DeepMind

Shakir Mohamed

@shakir_za shakir@deepmind.com

A Case Study in Data Science and AI

Predicting Organ Failure in Hospitals

#DSRD19

Machine Learning in Healthcare

Electronic
Records

Medical
Imaging

Business
Operations

Many areas for Machine Learning and Digital Platforms to play a role.

Triple Aims

'The Triple Aim' *Health Affairs*
Don Berwick

1 **Better clinical outcomes**

2 **Enhance patient and clinician experience**

3 **Reduce costs**

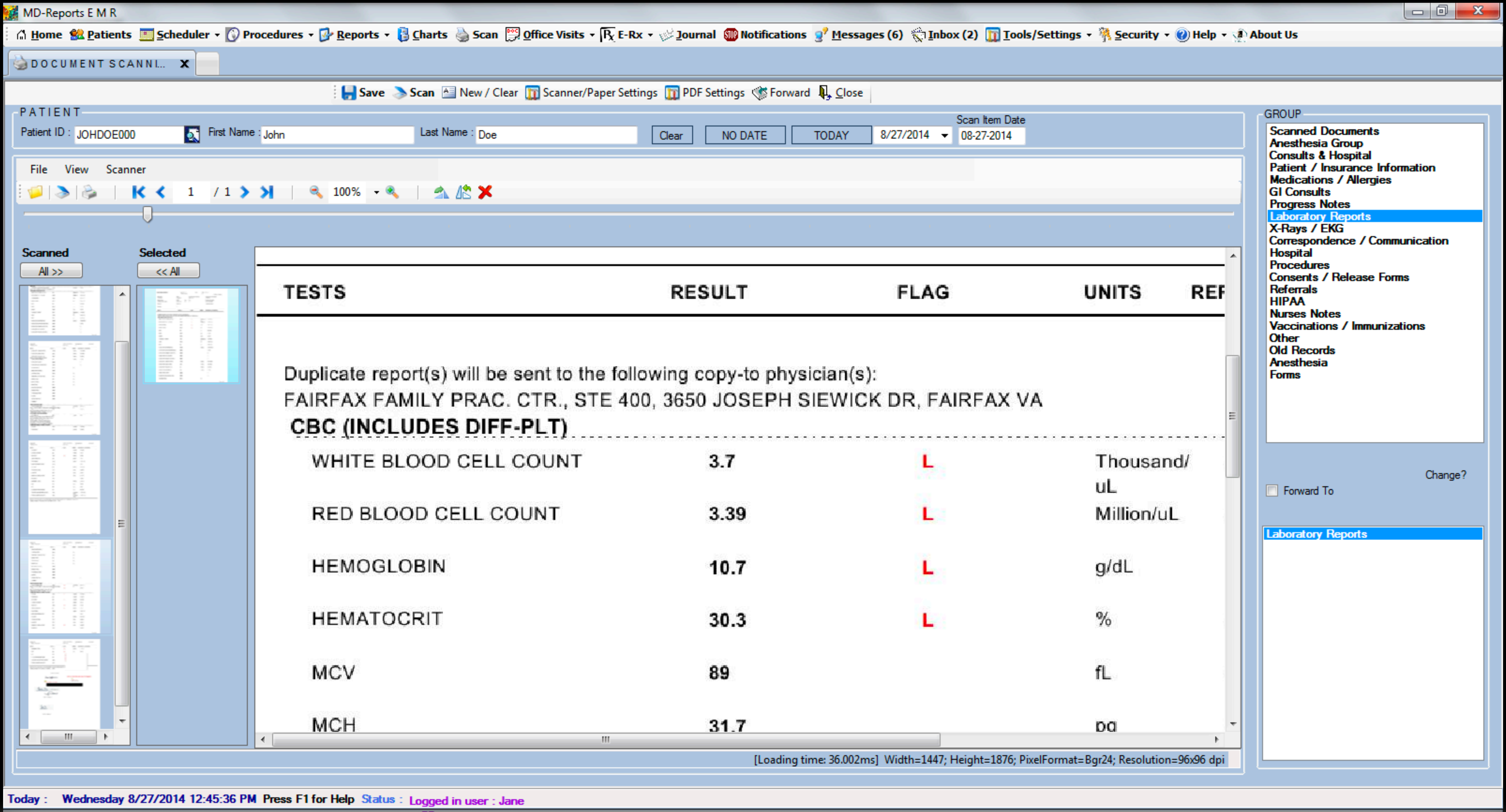
Systemic challenges

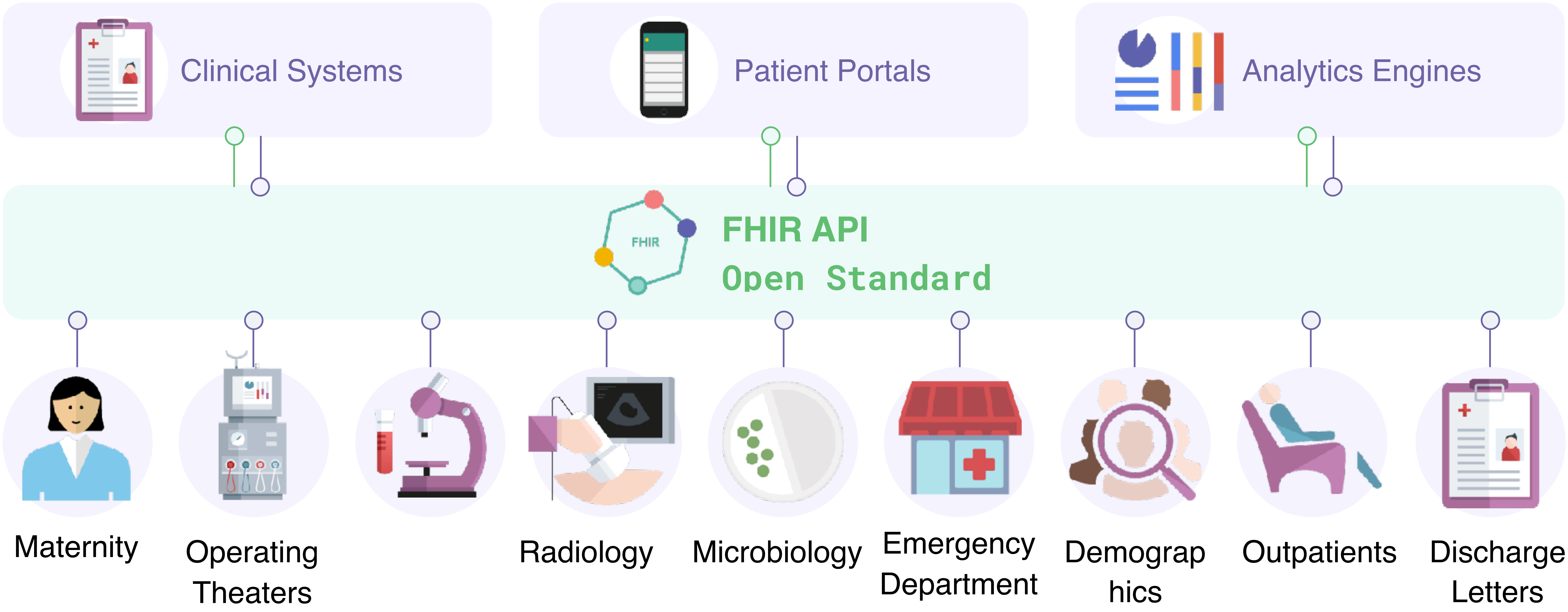
- >50% of healthcare not evidence based
- Staff burnout rates on the rise
- Care continues to be episodic vs integrated
- Intractable increases in healthcare costs
- Failure to deliver shared decision making for patients
- Unwarranted variation exists across healthcare delivery
- > 10% of patients experience harm in hospitals
- Focus and on illness at the expense of prevention

Detecting Deterioration

- Millions of people die every year from diseases that could be prevented with earlier detection.
- Worked with a hospital partner to look at AI for predicting patient deterioration.
- **Acute kidney injury (AKI),** a condition where a patient's kidney suddenly stops working properly. Affecting up to 1in 5 hospitalised patients in UK and US.

Data from these processes are captured within an electronic health record.


MD-Reports E M R
Home
Patients
Scheduler
Procedures
Reports
Charts
Scan
Office Visits
E-Rx
Journal
Notifications
Messages (6)
Inbox (2)
Tools/Settings
Security
Help
About Us
DOCUMENT SCANNI...
Save
Scan
New / Clear
Scanner/Paper Settings
PDF Settings
Forward
Close
PATIENT
Patient ID : JOHDOE000
First Name : John
Last Name : Doe
Clear
NO DATE
TODAY
8/27/2014
Scan Item Date
08-27-2014
File
View
Scanner
1 / 1
100%
Scanned
All >>
Selected
<< All
TESTS
RESULT
FLAG
UNITS
REF
Duplicate report(s) will be sent to the following copy-to physician(s):
FAIRFAX FAMILY PRAC. CTR., STE 400, 3650 JOSEPH SIEWICK DR, FAIRFAX VA
CBC (INCLUDES DIFF-PLT)
WHITE BLOOD CELL COUNT 3.7 L Thousand/uL
RED BLOOD CELL COUNT 3.39 L Million/uL
HEMOGLOBIN 10.7 L g/dL
HEMATOCRIT 30.3 L %
MCV 89 fL
MCH 31.7 pg
[Loading time: 36.002ms] Width=1447; Height=1876; PixelFormat=Bgr24; Resolution=96x96 dpi
GROUP
Scanned Documents
Anesthesia Group
Consults & Hospital
Patient / Insurance Information
Medications / Allergies
GI Consults
Progress Notes
Laboratory Reports
X-Rays / EKG
Correspondence / Communication
Hospital
Procedures
Consents / Release Forms
Referrals
HIPAA
Nurses Notes
Vaccinations / Immunizations
Other
Old Records
Anesthesia
Forms
Change?
Forward To
Laboratory Reports
Today : Wednesday 8/27/2014 12:45:36 PM Press F1 for Help Status : Logged in user : Jane



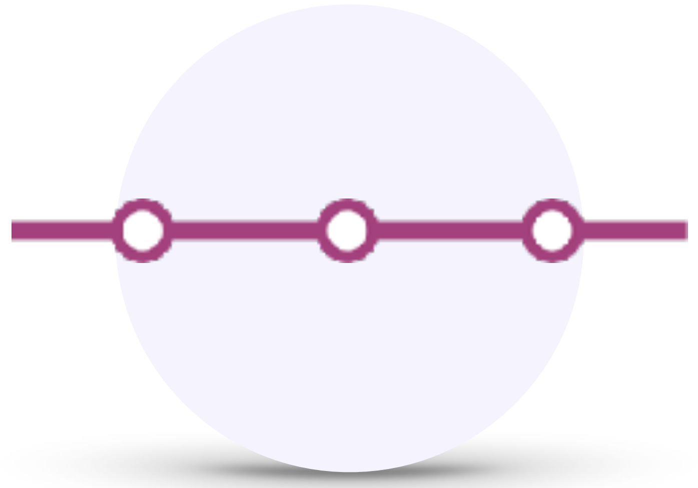
Characteristics

- Unstructured
- Noisy
- Recorded differently

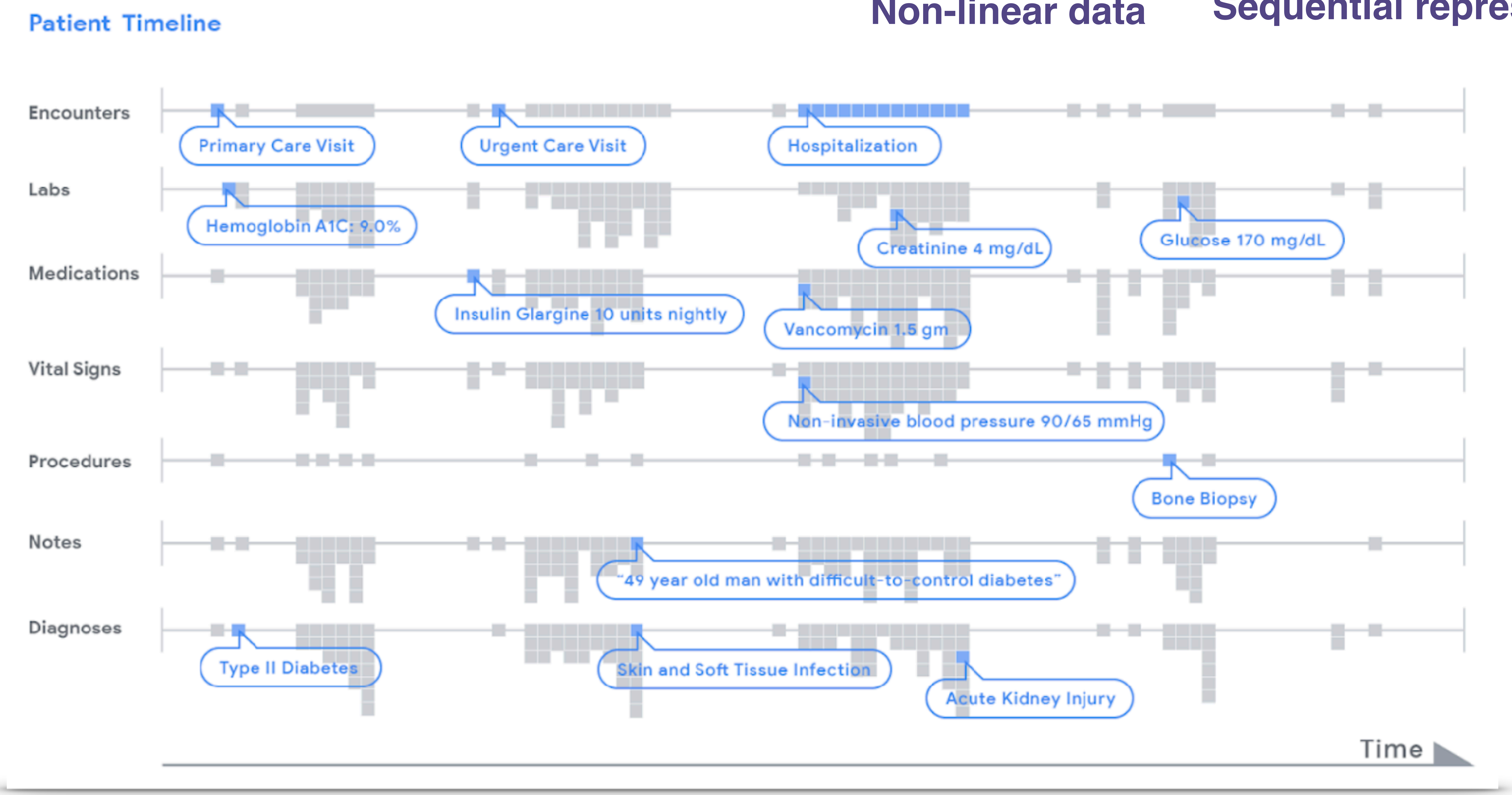
DS/AI Interactions:

- SWE and architects
- Security, privacy, law
- Clinical requirements

Non-linear data



Sequential representation



Data and Summarisation

Data from a large hospital partner

		Training	Validation	Calibration	Test
Patients					
Unique patients		562,507	35,277	35,317	70,681
Average age*		62.4	62.5	62.4	62.3
Ethnicity	Black	106,299 (18.9%)	6,544 (18.6%)	6,675 (18.6%)	13,183 (18.7%)
	Other	456,208 (81.1%)	28,733 (81.4%)	28,642 (81.4%)	57,498 (81.3%)
Gender	Female	35,855 (6.4%)	2,300 (6.5%)	2,252 (6.4%)	4,519 (6.4%)
	Male	526,652 (93.6%)	32,977 (93.5%)	33,065 (93.6%)	66,162 (93.6%)
Diabetes		56,958 (10.1%)	3,599 (10.2%)	3,702 (10.5%)	7,093 (10.0%)
Renal replacement therapy		22,284 (4.0%)	1,367 (3.9%)	1,384 (3.9%)	2,784 (3.9%)

Data Summarisation

Characteristics

- Sequential representation
- Sparse
- Missing data
- Included and excluded
- Handling time, alignment

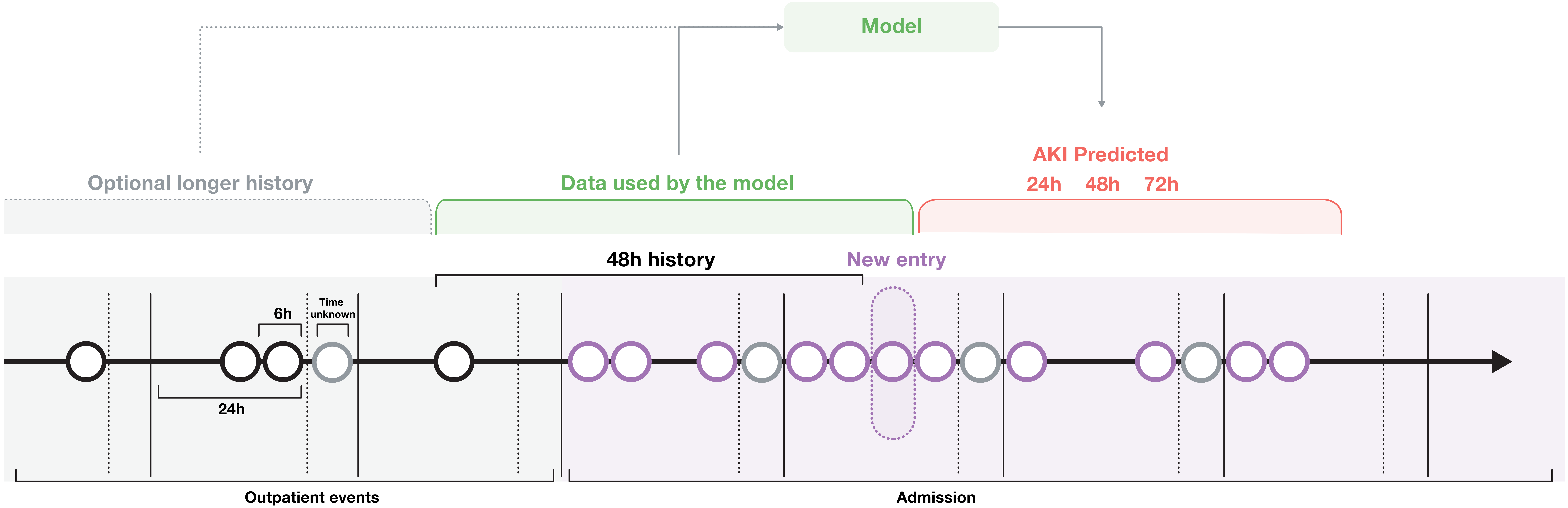
DS/AI Interactions:

- Important research questions; arise from practical considerations.
- Where do labels come from?
- What predictions and metrics are important to clinicians.

Summary of dataset

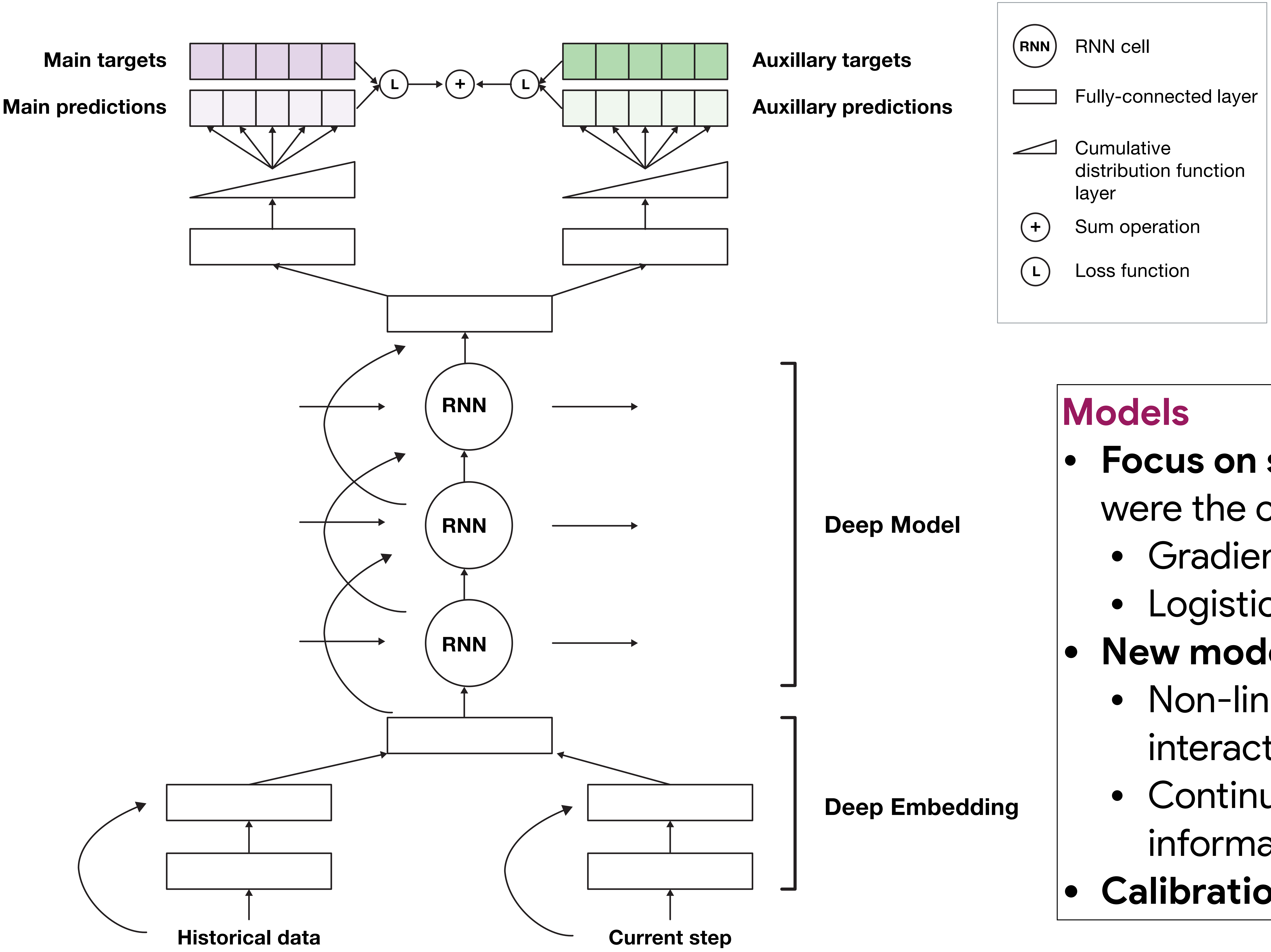
Admissions within a five year period					
Data center sites		130***	130***	130***	130***
Unique admissions		2,004,217	124,255	125,928	252,492
- per patient	Average	3.6	3.5	3.6	3.6
	Median	2	2	2	2
Duration (days)	Average	9.6	9.6	9.6	9.6
	Median	3.2	3.2	3.2	3.2
ICU admissions		214,644 (10.7%)	13,161 (10.6%)	13,411 (10.6%)	26,739 (10.6%)
Medical admissions		971,527 (48.5%)	60,762 (48.9%)	61,281 (48.7%)	121,675 (48.2%)
Surgical admissions		354,008 (17.7%)	21,857 (17.6%)	22,093 (17.5%)	44,766 (17.7%)
No creatinine measured		408,927 (20.4%)	25,162 (20.3%)	25,503 (20.3%)	51,484 (20.4%)
Chronic Kidney Disease	Any	746,692 (37.3%)	46,677 (37.5%)	46,622 (37.0%)	94,105 (37.3%)
	Stage 1**	8,409 (0.4%)	515 (0.4%)	576 (0.5%)	1,103 (0.4%)
	Stage 2	429,990 (21.5%)	27,162 (21.9%)	26,927 (21.4%)	54,476 (21.6%)
	Stage 3A	156,720 (7.8%)	9,837 (7.9%)	9,803 (7.8%)	19,548 (7.7%)
	Stage 3B	77,801 (3.9%)	4,675 (3.8%)	4,823 (3.7%)	9,760 (3.9%)
	Stage 4	31,646 (1.6%)	1,999 (1.6%)	2,003 (1.6%)	4,098 (1.6%)
	Stage 5	50,535 (2.5%)	3,004 (2.5%)	3,066 (2.5%)	6,223 (2.5%)
AKI present	Any AKI	267,396 (13.3%)	16,671 (13.4%)	16,760 (13.3%)	33,759 (13.4%)
	Stage 1	207,441 (10.4%)	12,794 (10.3%)	12,951 (10.3%)	26,215 (10.4%)
	Stage 2	43,446 (2.2%)	2,780 (2.2%)	2,783 (2.2%)	5,575 (2.2%)
	Stage 3	66,734 (3.3%)	4,267 (3.4%)	4,162 (3.3%)	8,453 (3.3%)

Future Prediction of AKI



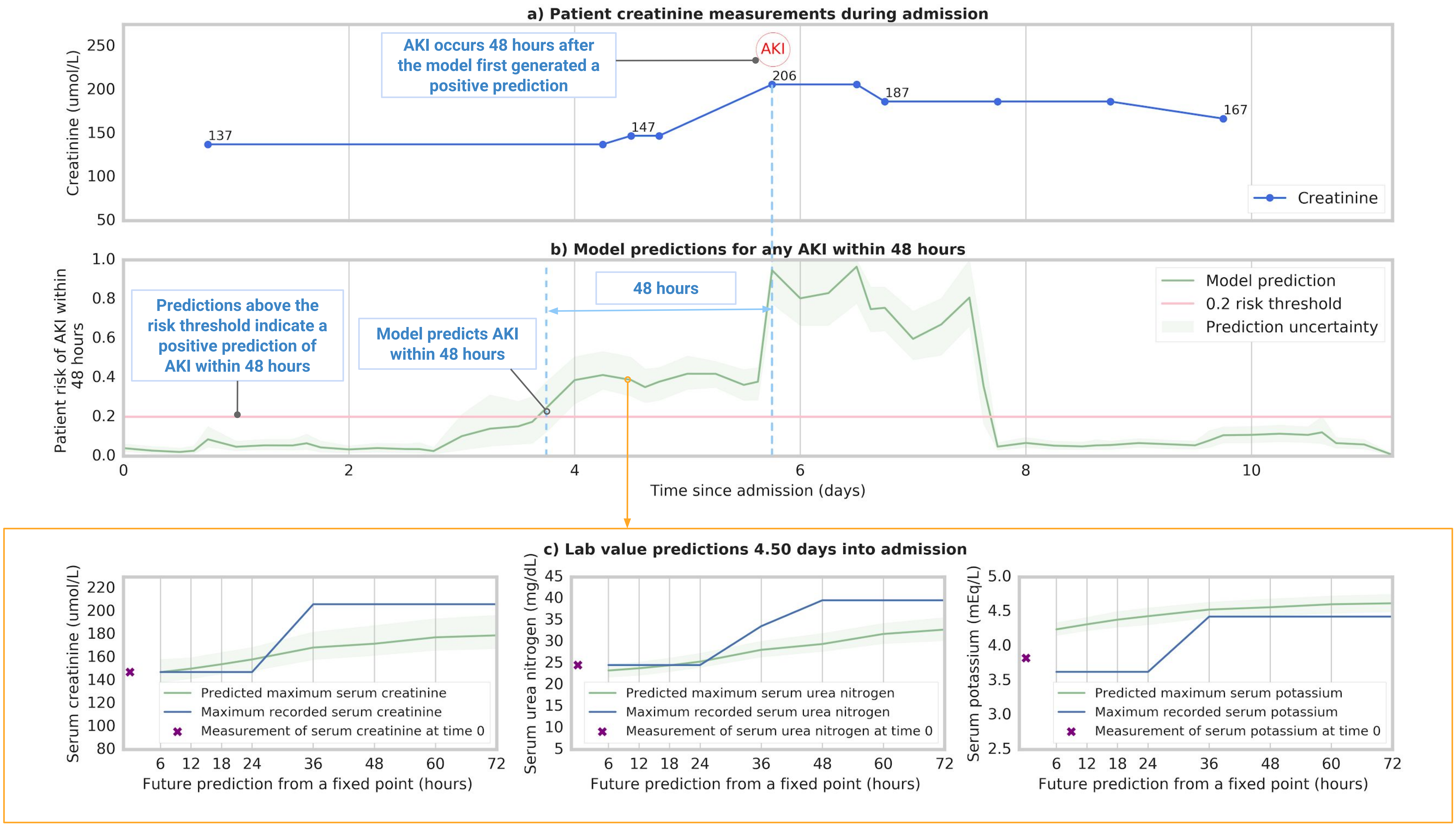
Useful predictions are those that are accurate and continuously updated, given with sufficient time to act, provide context for decision

Model on 700k features. Make predictions up to 48hrs ahead.



Models

- **Focus on strong baselines** that were the current state of the art.
  - Gradient Boosted Trees
  - Logistic regression
- **New models using Deep Learning**
  - Non-linear models and interactions
  - Continuous integration of information as they are received
- **Calibration, uncertainty**



a) Patient creatinine measurements during admission
Creatinine (umol/L)
250
200
150
100
50
AKI occurs 48 hours after the model first generated a positive prediction
AKI
137
147
206
187
167
Creatinine
b) Model predictions for any AKI within 48 hours
Patient risk of AKI within 48 hours
1.0
0.8
0.6
0.4
0.2
0.0
Predictions above the risk threshold indicate a positive prediction of AKI within 48 hours
Model predicts AKI within 48 hours
48 hours
Model prediction
0.2 risk threshold
Prediction uncertainty
0
2
4
6
8
10
Time since admission (days)
c) Lab value predictions 4.50 days into admission
Serum creatinine (umol/L)
220
200
180
160
140
120
100
80
Predicted maximum serum creatinine
Maximum recorded serum creatinine
Measurement of serum creatinine at time 0
6
12
18
24
36
48
60
72
Future prediction from a fixed point (hours)
Serum urea nitrogen (mg/dL)
45
40
35
30
25
20
15
10
5
Predicted maximum serum urea nitrogen
Maximum recorded serum urea nitrogen
Measurement of serum urea nitrogen at time 0
Serum potassium (mEq/L)
5.0
4.5
4.0
3.5
3.0
2.5
Predicted maximum serum potassium
Maximum recorded serum potassium
Measurement of serum potassium at time 0

A Clinically-applicable Approach to the Continuous Prediction of Future Acute Kidney

Summary:

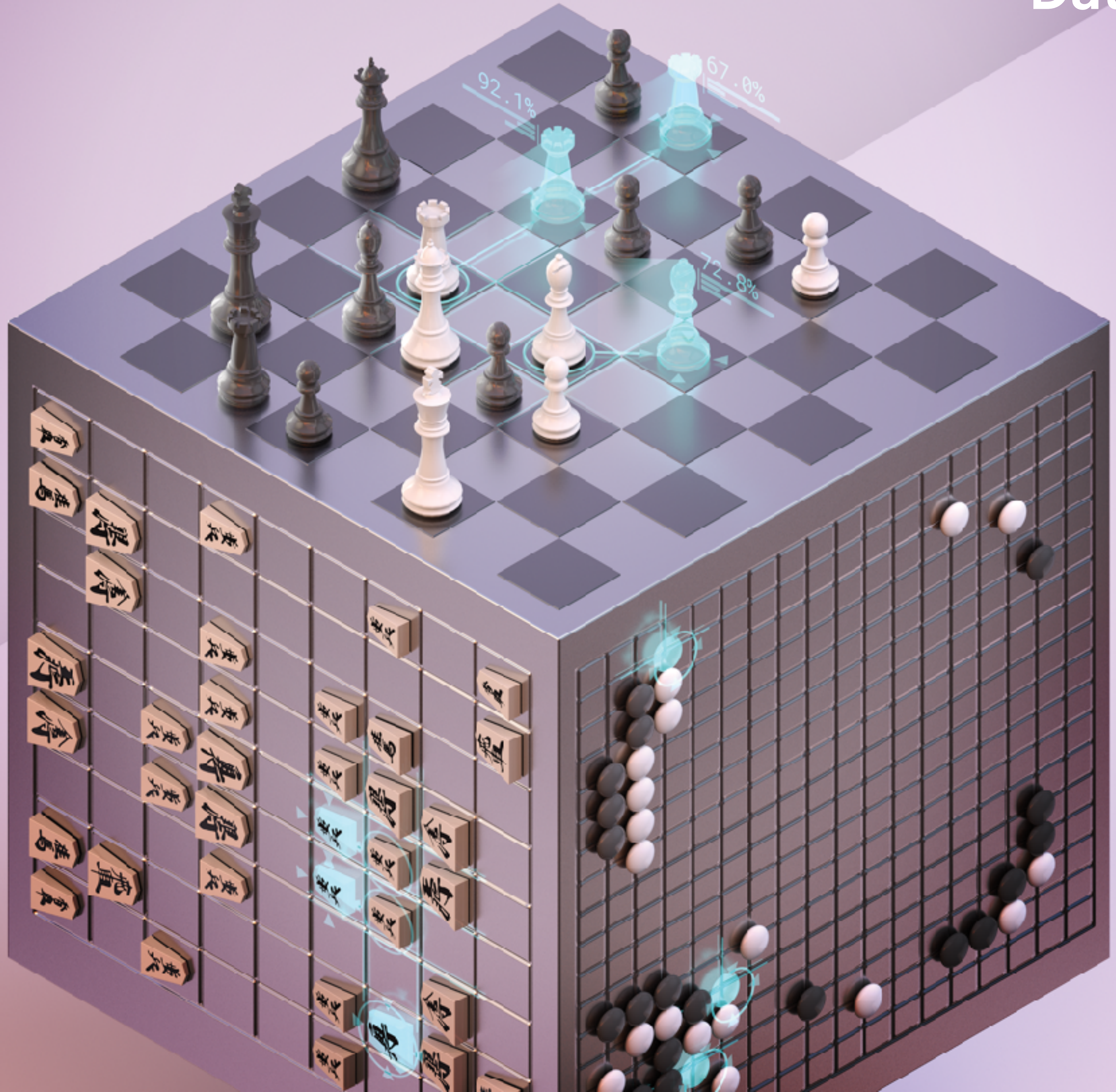
- Make predictions of AKI up to 48hr ahead.
- Provide a window for meaningful action.
- For the most severe cases, can detect up to 90% of cases.

Further considerations and limitations:

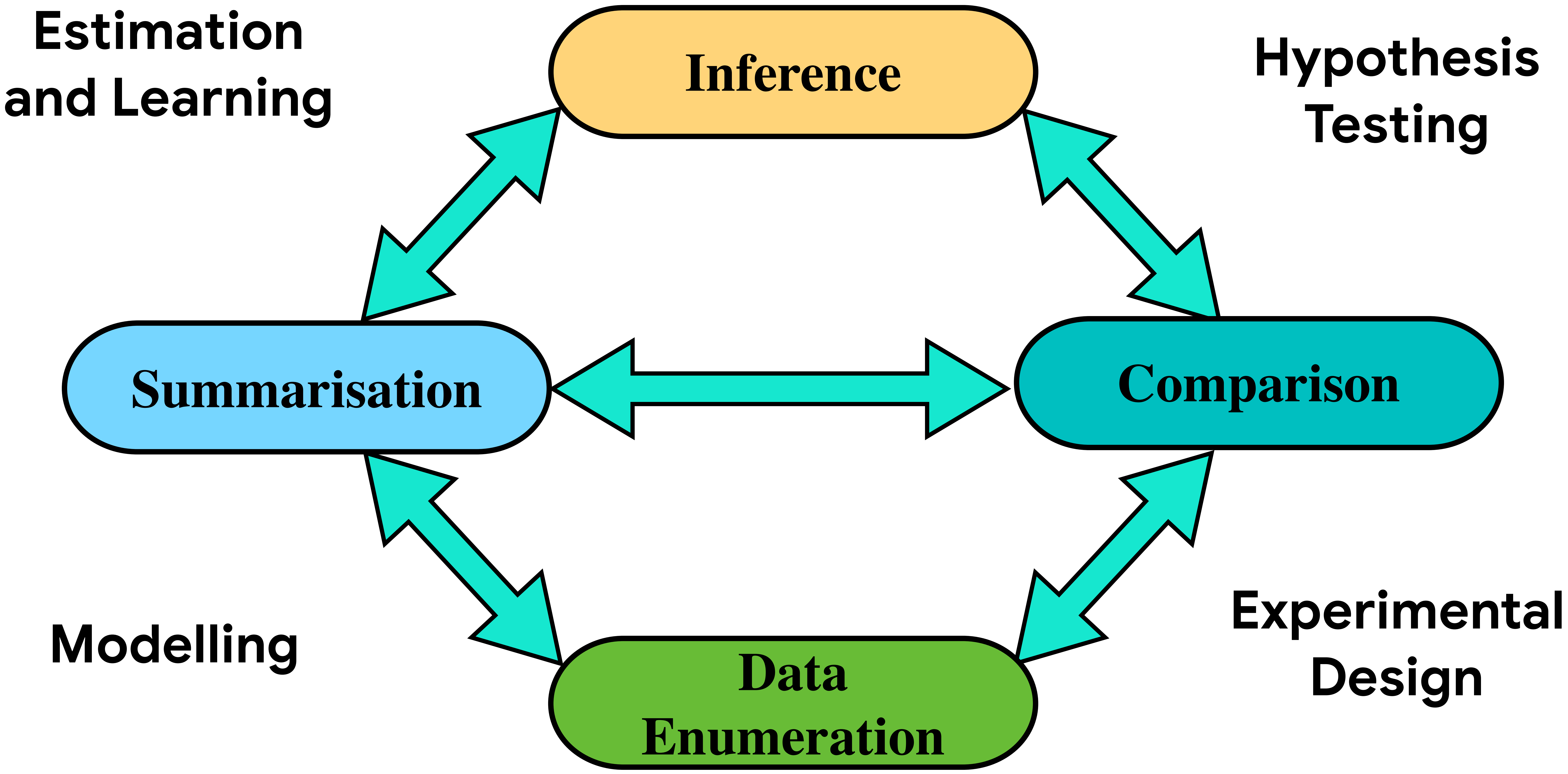
- Early or late predictions and alerting fatigue
- Generalisation needed to wider steps of hospitals, patient populations.
- Only a retrospective study.
- Need prospective studies to evaluate real clinical-use.

Tomasev et al. (2019)

Many Other Sources of Questions, Partners and Data



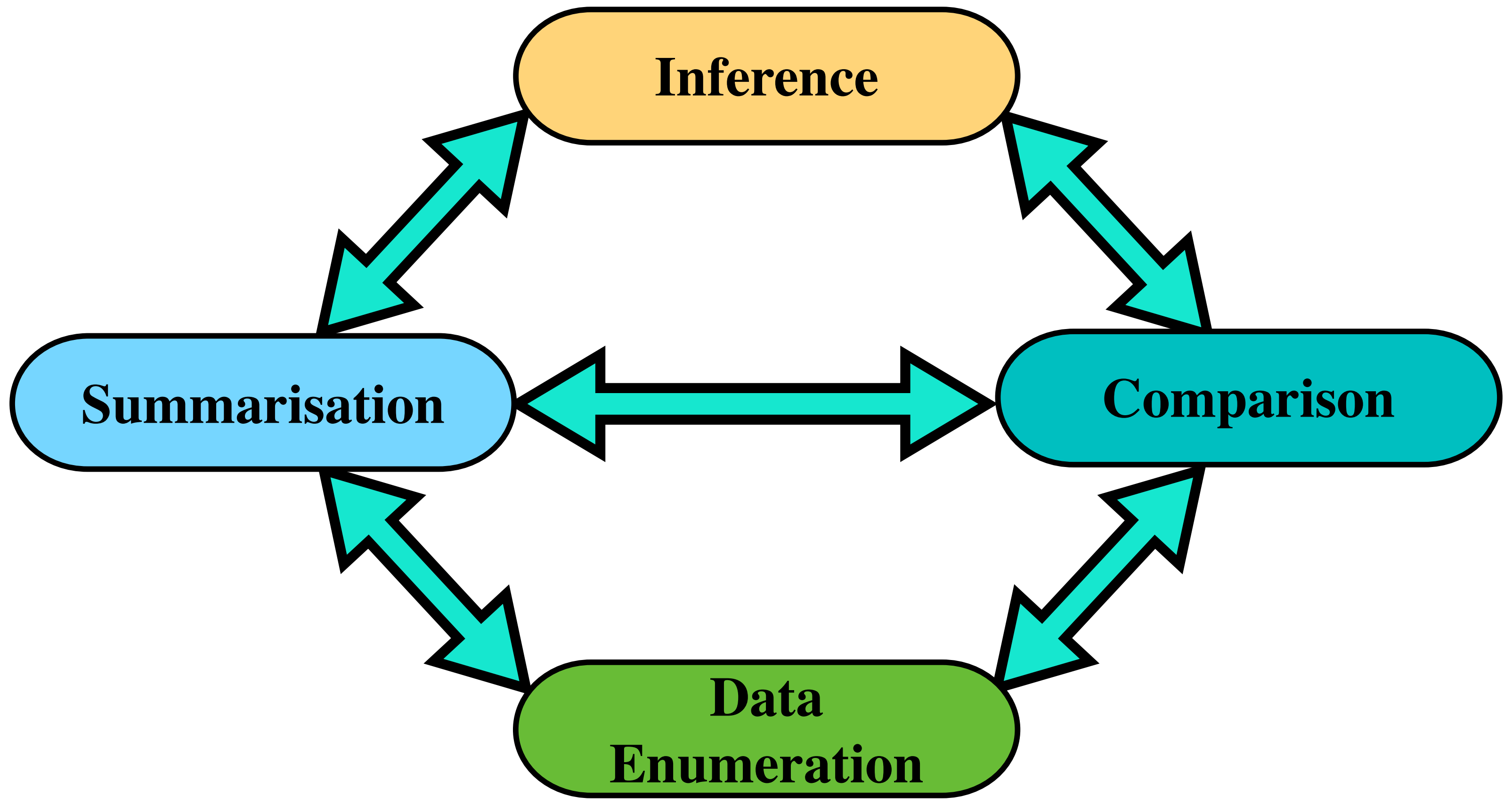
Statistical Operations



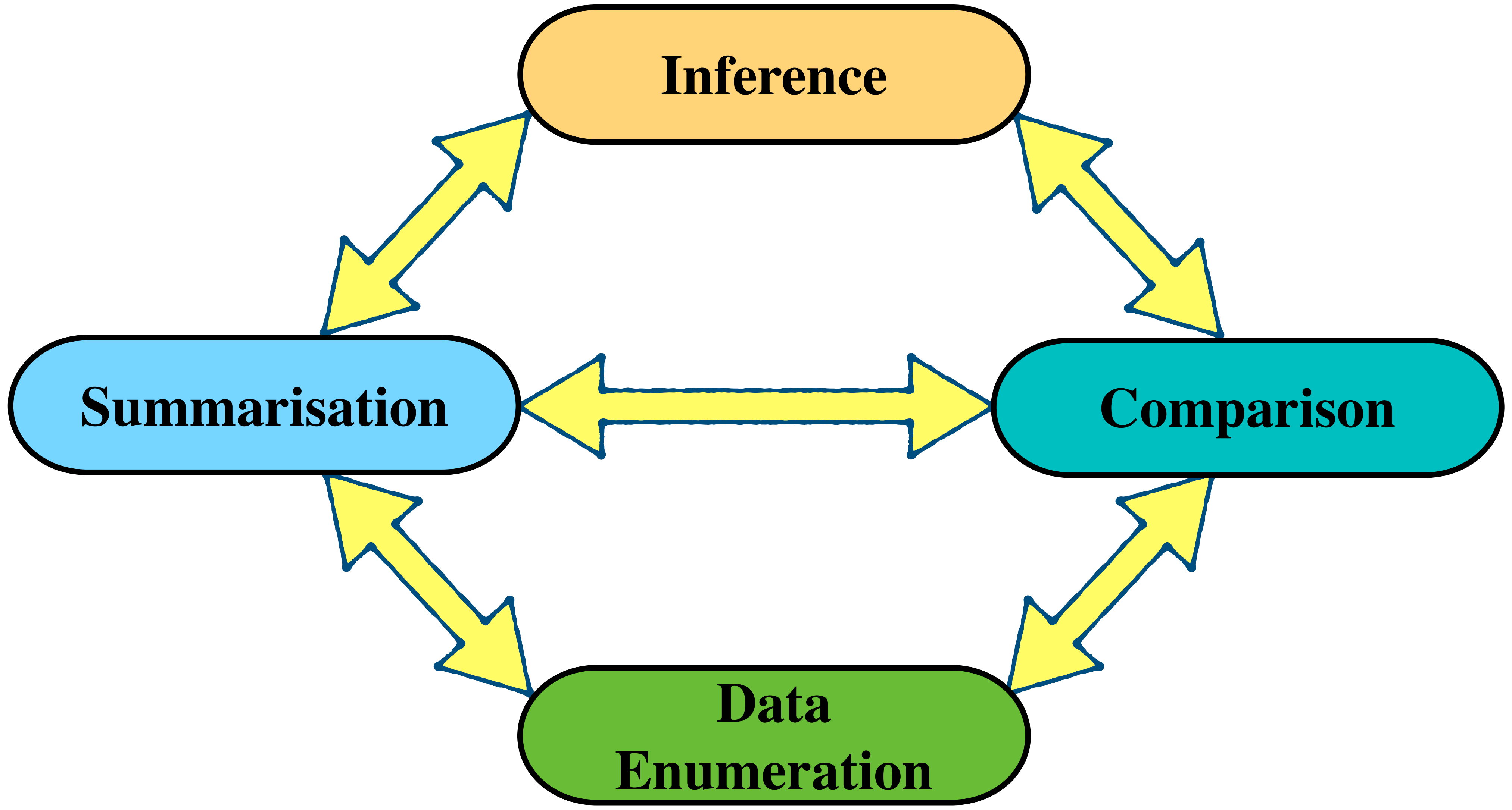
Statistical Operations

Inference
What we can *know* about our data

Decision-making
What we can *do* with our data.



Centrality of Inference



Artificial Intelligence will be the refined instantiation of these statistical operations.

The core questions of AI will be those of probabilistic inference

Principles to Products

Applications	Assistive Technology	Advancing Science	Climate and Energy	Healthcare	Fairness and Safety	Autonomous systems
Reasoning	Planning	Explanation	Rapid Learning	World Simulation	Objects and Relations	
Information	Uncertainty	Information Gain	Causality	Prediction		
Principles	Probability Theory	Bayesian Analysis	Hypothesis Testing	Estimation Theory	Asymptotics	

Neutrality and Universality

Neutrality Traps

- **The Portability Trap:** solutions designed for one social context may be inaccurate / do harm when applied to a different context.
- **The Formalism Trap:** Failure to account for the full meaning of social concepts such as fairness, and think they can be resolved through mathematical formalisms.
- **The Ripple Effect Trap**: Inserting technology into an existing social system changes the behaviours and embedded values of the pre-existing system .
- **The Solutionism Trap**: Failure to recognise the possibility that the best solution to a problem may not involve technology.

Universality

‘A mono-cultural view of ethics conceives itself as the only valid one. In order to **avoid this kind of ethical chauvinism and colonialism** it is necessary that transcultural ethics arise from an intercultural dialogue instead of thinking of itself as universal without noticing its own cultural bias.’ Capurro, 2004

SUSTAINABLE DEVELOPMENT GOALS

DeepMind

Shakir Mohamed

@shakir_za

A Case Study in Data Science and AI

Predicting Organ Failure in Hospitals

#DSRD19