AI in Medicine: Using What We Know

Zina Ibrahim
Senior Lecturer in AI for Medicine
Department of Biostatistics and Health Informatics
King’s College London

Demis Hassabis & John Jumper awarded Nobel Prize in Chemistry

9 OCTOBER 2024

Share

Article | Open access | Published: 15 July 2021

Highly accurate protein structure prediction with AlphaFold

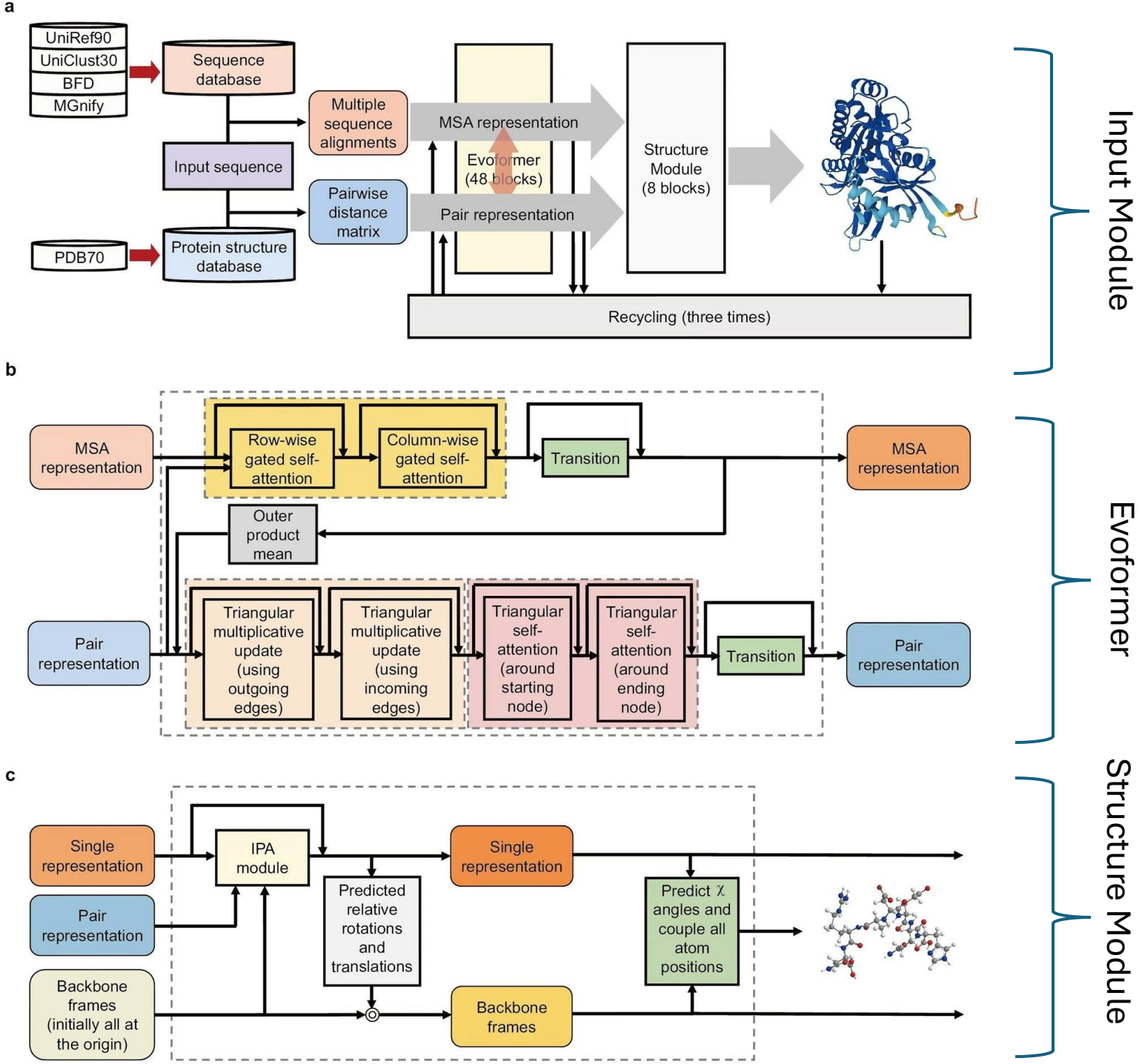
John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, ... Demis Hassabis + Show authors

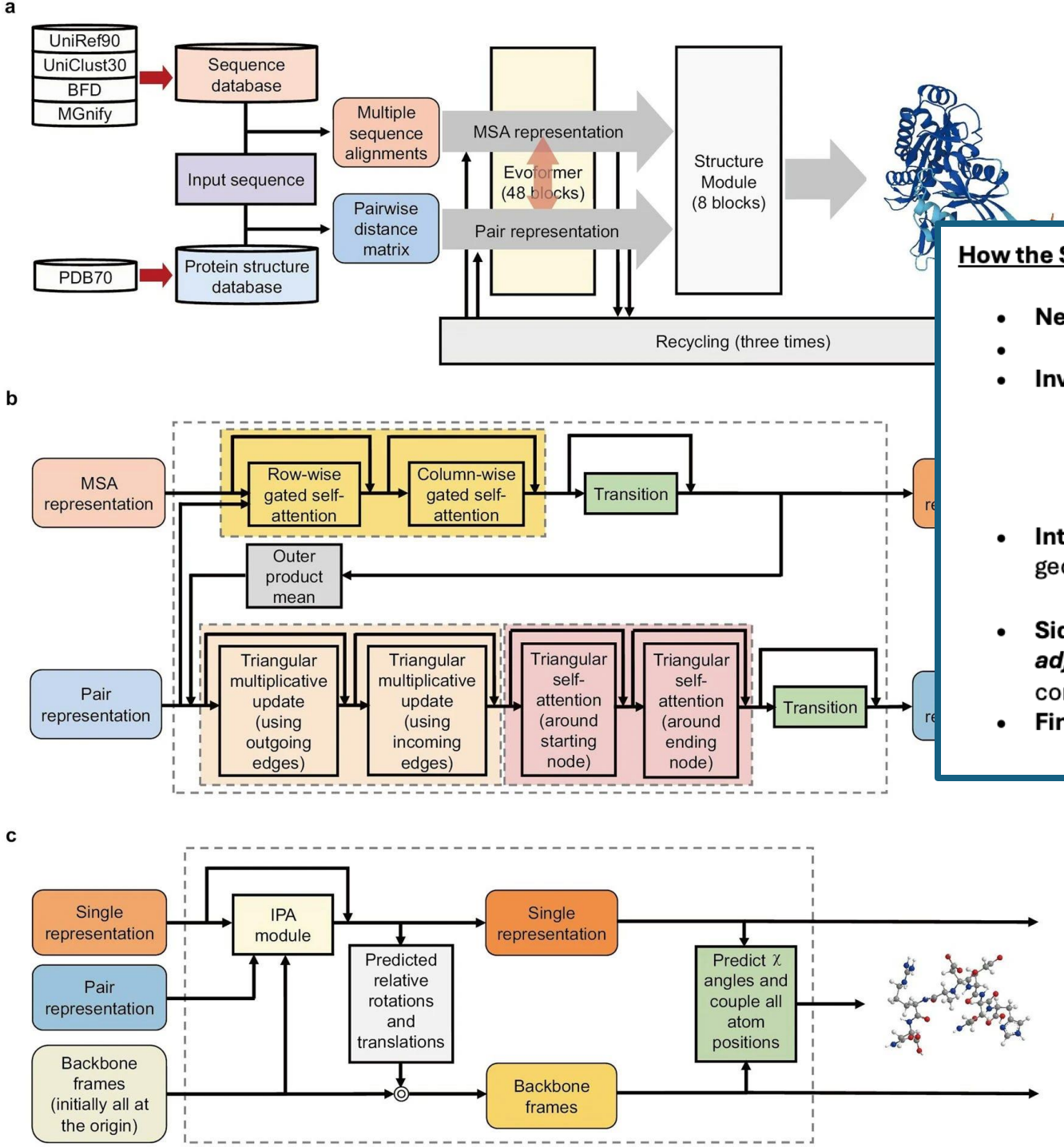
Nature **596**, 583–589 (2021) | Cite this article

Abstract

Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function. Through an enormous experimental effort[1,2,3,4], the structures of around 100,000 unique proteins have been determined[5], but this represents a small fraction of the billions of known protein sequences[6,7]. Structural coverage is bottlenecked by the months to years of painstaking effort required to determine a single protein structure. Accurate computational approaches are needed to address this gap and to enable large-scale structural bioinformatics. Predicting the three-dimensional structure that a protein will adopt based solely on its amino acid sequence—the structure prediction component of the 'protein folding problem'[8]—has been an important open research problem for more than 50 years[9]. Despite recent progress[10,11,12,13,14], existing methods fall far short of atomic accuracy, especially when no homologous structure is available. Here we provide the first computational method that can regularly predict protein structures with atomic accuracy even in cases in which no similar structure is known. We validated an entirely redesigned version of our neural network-based model, AlphaFold, in the challenging 14th Critical Assessment of protein Structure Prediction (CASP14)[15], demonstrating accuracy competitive with experimental structures in a majority of cases and greatly outperforming other methods. Underpinning the latest version of AlphaFold is a novel machine learning approach that incorporates physical and biological knowledge about protein structure, leveraging multi-sequence alignments, into the design of the deep learning algorithm.

The Alphafold Architecture





Input M

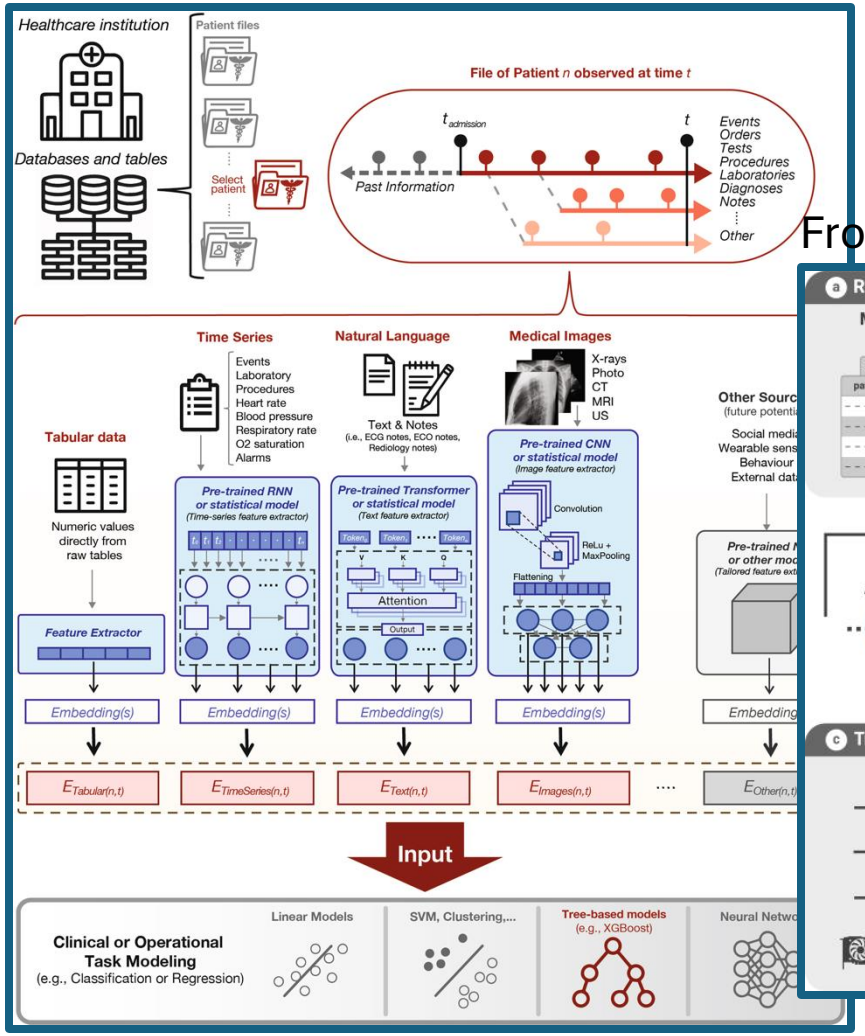
How the Structure Module Works:

- **Neural Input**: learned sequence embeddings from Evoformer
-
- **Invariant Point Attention (IPA)**: is a geometry-aware attention operation:
  - Models ***spatial relationships between residues***
  - Ensures predictions are ***invariant to rotations and translations***.
  - Guides how residues should interact with 3D space.
- **Interaction with Neural Layers**: IPA refines neural predictions iteratively, using geometric constraints to align residues in realistic 3D positions.
- **Side Chain and Backbone Refinement**: ***IPA alternates between spatial adjustments and neural predictions*** to refine both backbone and side chain conformations.
- **Final Output**: Produces a 3D structure

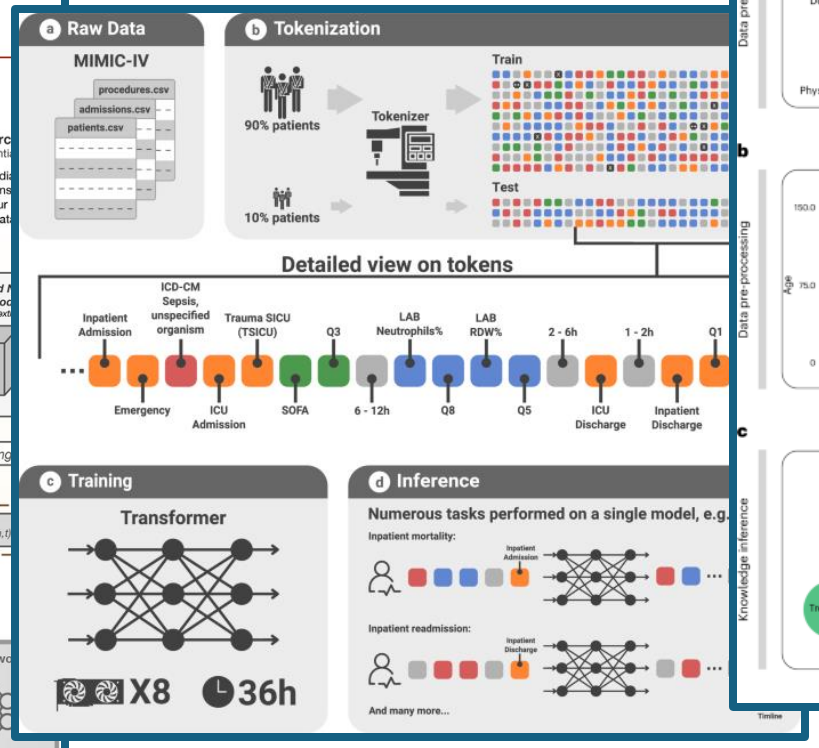
Structure Module

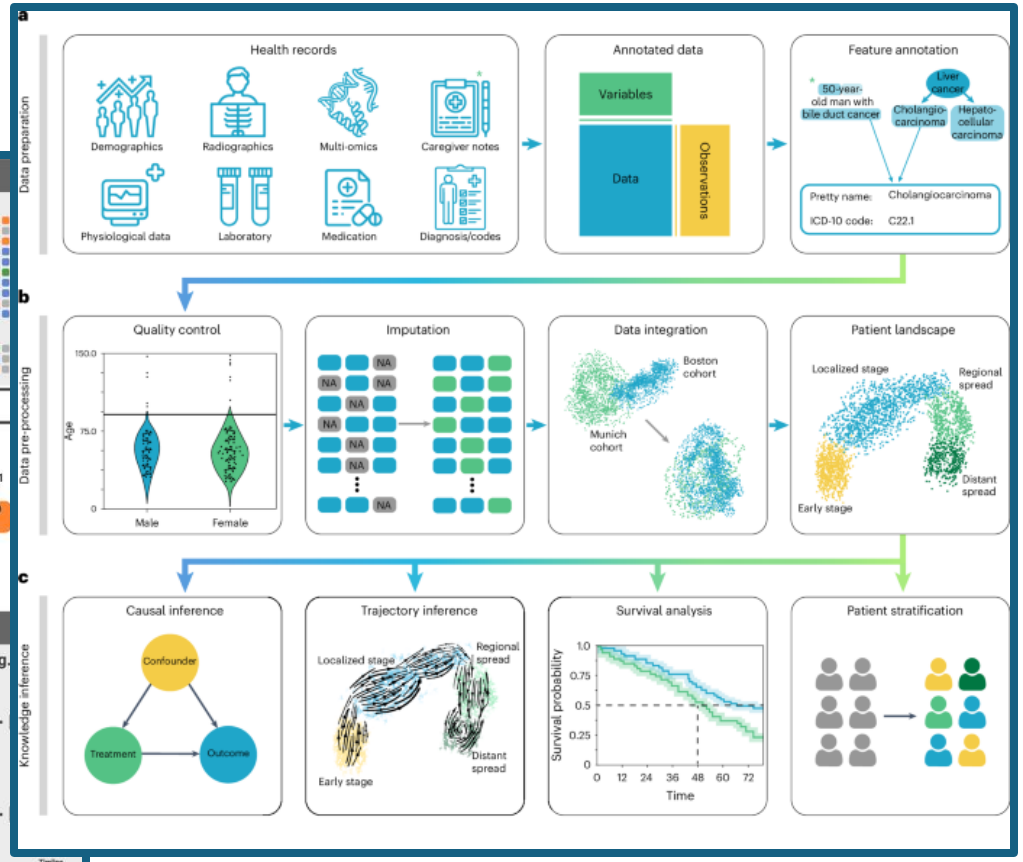
In Medicine: Patient Trajectory Prediction

From NPJ



From Nature

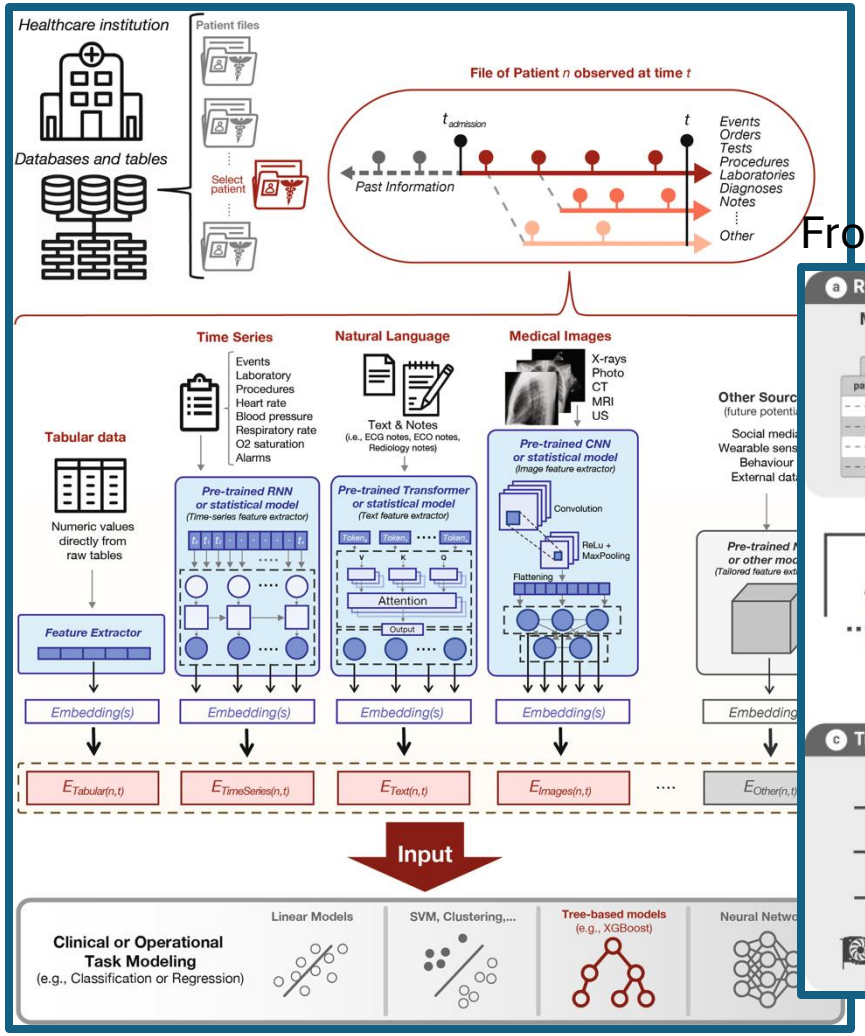




From Nature

In Medicine: Patient Trajectory Prediction

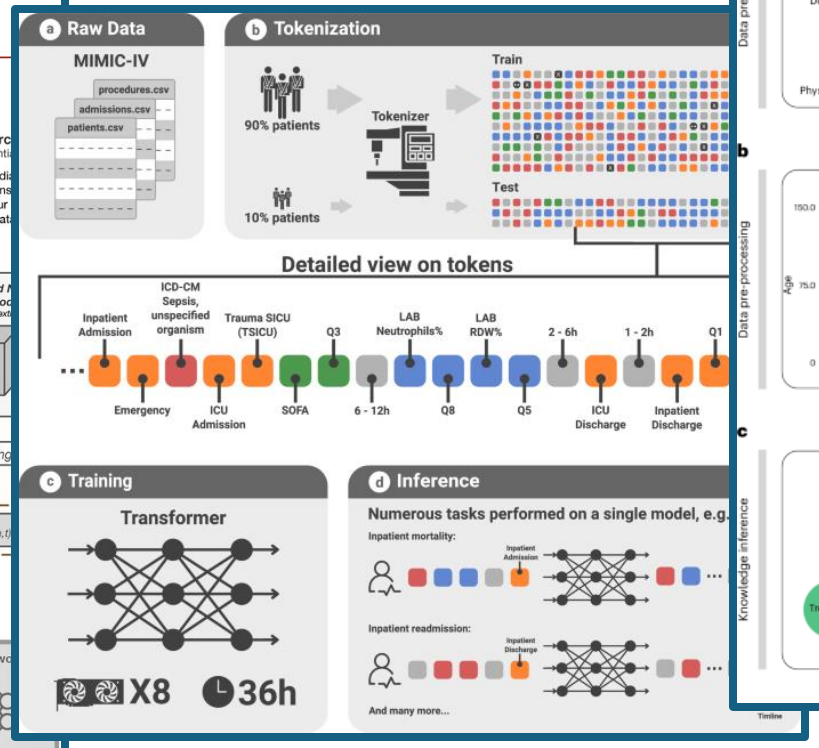
From NPJ

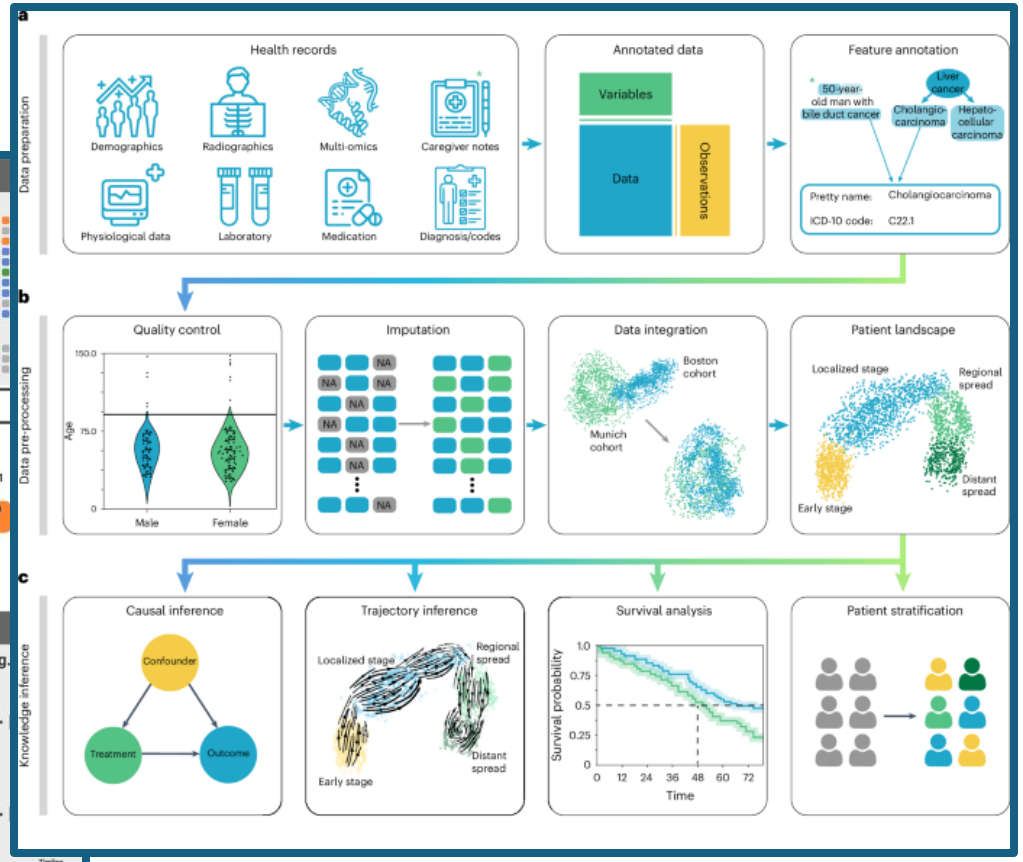


Medicine is as abundant in knowledge as it is in data.

So why aren't we using what we already know?

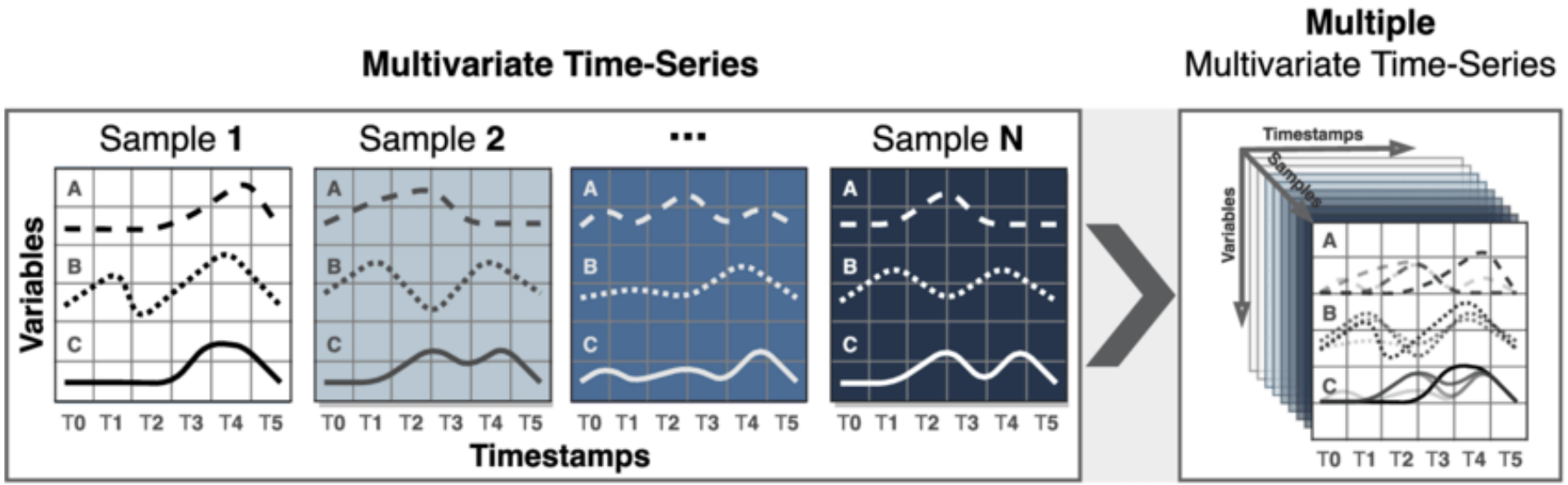
From Nature



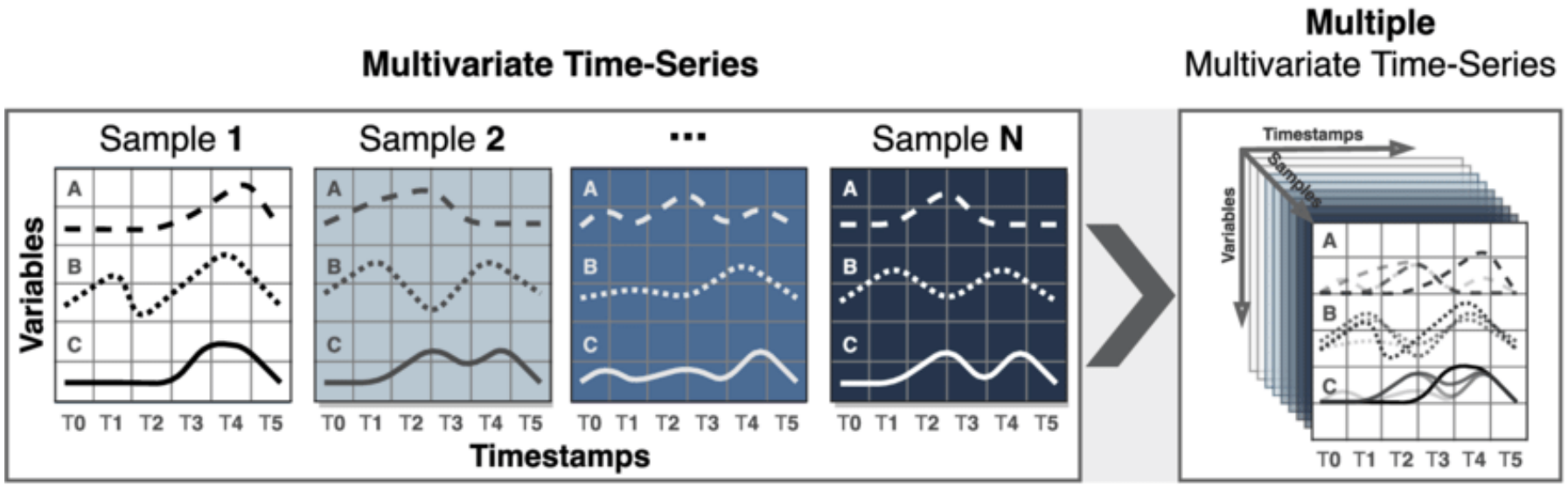


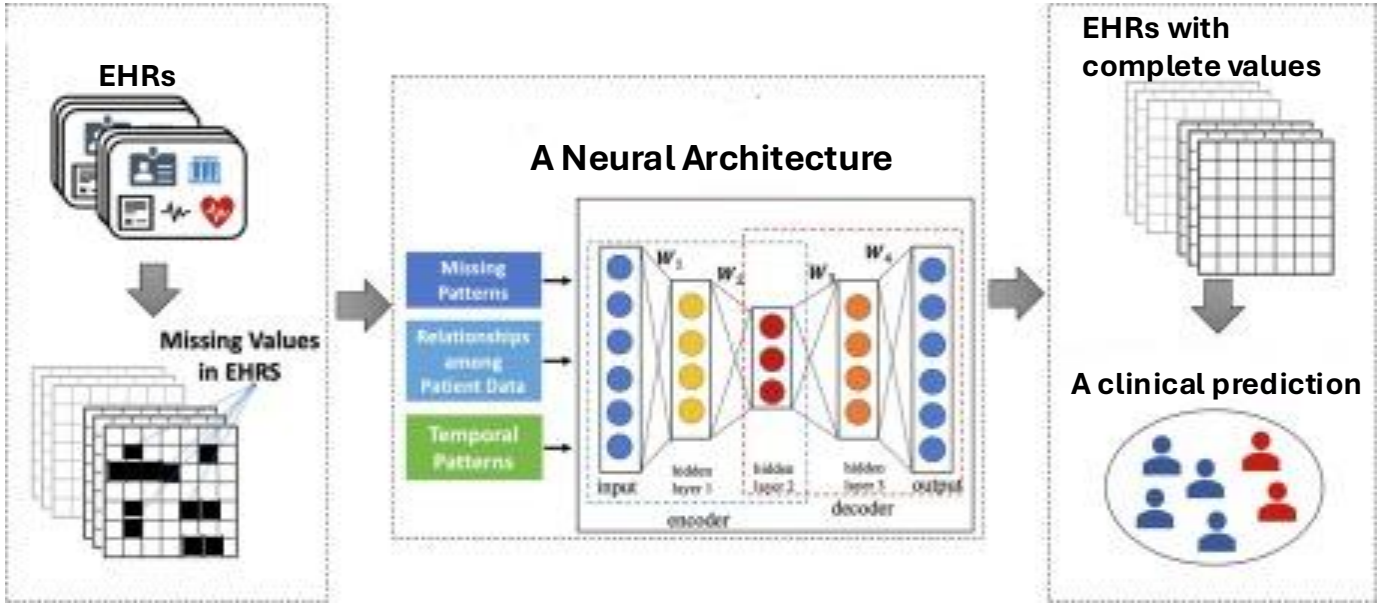
From Nature

The Structured Data Case (e.g. labs & vitals)



The Structured Data Case (e.g. labs & vitals)





BRITS: Bidirectional Recurrent Imputation for Time Series

Wei Cao*
Tsinghua University
Bytedance AI Lab
cao-13@tsinghua.org.cn

Dong Wang
Duke University
dong.wang363@duke.edu

Jian Li
Tsinghua University
lijian83@mail.tsinghua.edu.cn

Hao Zhou
Bytedance AI Lab
haozhou0806@gmail.com

Yitan Li
Bytedance AI Lab
liyitan@bytedance.com

Lei Li
Bytedance AI Lab
lileilab@bytedance.com

GP-VAE: Deep Probabilistic Multivariate Time Series Imputation

Vincent Fortuin[1,2]
ETH Zürich, Switzerland

Dmitry Baranchuk[1,3]
Yandex, Russia

Gunnar Rätsch
ETH Zürich, Switzerland

Stephan Mandt
UC Irvine, USA

HHS Public Access
Author manuscript
ACM BCB. Author manuscript; available in PMC 2021 September 30.

Published in final edited form as:
ACM BCB. 2021 August ; 2021: . doi:10.1145/3459930.3469512.

CSDI: Conditional Score-based Diffusion Models fo Probabilistic Time Series Imputation

Yusuke Tashiro[123*], **Jiaming Song**[1], **Yang Song**[1], **Stefano Ermon**[1]
[1]Department of Computer Science, Stanford University, Stanford, CA, USA
[2]Mitsubishi UFJ Trust Investment Technology Institute, Tokyo, Japan
[3]Japan Digital Design, Tokyo, Japan
{ytashiro,tsong,songyang,ermon}@cs.stanford.edu

Concurrent Imputation and Prediction on EHR data using Bi-Directional GANs:

Bi-GANs for EHR imputation and prediction

Mehak Gupta*,
University of Delaware Newark, Delaware, USA

H. Timothy Bunnell,
Nemours Children's Health System Willmington, Delaware, USA

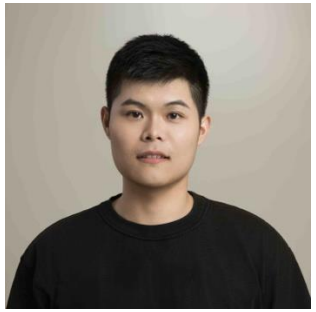
Thao-Ly T. Phan,
Nemours Children's Health System Willmington, Delaware, USA

Rahmatollah Beheshti
University of Delaware Newark, Dela

A Knowledge Distillation Ensemble Framework for Predicting Short and Long-term Hospitalisation Outcomes from Electronic Health Records Data

Zina M Ibrahim, Daniel Bean, Thomas Searle, Linglong Qian, Honghan Wu, Anthony Shek, Zeljko Kraljevic, James Galloway, Sam Norton, James T Teo, Richard JB Dobson

CSAI –Domain-informed Imputation & Prediction



Hugh Logan Ellis Linglong Qian

Principle of temporal decay: influence of past recordings on missing data decreases over time, with features imputed closer to default values if their last observation occurred a long time ago

$$\gamma_{th} = \exp\left(-\max(0, \boldsymbol{W}_\gamma \boldsymbol{\delta}_t + \boldsymbol{b}_\gamma)\right)$$

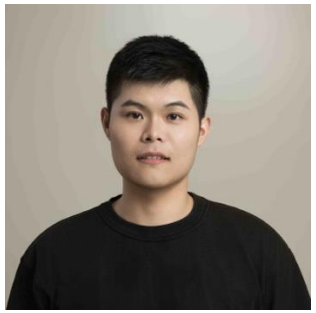
	t_1	t_2	t_3	t_4	t_5
f_1	5	4	/	8	9
f_2	7	/	/	/	9
f_3	2	4	1	6	/

Time

CSAI's domain-informed temporal decay: influence of past recordings on missing data decreases over time, constrained by each feature's ***median time-gap.***
The probability of a feature being imputed closer to default values only increases if the last observation happened earlier than the feature time gap.

$$\gamma_t^d = \exp(-\max(0, W_\gamma(\delta_t^d - \tau_d) + b_\gamma))$$

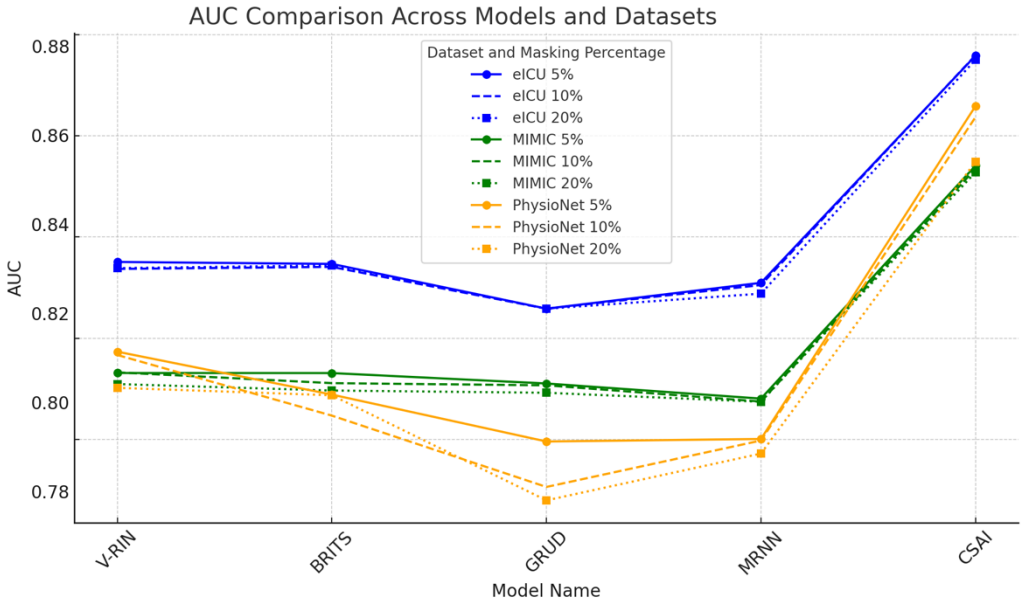
CSAI –Domain-informed Imputation & Prediction



Hugh Logan Ellis Linglong Qian

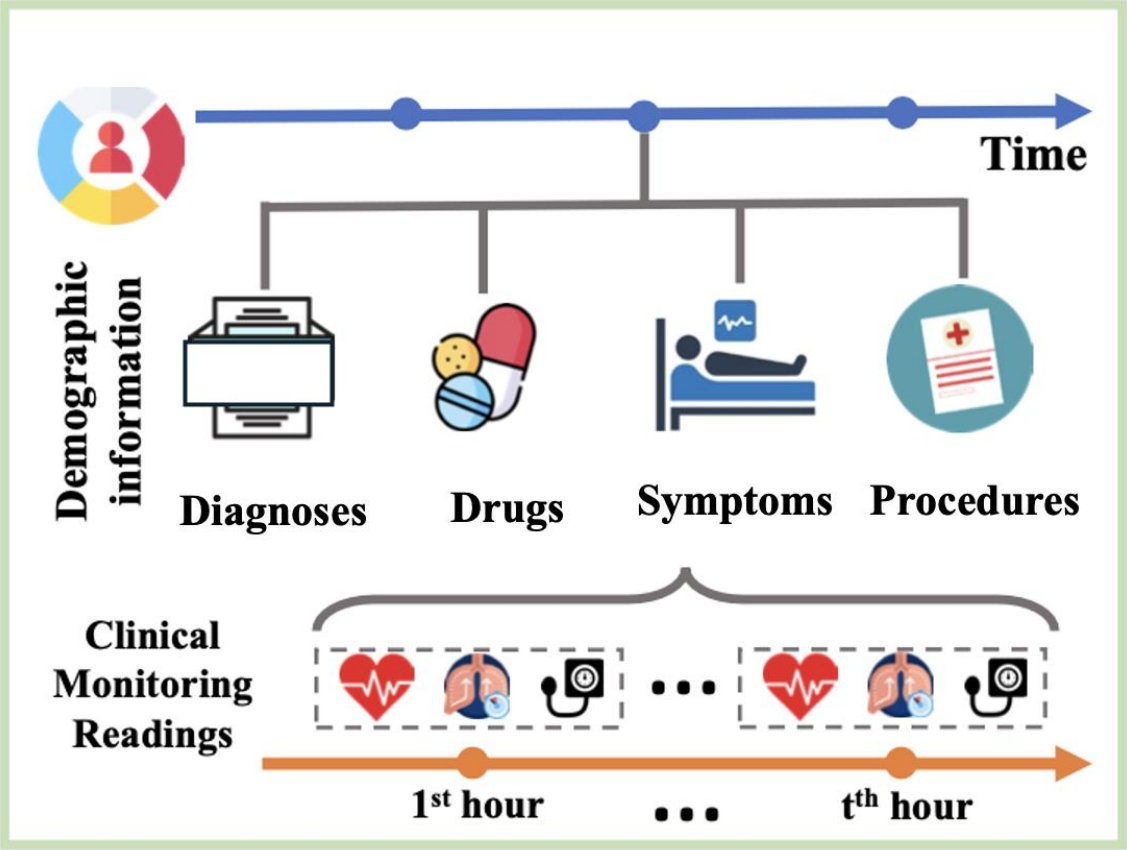
IMPUTATION PERFORMANCE USING 5%, 10%, AND 20% MASKING RATIOS ON THREE DATASETS. THE MODEL WITH THE LOWEST MAE IN EACH SETUP IS HIGHLIGHTED IN BOLD.

	eICU (MAE)			MIMIC_59 (MAE)			PhysioNet (MAE)		
	5%	10%	20%	5%	10%	20%	5%	10%	20%
V-RIN (N)	0.24161 ± 0.015	0.24254 ± 0.013	0.25214 ± 0.019	0.15457 ± 0.007	0.13818 ± 0.017	0.33697 ± 0.010	0.26163 ± 0.015	0.27372 ± 0.010	0.29997 ± 0.018
V-RIN (R)	0.24070 ± 0.020	0.24974 ± 0.011	0.26086 ± 0.008	0.20929 ± 0.019	0.22198 ± 0.012	0.44361 ± 0.016	0.26126 ± 0.012	0.27305 ± 0.019	0.29674 ± 0.015
BRITS (N)	0.16699 ± 0.014	0.17053 ± 0.020	0.17681 ± 0.009	0.15195 ± 0.018	0.14023 ± 0.009	0.34039 ± 0.019	0.25634 ± 0.013	0.26762 ± 0.017	0.28722 ± 0.014
BRITS (R)	0.17330 ± 0.019	0.18146 ± 0.012	0.19125 ± 0.017	0.19643 ± 0.018	0.19756 ± 0.008	0.40886 ± 0.015	0.25471 ± 0.016	0.26631 ± 0.011	0.28376 ± 0.020
GRUD (N)	0.22274 ± 0.018	0.22560 ± 0.010	0.23098 ± 0.020	0.30447 ± 0.012	0.28704 ± 0.014	0.48670 ± 0.017	0.49406 ± 0.015	0.49779 ± 0.020	0.50952 ± 0.018
GRUD (R)	0.22274 ± 0.016	0.22560 ± 0.015	0.23098 ± 0.018	0.28624 ± 0.014	0.24562 ± 0.012	0.39322 ± 0.015	0.49403 ± 0.020	0.49775 ± 0.011	0.50997 ± 0.019
MRNN (N)	0.47036 ± 0.015	0.47998 ± 0.017	0.50065 ± 0.020	0.30573 ± 0.013	0.28342 ± 0.012	0.47198 ± 0.015	0.54671 ± 0.013	0.55647 ± 0.014	0.57230 ± 0.017
MRNN (R)	0.47059 ± 0.019	0.48028 ± 0.020	0.50665 ± 0.016	0.31242 ± 0.017	0.30907 ± 0.010	0.50181 ± 0.018	0.54738 ± 0.019	0.55727 ± 0.017	0.57929 ± 0.019
CSAI (N)	**0.14967 ± 0.017**	**0.14149 ± 0.011**	**0.14637 ± 0.015**	**0.13119 ± 0.009**	**0.11291 ± 0.008**	**0.22976 ± 0.014**	**0.22602 ± 0.014**	**0.22747 ± 0.017**	**0.23476 ± 0.019**
CSAI (R)	0.14426 ± 0.020	0.14944 ± 0.016	0.14960 ± 0.017	0.16145 ± 0.019	0.15677 ± 0.011	0.36465 ± 0.018	0.24594 ± 0.016	0.25663 ± 0.015	0.27442 ± 0.016



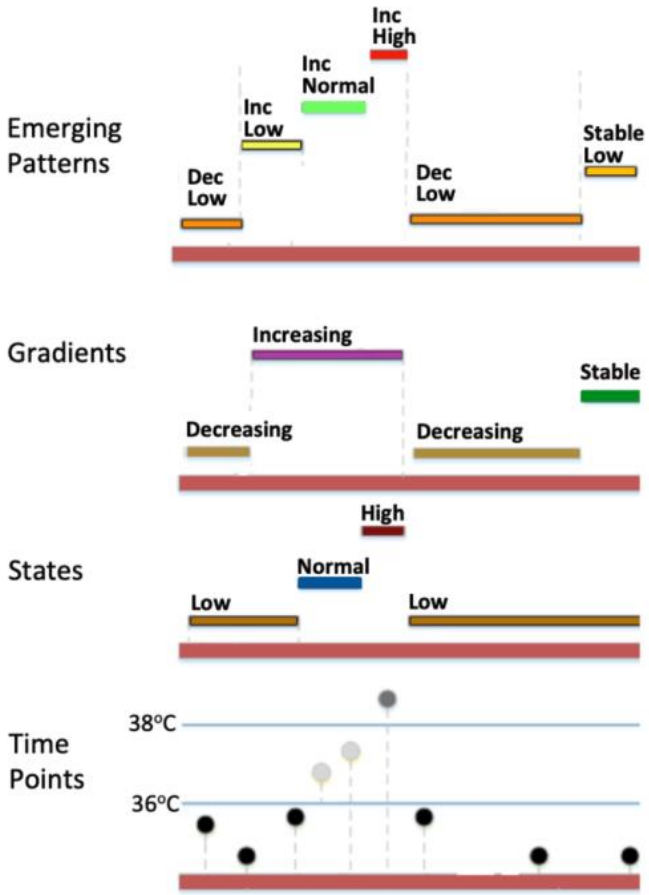
The Unstructured Case

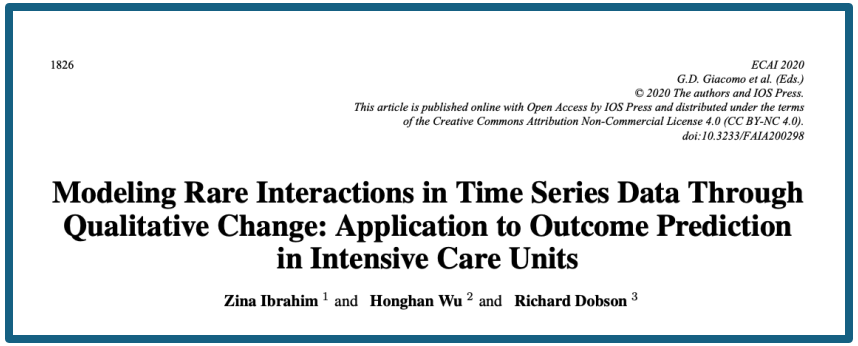
- Foresight does not operate on numerical data. It works with SNOMED codes.



Enriching Foresight with Domain-informed Measurements – the Qualitative Interaction Graph (QIG)

- Mine textual descriptions of vitals and lab test values for the CA cohort from the structured records
- Use the generated text as additional input to Foresight





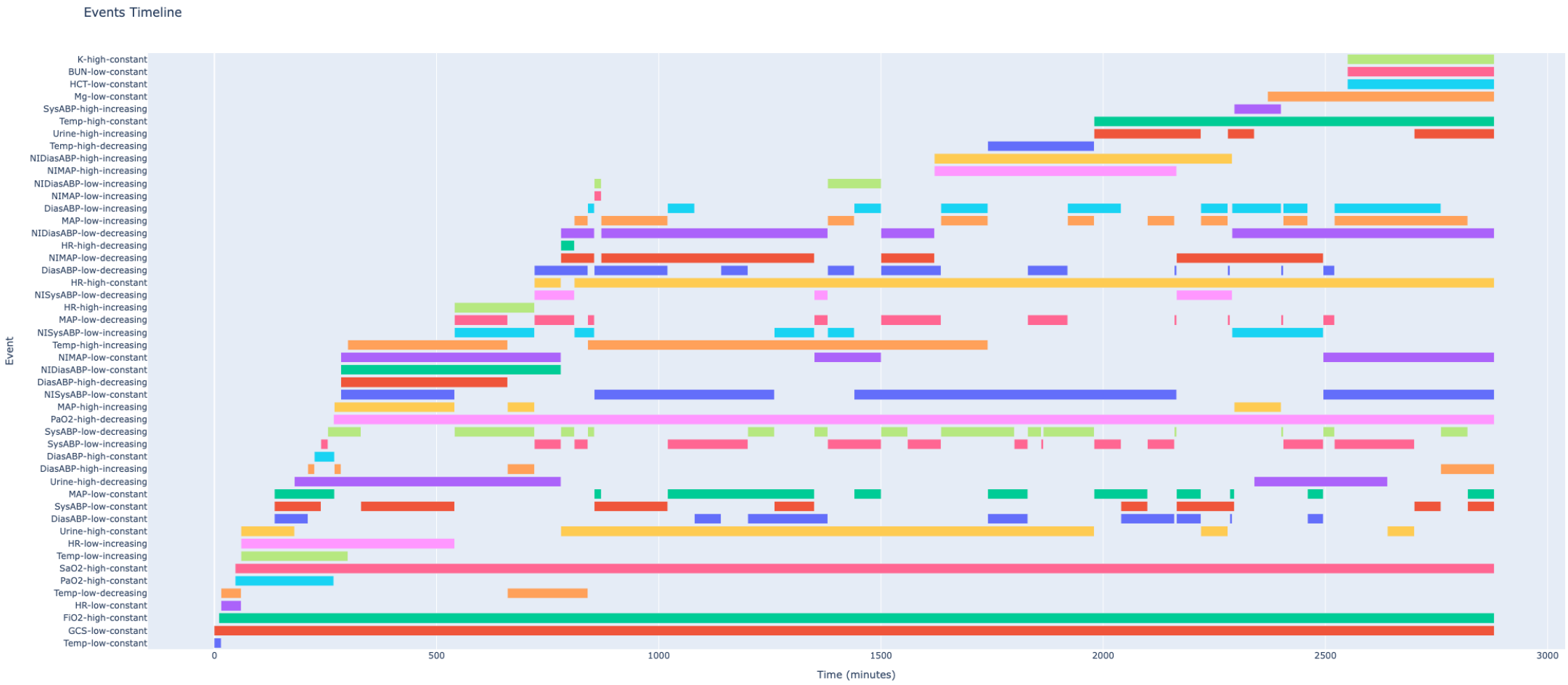
1826

ECAI 2020
G.D. Giacomo et al. (Eds.)
© 2020 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA200298

Modeling Rare Interactions in Time Series Data Through Qualitative Change: Application to Outcome Prediction in Intensive Care Units

Zina Ibrahim [1] and **Honghan Wu** [2] and **Richard Dobson** [3]

QIG for One Patient



QIG Textual Descriptor

- Currently…
  - Training a foundational model based on QIG descriptors

In [ ]:
```
descriptions
```

Out[ ]:
```
[{'event': 'Temp-low-constant starts at order 0, ends at order 0, lasts for 0.25 h',
  'relations': ['is equal to DiasABP-normal-constant, which starts at order 0, ends at order 0, lasts for 0.25 h',
   'is equal to SysABP-normal-constant, which starts at order 0, ends at order 0, lasts for 0.25 h',
   'is equal to HR-normal-constant, which starts at order 0, ends at order 0, lasts for 0.25 h',
   'is equal to MAP-normal-constant, which starts at order 0, ends at order 0, lasts for 0.25 h',
   'starts Urine-normal-constant, which starts at order 0, ends at order 3, lasts for 1.00 h',
   'starts GCS-low-constant, which starts at order 0, ends at order 11, lasts for 3.00 h',
   'overlaps with FiO2-high-constant, which starts at order 1, ends at order 3, lasts for 0.83 h',
   'meets HR-low-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets DiasABP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets SysABP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets MAP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets Temp-low-decreasing, which starts at order 2, ends at order 3, lasts for 0.75 h']},
 {'event': 'DiasABP-normal-constant starts at order 0, ends at order 0, lasts for 0.25 h',
  'relations': ['is equal to SysABP-normal-constant, which starts at order 0, ends at order 0, lasts for 0.25 h',
   'is equal to HR-normal-constant, which starts at order 0, ends at order 0, lasts for 0.25 h',
   'is equal to MAP-normal-constant, which starts at order 0, ends at order 0, lasts for 0.25 h',
   'starts Urine-normal-constant, which starts at order 0, ends at order 3, lasts for 1.00 h',
   'starts GCS-low-constant, which starts at order 0, ends at order 11, lasts for 3.00 h',
   'overlaps with FiO2-high-constant, which starts at order 1, ends at order 3, lasts for 0.83 h',
   'meets HR-low-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets DiasABP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets SysABP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets MAP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets Temp-low-decreasing, which starts at order 2, ends at order 3, lasts for 0.75 h']},
 {'event': 'SysABP-normal-constant starts at order 0, ends at order 0, lasts for 0.25 h',
  'relations': ['is equal to HR-normal-constant, which starts at order 0, ends at order 0, lasts for 0.25 h',
   'is equal to MAP-normal-constant, which starts at order 0, ends at order 0, lasts for 0.25 h',
   'starts Urine-normal-constant, which starts at order 0, ends at order 3, lasts for 1.00 h',
   'starts GCS-low-constant, which starts at order 0, ends at order 11, lasts for 3.00 h',
   'overlaps with FiO2-high-constant, which starts at order 1, ends at order 3, lasts for 0.83 h',
   'meets HR-low-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets DiasABP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets SysABP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets MAP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets Temp-low-decreasing, which starts at order 2, ends at order 3, lasts for 0.75 h']},
 {'event': 'HR-normal-constant starts at order 0, ends at order 0, lasts for 0.25 h',
  'relations': ['is equal to MAP-normal-constant, which starts at order 0, ends at order 0, lasts for 0.25 h',
   'starts Urine-normal-constant, which starts at order 0, ends at order 3, lasts for 1.00 h',
   'starts GCS-low-constant, which starts at order 0, ends at order 11, lasts for 3.00 h',
   'overlaps with FiO2-high-constant, which starts at order 1, ends at order 3, lasts for 0.83 h',
   'meets HR-low-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets DiasABP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets SysABP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets MAP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets Temp-low-decreasing, which starts at order 2, ends at order 3, lasts for 0.75 h']},
 {'event': 'MAP-normal-constant starts at order 0, ends at order 0, lasts for 0.25 h',
  'relations': ['starts Urine-normal-constant, which starts at order 0, ends at order 3, lasts for 1.00 h',
   'starts GCS-low-constant, which starts at order 0, ends at order 11, lasts for 3.00 h',
   'overlaps with FiO2-high-constant, which starts at order 1, ends at order 3, lasts for 0.83 h',
   'meets HR-low-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets DiasABP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets SysABP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets MAP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'meets Temp-low-decreasing, which starts at order 2, ends at order 3, lasts for 0.75 h']},
 {'event': 'Urine-normal-constant starts at order 0, ends at order 3, lasts for 1.00 h',
  'relations': ['starts GCS-low-constant, which starts at order 0, ends at order 11, lasts for 3.00 h',
   'is finished by FiO2-high-constant, which starts at order 1, ends at order 3, lasts for 0.83 h',
   'contains HR-low-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'contains DiasABP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'contains SysABP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'contains MAP-normal-decreasing, which starts at order 2, ends at order 1, lasts for 0.25 h',
   'is finished by Temp-low-decreasing, which starts at order 2, ends at order 3, lasts for 0.75 h',
   'contains SysABP-normal-increasing, which starts at order 3, ends at order 2, lasts for 0.25 h',
   'contains MAP-normal-increasing, which starts at order 3, ends at order 2, lasts for 0.25 h',
   'contains DiasABP-normal-increasing, which starts at order 3, ends at order 2, lasts for 0.25 h',
   'is finished by HR-low-constant, which starts at order 3, ends at order 3, lasts for 0.50 h',
```

Take home message

- Stacking (deep models) without a design can only get us so far…
- The answer does not only lie in more data
  - Knowledge-informed models can be trained with less data
- Collaboration is key

Thank you!