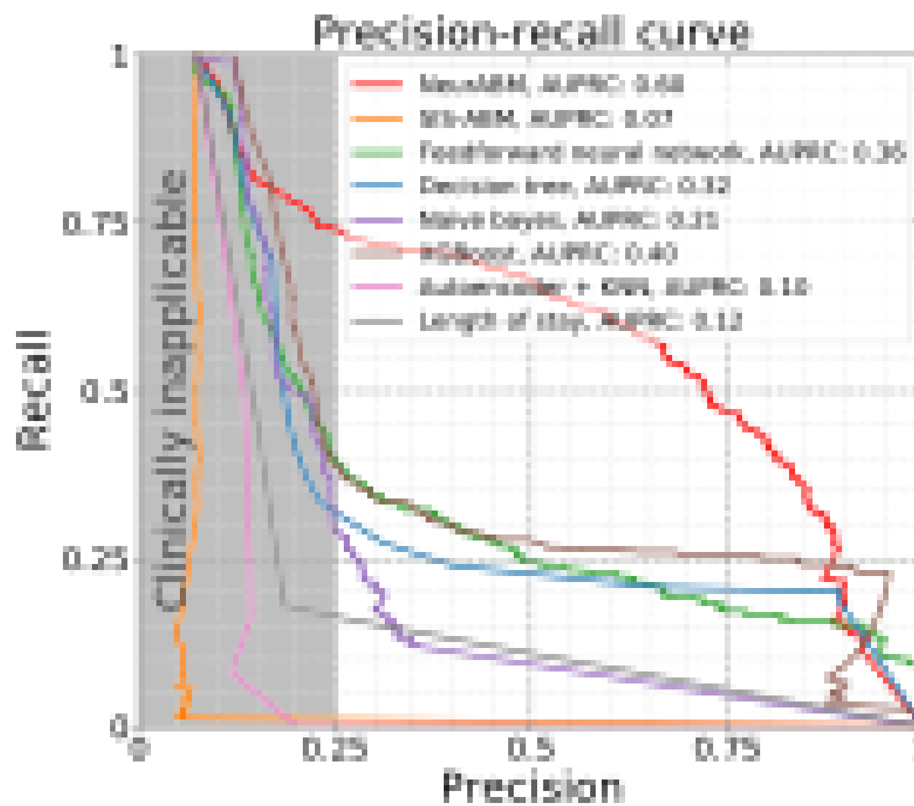
Both methods do not generalize well!

[1] Sen et al. Identifying asymptomatic spreaders of antimicrobial-resistant pathogens in hospital settings. PNAS 2021.

Our idea: NeurABM (Neural Networks + ABMs)

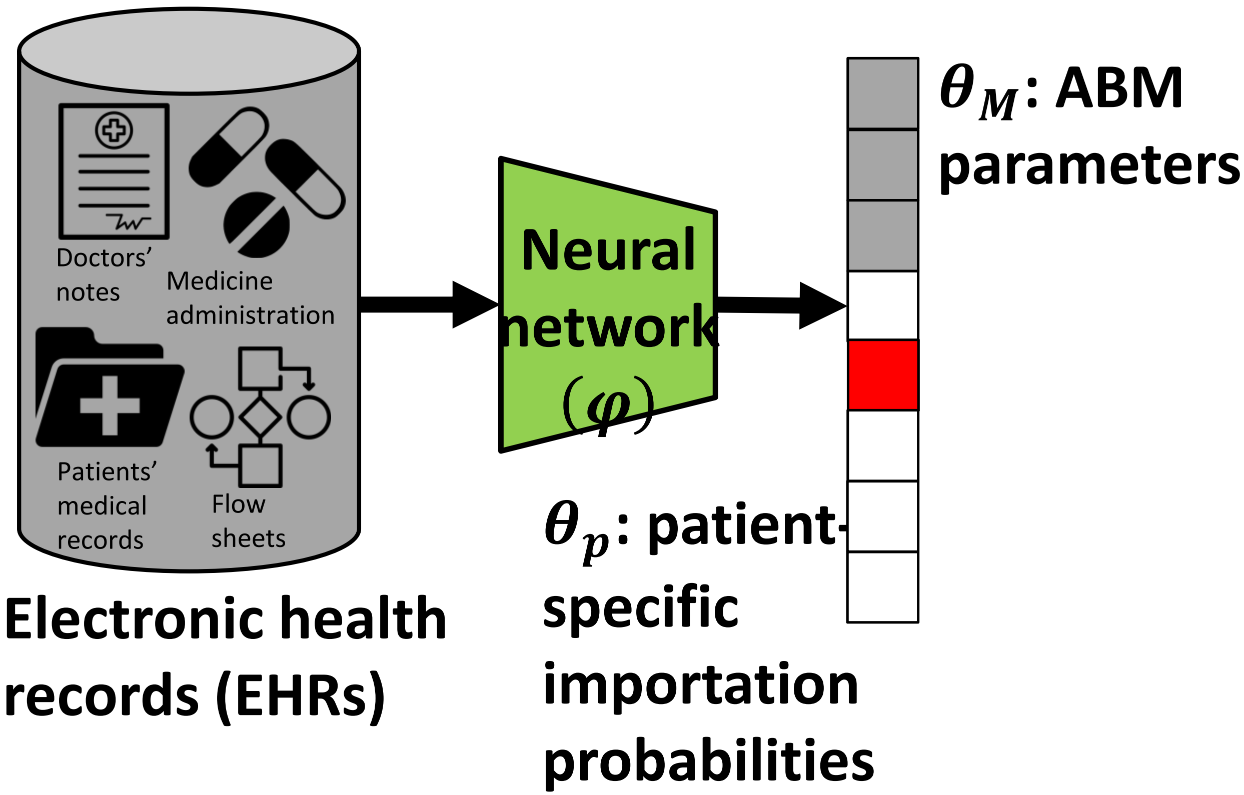
- Incorporate agent-based models
  - Leverage epidemiological knowledge
- Incorporate ML to use different datasets
  - Give patient-specific prediction

UVAHealth



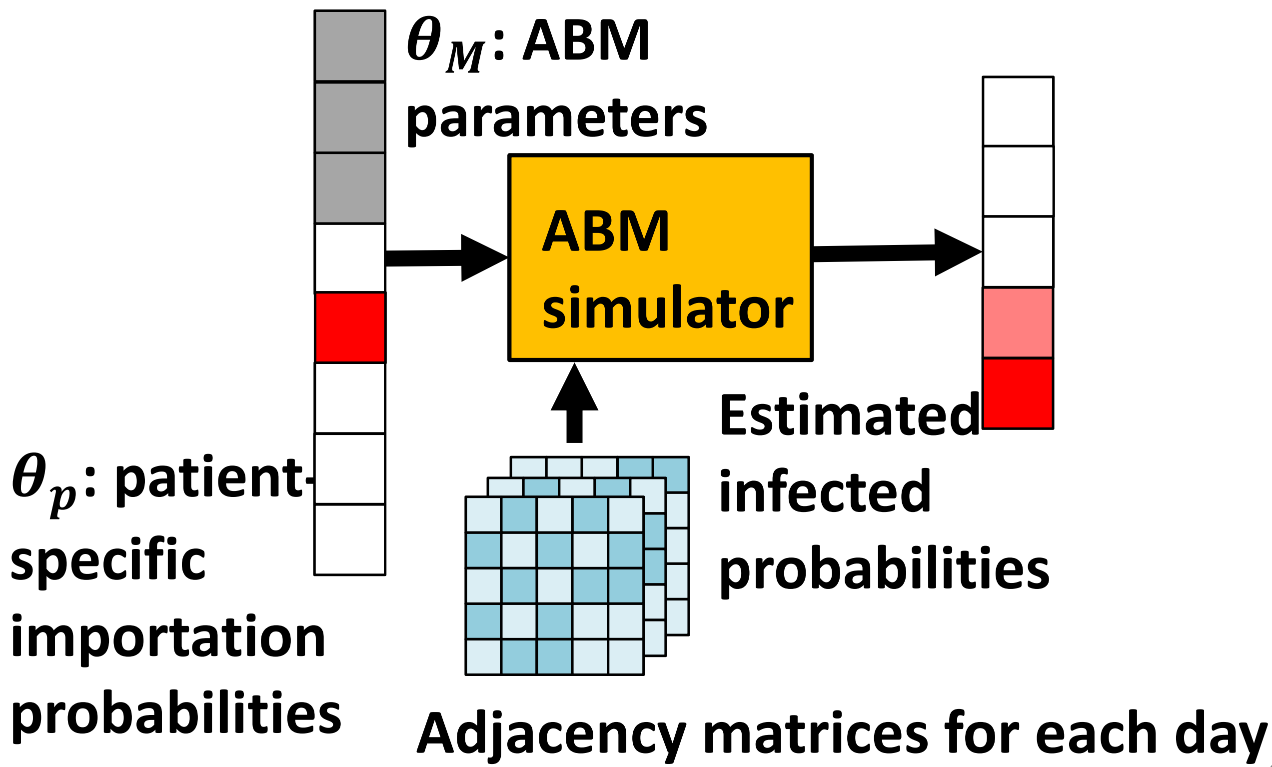
Our framework: Overview

- Neural network part: Leverage patients' features



Our framework: Overview

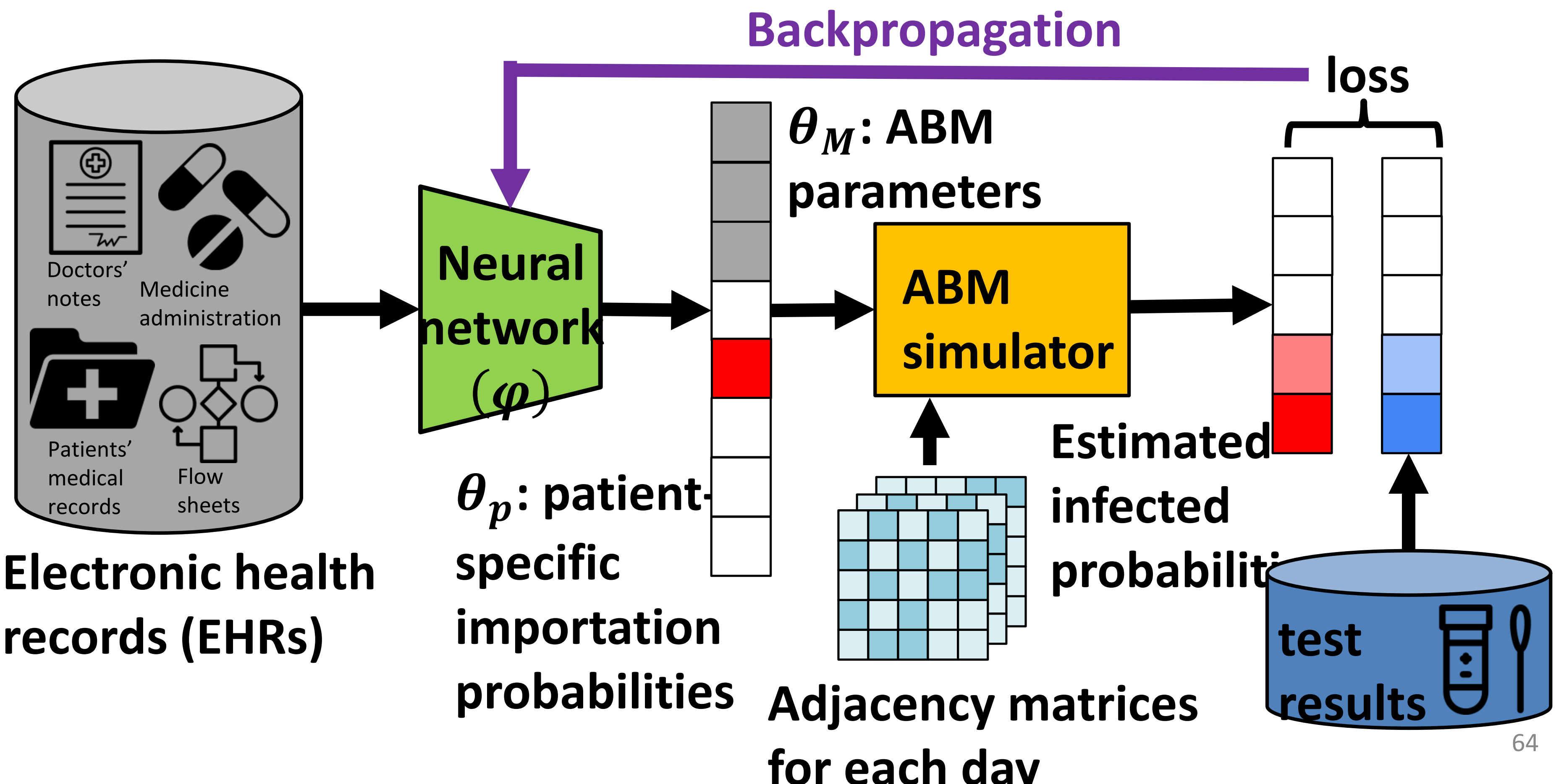
- ABM simulator part: Leverage epi knowledge



Georgia Tech

Our framework: Overview

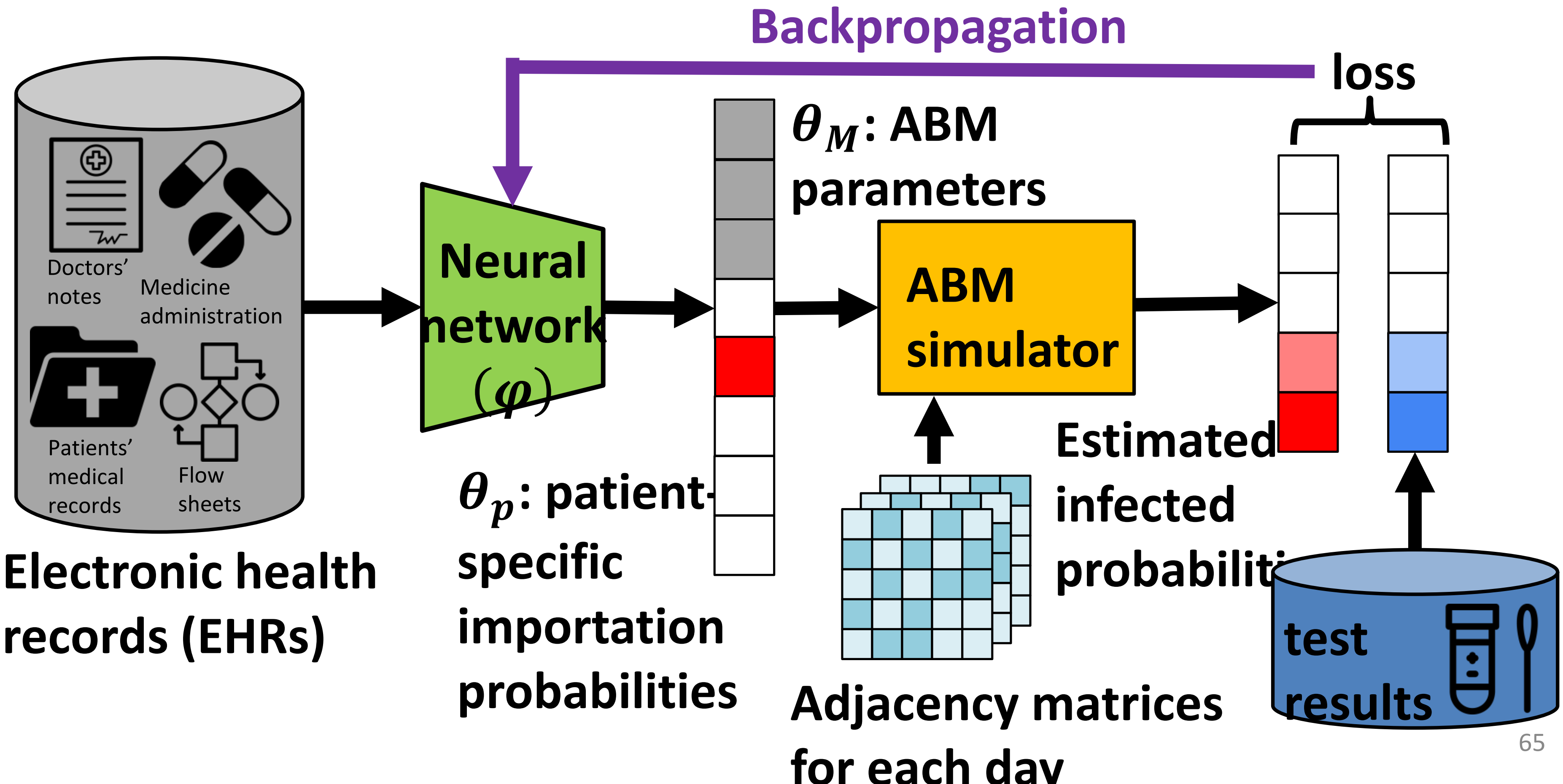
- Pipeline two parts together



Our framework: Overview

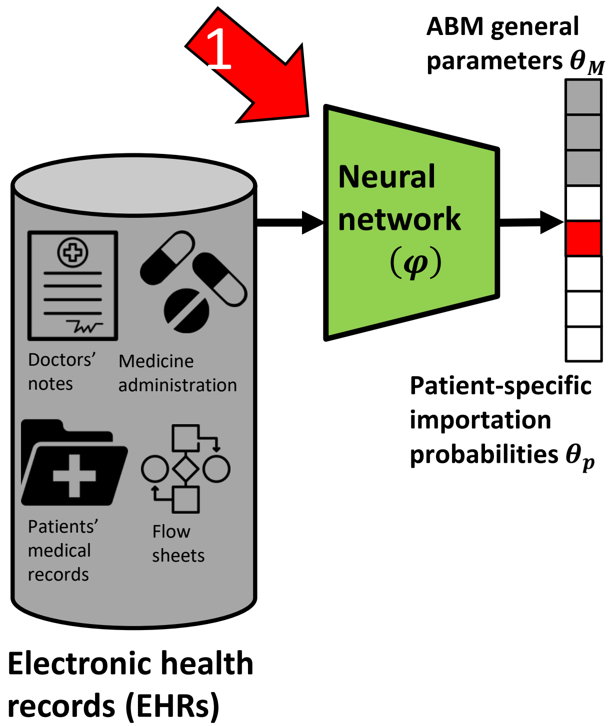
- Pipeline two parts together

ABM simulator needs to be differentiable for backpropagation



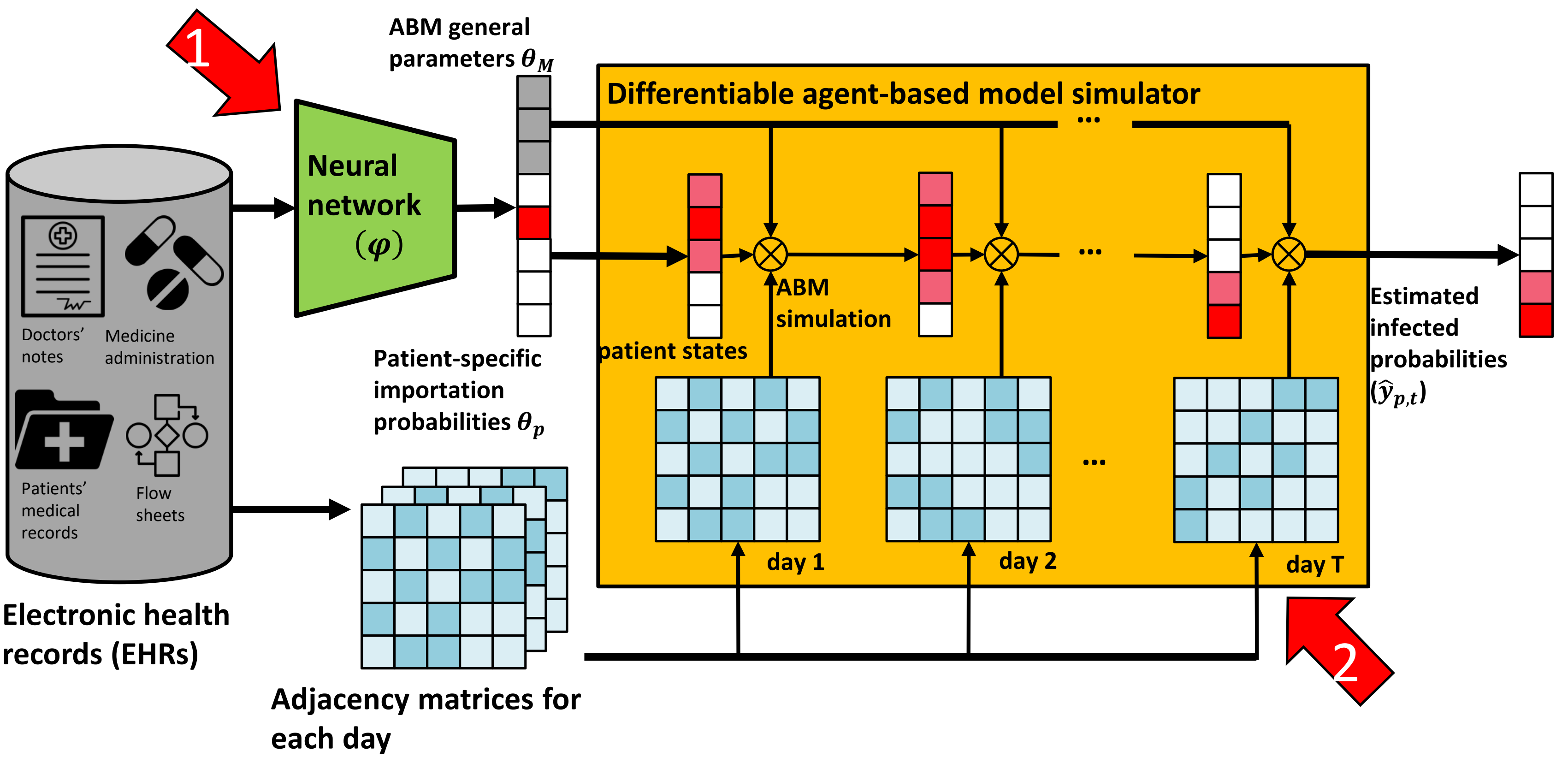
NeurABM: Incorporate NN into ABM

- Our framework



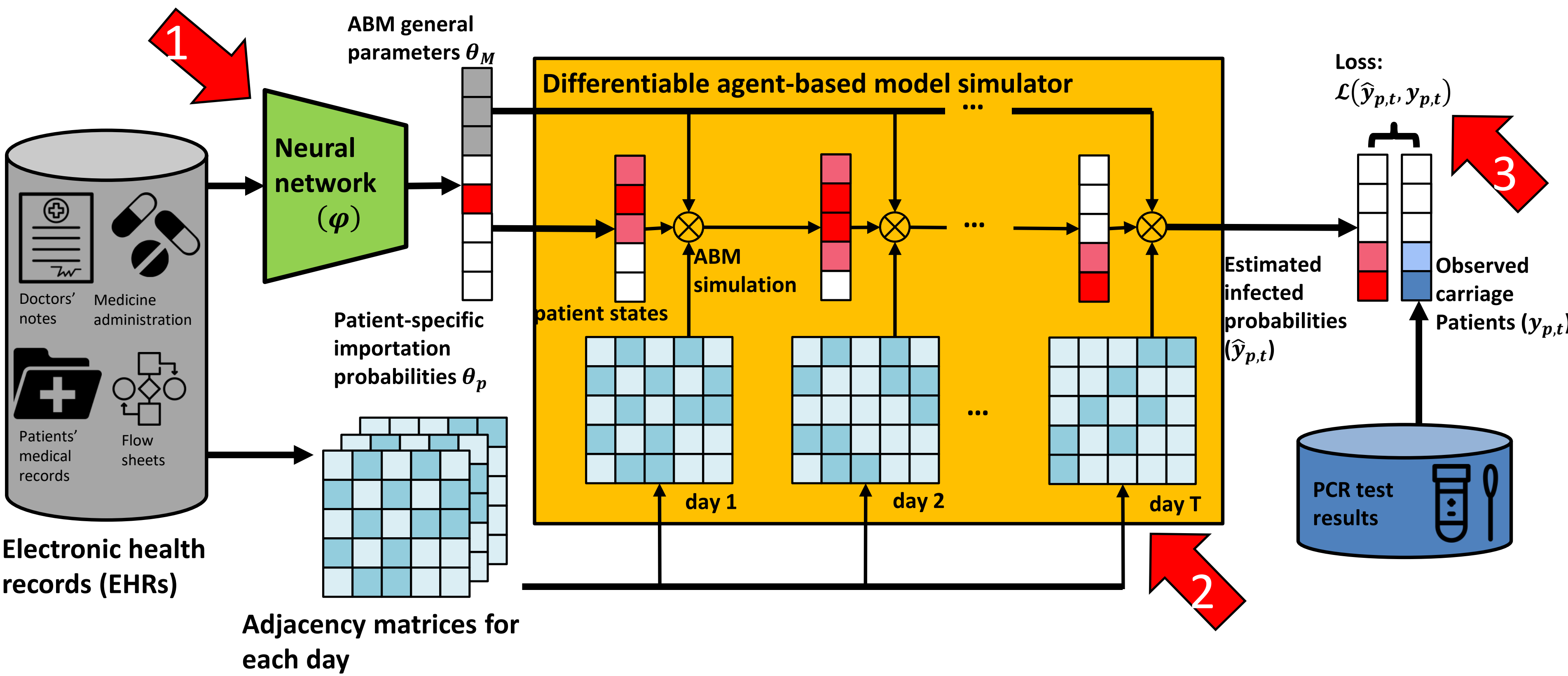
NeurABM: Incorporate NN into ABM

- Our framework



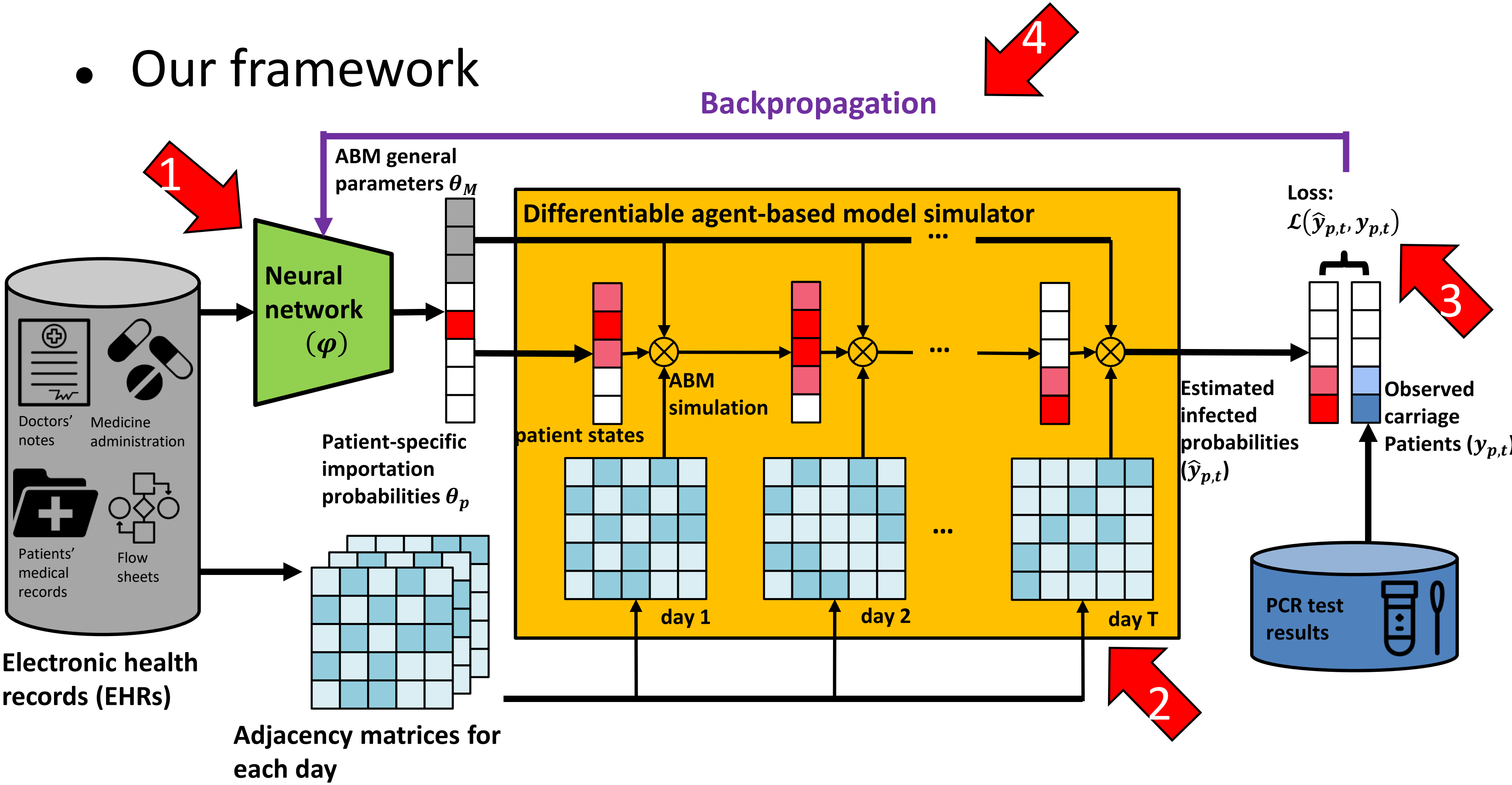
NeurABM: Incorporate NN into ABM

- Our framework



NeurABM: Incorporate NN into ABM

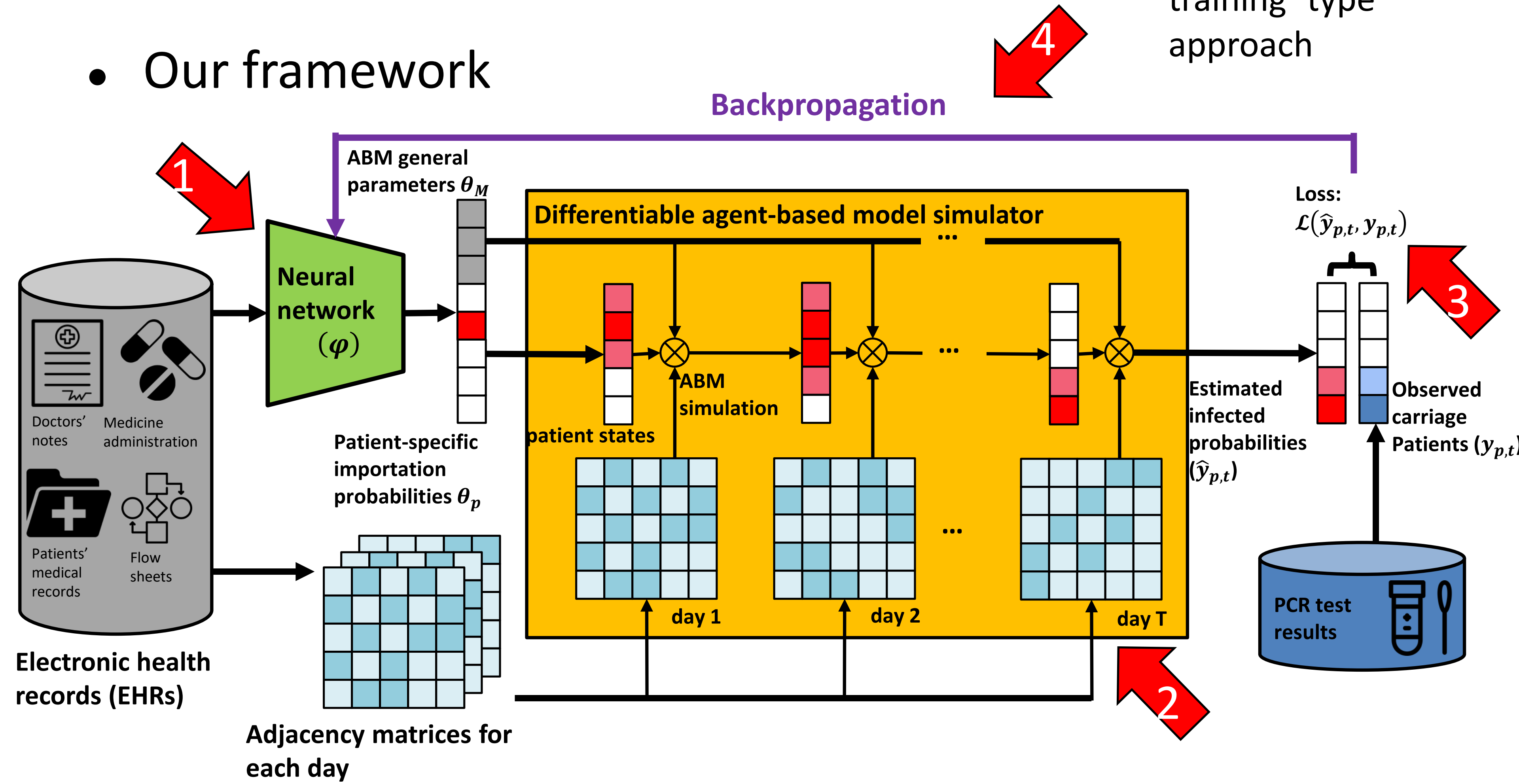
- Our framework



NeurABM: Training Issues

- Deep model (1 layer per day)
- Iterative ‘pre-training’ type approach

- Our framework



Experiment Setup

- Real-world data and tasks

Target 1: Importation cases

- Mechanistic model: **2-Mode-SIS**
- ICU in UVA hospital
- Evaluation on 6 months in 2019

Experiment Setup

- Real-world data and tasks

Target 1: Importation cases	Target 2: Nosocomial infection cases
• Mechanistic model: **2-Mode-SIS** • ICU in UVA hospital • Evaluation on 6 months in 2019	• Mechanistic model: **2-Mode-SIS** • ICU in UVA hospital • Evaluation on 6 months in 2019

Georgia Tech

Experiment Setup

- Real-world data and tasks

Target 1: Importation cases	Target 2: Nosocomial infection cases
• Mechanistic model: **2-Mode-SIS** • ICU in UVA hospital • Evaluation on 6 months in 2019	• Mechanistic model: **2-Mode-SIS** • ICU in UVA hospital • Evaluation on 6 months in 2019

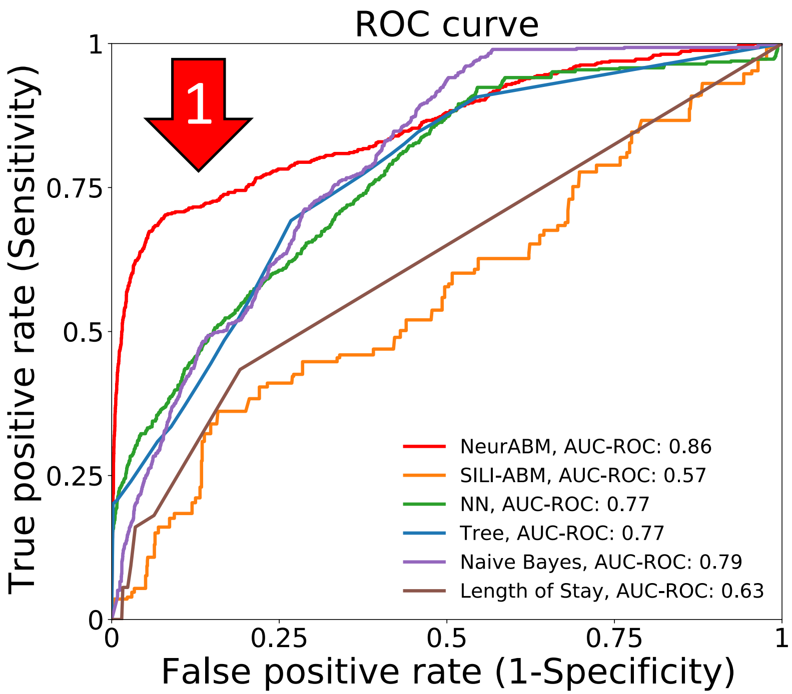
- Baselines:
  - ML categories: NN, decision tree, naïve bayes
  - Modeling categories: SILI model [1]
  - Clinical heuristic categories: Length of stay in hospital

[1] Sen et al. Identifying asymptomatic spreaders of antimicrobial-resistant pathogens in hospital settings. PNAS 2021.

Results

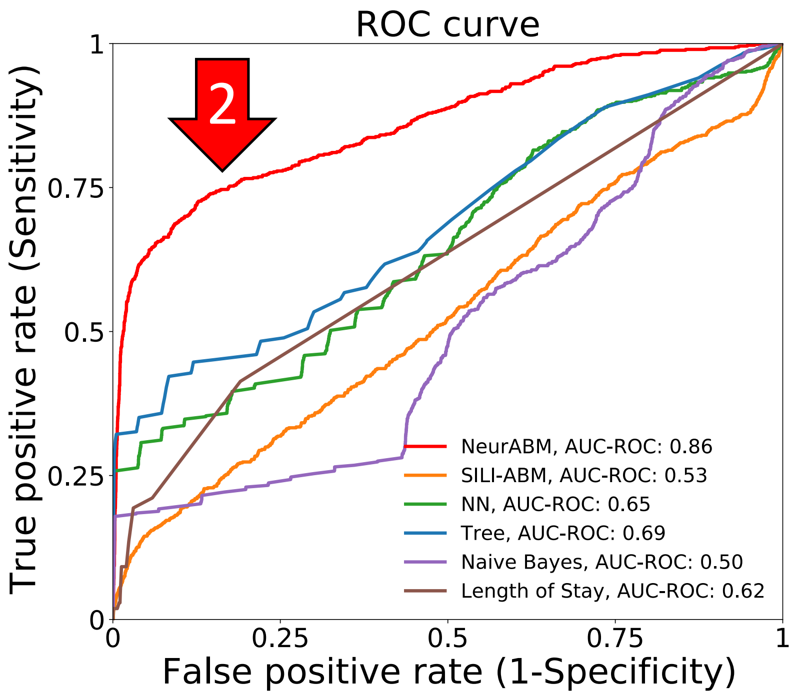
UVA Health

Identify importation cases



Improvement: 9%

Identify nosocomial infection cases



Improvement: 25%

- NeurABM
- ABM (SILI model)
- NN
- Tree
- Naive Bayes
- Length of Stay

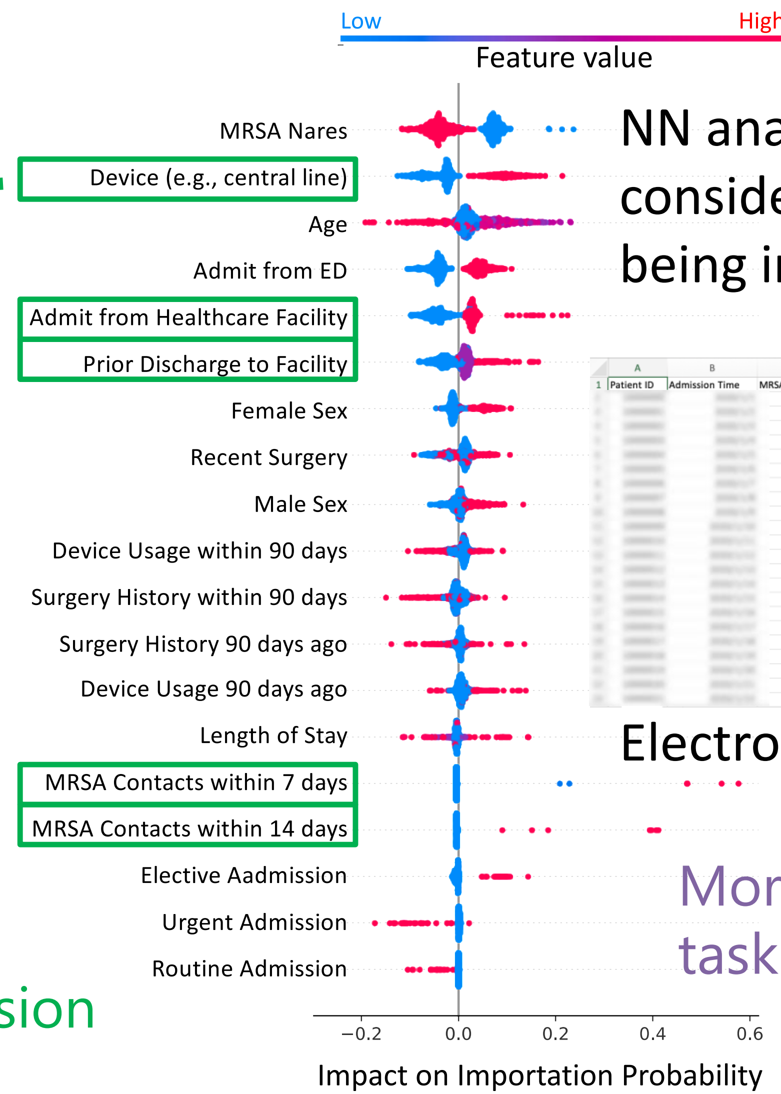
Results (Cont.)

UVA Health

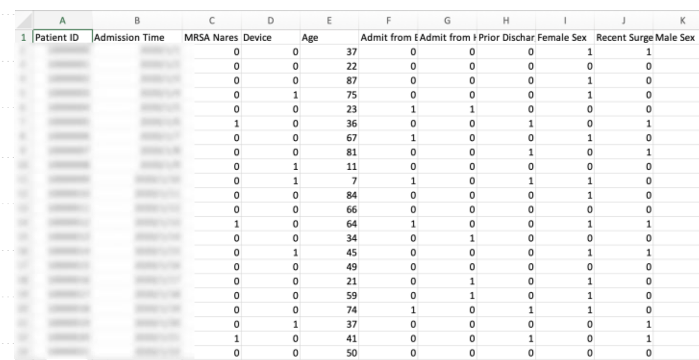
Medical device use history

From/to long-term healthcare facilities

Contacted with MRSA patients before admission



NN analysis: Features considered as high risk for being importation cases



	A	B	C	D	E	F	G	H	I	J	K
1	Patient ID	Admission Time	MRSA Nares	Device	Age	Admit from E	Admit from H	Prior Dischar	Female Sex	Recent Surge	Male Sex
[illegible]	[illegible]	[illegible]	0	0	37	0	0	0	1	1	0
[illegible]	[illegible]	[illegible]	0	0	22	0	0	0	0	0	1
[illegible]	[illegible]	[illegible]	0	0	87	0	0	0	1	0	0
[illegible]	[illegible]	[illegible]	0	1	75	0	0	0	1	0	0
[illegible]	[illegible]	[illegible]	0	0	23	1	1	0	0	0	1
[illegible]	[illegible]	[illegible]	1	0	36	0	0	1	0	1	1
[illegible]	[illegible]	[illegible]	0	0	67	1	0	0	1	0	0
[illegible]	[illegible]	[illegible]	0	0	81	0	0	1	0	1	1
[illegible]	[illegible]	[illegible]	0	1	11	0	0	0	0	0	1
[illegible]	[illegible]	[illegible]	0	1	7	1	0	1	1	0	0
[illegible]	[illegible]	[illegible]	0	0	84	0	0	0	1	0	0
[illegible]	[illegible]	[illegible]	0	0	66	0	0	0	0	0	1
[illegible]	[illegible]	[illegible]	1	0	64	1	0	0	1	1	0
[illegible]	[illegible]	[illegible]	0	0	34	0	1	0	0	0	1
[illegible]	[illegible]	[illegible]	0	1	45	0	0	0	1	1	0
[illegible]	[illegible]	[illegible]	0	0	49	0	0	0	0	0	1
[illegible]	[illegible]	[illegible]	0	0	21	0	1	0	1	0	0
[illegible]	[illegible]	[illegible]	0	0	59	0	1	0	1	0	0
[illegible]	[illegible]	[illegible]	0	0	74	1	0	1	1	0	0
[illegible]	[illegible]	[illegible]	0	1	37	0	0	0	0	1	1
[illegible]	[illegible]	[illegible]	1	0	41	0	0	1	0	1	1
[illegible]	[illegible]	[illegible]	0	0	50	0	0	0	0	0	1

Electronic Health Record

More results for other tasks in the paper

Georgia Tech

WHAT'S NEXT

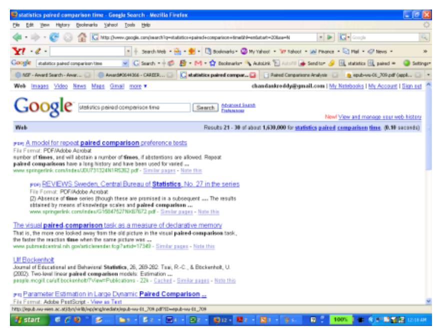
Many open technical questions

- handling different sources of uncertainty,
- behavior?
- bridging with mechanistic ode/agent models: incorporate more granular individualized datasets
- Going beyond prediction, other tasks
- Handling dynamically changing data streams …

• Other settings: hospitals, campus, countries etc.

Increasing data collection

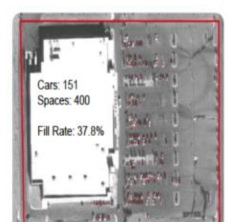
- Genomic
- Mobility
- Point of care
- Line lists
- Surveys
- Social Media
- …



Medical record

GAATTCATACCAGATCACCGGATTCCCGACTCCAAATGTGTCCCCCTCACAC
TCCCCCGATTACCGTCTTCTGCTCTTAGACCACTCTACCCTATTCCCCACACT
CACCGGAGCCAAAGCCGCGGCCCTTCCGTTCCGATTACCGAAAAGACCCCA
CCCGTAGGTGGCAAGCTAGCTTAAGTAACGCCACTTCGATTAACGAGGAAA
AATACATAACTGACCTATTATCGAGTTCAGATCAAGGTCAGGAACAAAGAA
ACACCGATTACCGTAACCGTAAGATARTGGTATCGATACGTAGACAGTTTA

Estimating "fill rate" in parking lots



Why Data Science and AI?

- IN ADDTION to increasing data collection:
  - Questions about epidemic spread naturally have a large spatial and temporal scales
    - And multiple such scales!
  - Networks are everywhere
  - Models have become complex
  - Small and big data, noisy and incomplete
  - New data science and AI techniques which can handle end-to-end learning, stochastic optimization and are responsive to dynamic changes
  - New computing tools can work in conjunction with epidemiologists, biologists, engineers, decision makers

Georgia Tech

Incorporating Behaviors: Performative Predictions

[SIGKDD 2025]

Leo Zhao

Figure 1: Example performative feedback loop for COVID-19 mortality and mobility.

Papers and code: cc.gatech.edu/~badityap
github.com/AdityaLab/
badityap@cc.gatech.edu

Thanks!

- Students
  - Arjun Choudhry
  - Harsha Kamarthi
  - Shangqing Xu
  - Pulak Agarwal
  - Javen Ho
  - Haoxin Liu
  - Zhiyuan (Leo) Zhao
  - Srikar Bulusu
  - Rishi Raman
  - June Cho
  - Megha Sharma
  - Dr. Jiajia Xie (graduated → Tesla)
  - Dr. Jiaming Cui (graduated → Asst. Prof. CS, Virginia Tech)
  - Dr. Alexander Rodriguez (→ Asst. Prof. CSE, U. of Michigan – Ann Arbor)
  - Dr. Anika Tabassum (→ Oak Ridge National Lab)
  - Dr. Bijaya Adhikari (→ Asst. Prof. UIowa)
  - Dr. Stacey Clifton (→ Asst. Prof. Radford)
  - Dr. Sorour Amiri (Google), Dr. Xinfeng Xu (Johns Hopkins), Dr. Liangzhe Chen (Pinterest), Dr. Yao Zhang (Yahoo)

- Collaborators (subset): Vyas Sekar (CMU), Kai Wang (GaTech), Ramesh Raskar (MIT), Rajat Sen (Google), Shinobu Kitayama (Michigan), Gerg Wellenius (BU/Google), Anil Vullikanti (UVA), Naren Ramakrishnan (VT), Gregory Abowd (GaTech), Aditya Parameswaran (Berkeley), Chao Zhang (GaTech), Supriya Chinthavali (ORNL), Hanghang Tong (UIUC), Jilles Vreeken (MPI), Sriram Pemmaraju (Iowa)

- Supported by:

  NSF (Expeditions, IIS, IIS, IIS, RAPID, CAREER, CCF), NIH R01, GTRI, ORNL, Facebook, CDC MIND, Dow, Walmart. Dolby..