



NEW PUBLICATION NEWS

clairvoyance alpha: the first pipeline toolkit for medical time series



Mihaela van der Schaar



Jinsung Yoon



Zhaozhi Qian



Dan Jarrett



Ioana Bica

📅 May 15, 2020 ⌚ 8 min read

clairvoyance

The van der Schaar Lab has released an alpha version of clairvoyance, a ground-breaking package that represents the culmination of years of research, development, and real-world testing. As a unified, end-to-end pipeline for time-series data, clairvoyance is unmatched in its versatility and capability. While primarily developed to aid clinical research and decision support, clairvoyance can also be used to great effect in non-medical contexts, thanks to its ability to facilitate complex inference workflows in a transparent, reproducible, and efficient manner.

clairvoyance is an enormously important project for our team, and is the result of years of work across a number of different areas. It's the first of its kind: an end-to-end pipeline that can produce personalized and interpretable predictions and recommendations using time-series data. I have no doubt that





to the machine learning community.

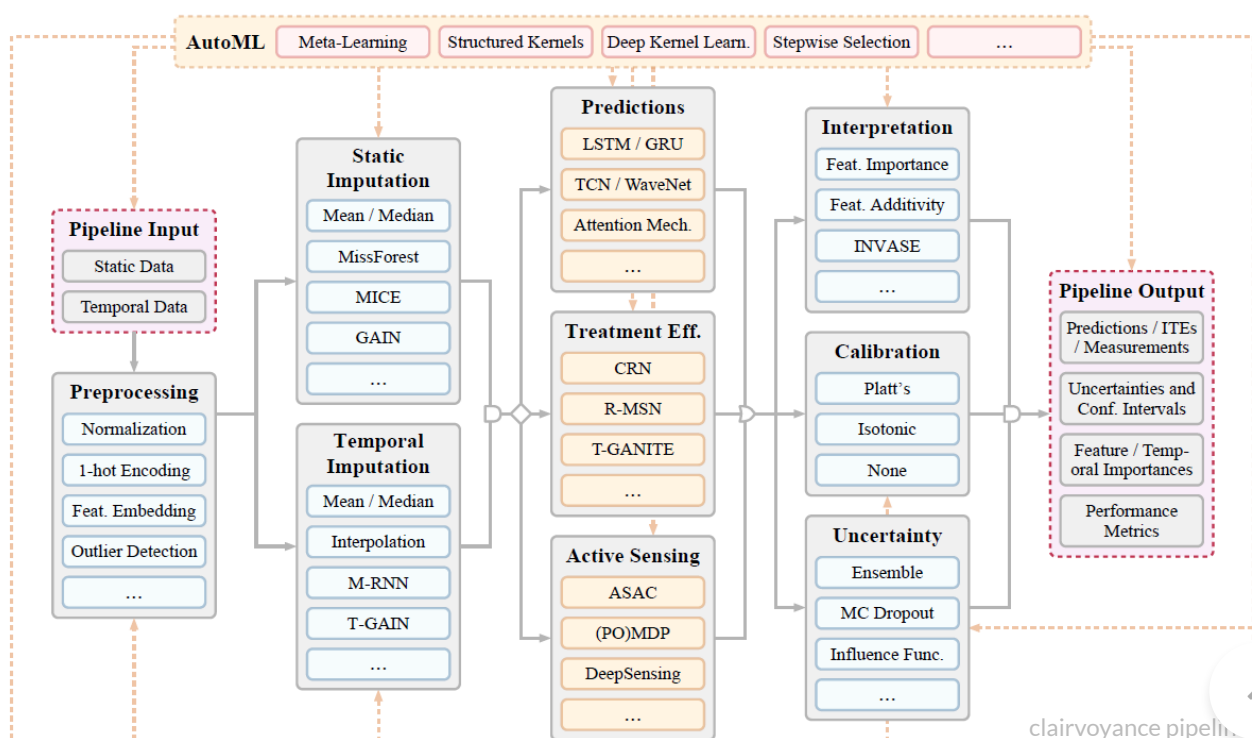
- PROF. MIHAELA VAN DER SCHAAR

A crucial breakthrough in the use of time-series data

Time-series data is the bread and butter of evidence-based clinical decision support, as it offers significantly more valuable insight than the “snapshots” presented by static data. With the increasing availability of electronic patient records, there is enormous untapped potential to apply machine learning to time-series data, providing accurate and actionable predictive models for real-world concerns.

At the same time, medical time-series problems in the wild are challenging due to their highly composite nature. Existing applications of machine learning to such problems have treated these component tasks as separate problems, leading to a siloed and stylized development approach that often fails to account for complexities and interdependencies within the real-world machine learning lifecycle. Regrettably, this has resulted in a remarkable gap between the inherent capabilities of machine learning methods and their actual effectiveness in clinical research and decision support.

clairvoyance offers a systematic and full-featured approach to personalized dynamic predictions, personalized information acquisition, personalized monitoring, and personalized treatment plans, while also offering interpretations.





problem definitions, (iv) handling missing or irregular samples in both static and temporal contexts, (v) conducting feature selection, (vi) fitting prediction models, performing (vii) calibration and (viii) uncertainty estimation of model outputs, (ix) applying global or instance-wise methods for interpreting learned models, and (x) computing evaluation metrics.

Existing time-series packages typically concentrate on implementing algorithms for specific problems, such as imputation, dynamic forecasting or feature extraction. By contrast, clairvoyance enables end-to-end development along every step of the inference workflow, including components critical to medical problems.

– DR. JINSUNG YOON (VAN DER SCHAAR LAB; GRADUATED 2020)

One solution to 3 problems

Designing real-world project lifecycles poses challenges in terms of engineering (difficult to build), evaluation (difficult to assess), and efficiency (difficult to optimize). clairvoyance was conceived to tackle all of these at once.

The **engineering problem** is that inference procedures involves significant investment, yet few resources are available for clinical practitioners and domain experts to easily develop and validate complete workflows. As a software toolkit, clairvoyance enables workflow development within a single unified interface: having modular and composable structures enables rapid experimentation and deployment by clinical practitioners and engineers alike, while simplifying collaboration and code-sharing.

The **evaluation problem** is that, while the performance of each component in the workflow may depend on the broader context, current empirical practices tend to examine the merits of each component individually; surrounding steps are largely configured per convenience to enforce “all else equal” conditions for assessment. This prevents meaningful assessment of the performance of the workflow as a whole, and does not meaningfully support integrated workflow development. As an empirical standard, clairvoyance serves as a complete experimental benchmarking environment: having a standardized, extensible pipeline paradigm provides realistic and systematic context for evaluating novel component models, ensuring that comparisons are fair, transparent, and reproducible.





computational difficulty compounded by the combination of multiple models, as well as by potential temporal distribution shifts in time series. Finally, through automated machine learning, clairvoyance takes care of model configuration and stepwise selection, efficiently considering interdependencies among components, algorithms, and time steps.

Estimating personalized treatment effects


Identifying when to give treatments to patients, and how to select among multiple treatments over time, are important medical problems with a few existing solutions. As clinical decision-makers are often faced with the problem of choosing between treatment alternatives for patients, reliably estimating their effects is paramount. An integral aspect of clairvoyance is the use of a novel counterfactual recurrent network (CRN) approach, developed by van der Schaar Lab's researchers, to estimate future treatment outcomes. This approach leverages recent advances in representation learning and domain adversarial training to overcome the problems of existing methods for causal inference over time. CRN achieves this in a manner that is free from the bias introduced by time-varying confounders. These developments were presented in a session during the 2020 International Conference on Learning Representations (ICLR 2020); further details can be found [here](#).

What is unique here is that we go beyond dynamic forecasting and provide actionable intelligence by estimating individualized treatment effects. clairvoyance is capable of predicting counterfactual trajectories for each patient under different possible treatment strategies, thus enabling us to determine when to give treatments to patients and how to select among multiple treatments over time.

- IOANA BICA (PH.D. STUDENT, VAN DER SCHAAR LAB)

Interpretable and actionable outputs

As a group that works extensively with clinicians in formulating problems and developing solutions, the van der Schaar Lab places particular emphasis on ensuring that models are interpretable and can be trusted by users without extensive machine learning expertise. This is an essential factor in being able to narrow the divide between the builders and users of machine learning models. Interpretability has therefore been built into clairvoyance through the inclusion of numerous interpretability features including INVASE, an in-house





In addition to interpretability, [clairvoyance](#) offers confidence estimates, ensuring that users are given an indication of the degree of certainty accompanying all recommendations or predictions.

Applications in medicine

clairvoyance is an inherently versatile tool, and can be applied to practically any disease or condition for which high-quality time-series patient datasets are available. It could also be adapted beyond a medical setting with relative ease.

Precursors to clairvoyance have been tested and validated using datasets for breast cancer and cystic fibrosis patients, as well as for ICU admission prediction (initially using past patient data in the U.S. and subsequently using live patient data in the U.K.). Results of testing have consistently demonstrated predictive capabilities exceeding those of existing statistical techniques or current state-of-the-art machine learning models. A more detailed analysis of the performance of the current version of clairvoyance can be found [here](#).

One potential application of clairvoyance in the near future could be the current COVID-19 pandemic. While the van der Schaar Lab is [already providing](#) AI-enabled predictive tools to the UK's National Health Service to help hospital capacity management, clairvoyance could offer even more comprehensive support for healthcare professionals, including:

- personalized predictions (mortality, discharge, ICU admission [including early warning] and readmission) and recommendations, both at time of admission and throughout the period of hospitalization;
- personalized information acquisition, including estimations of the value of certain information and determination of which tests should be conducted;
- personalized monitoring to determine which tests should be conducted and when to make forecasts, predictions, and treatment recommendations;
- personalized treatment plans to determine whether, how and when to admit patients to ICU, to use specific kinds of mechanical equipment, or to provide medication, as well as the likely impact of doing so; and
- transfer learning, enabling i) comparison of disease trajectories for patients before and after COVID-19, and ii) learning which information acquired from the numerous trajectories of non-COVID-19 patients can be effectively transferred to COVID-19 patients, and which information cannot.

Say a new patient walks in for something. As a matter of basic decision support, you'd want to be able to forecast some indicators, assess possible treatments, and decide what more





Looking to the future of clairvoyance

clairvoyance alpha is publicly available, and can be downloaded and tested. Development continues at pace, and the van der Schaar Lab team expects to release a beta version in the near future, followed by a full release subsequently. Further iterations will continue to be publicly available at no cost.

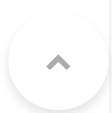
It is the research team's firm belief that, even in alpha, clairvoyance offers versatility and integration that cannot be found elsewhere. While today's announcement marks the release of the alpha version and a limited amount of supplementary information, further details will be shared during the coming months.

The alpha software can be found [here](#), and a more thorough introduction to clairvoyance is available [here](#).

Appendix: clairvoyance compared with related software

	cesium	tslearn	seglearn	tsfresh	pysf	sktime	pyts	clairvoyance
Preprocessing	✓	✓	✓	✓	✗	✓	✓	✓
Imputation	✓	✓	✓	✓	✓	✗	✓	✓
Feature Selection	✗	✗	✗	✓	✗	✗	✗	✓
Predictions	Context Data	✗	✓	✗	✗	✗	✗	✓
	Online Targets	✗	✗	✓	✗	✗	✓	✓
	Predictions	✓	✓	✓	✓	✓	✓	✓
	Treatment Effects	✗	✗	✗	✗	✗	✗	✓
	Active Sensing	✗	✗	✗	✗	✗	✗	✓
Interp.	Feat. Importance	✗	✗	✗	✗	✗	✗	✓
	Feat. Additivity	✗	✗	✗	✗	✗	✗	✓
	Instance-wise	✗	✗	✗	✗	✗	✗	✓
Calibration	✗	✗	✗	✗	✗	✗	✗	✓
Uncertainty	✗	✗	✗	✗	✗	✗	✗	✓
Pipelining	✗	✗	✗	✗	✗	✗	✗	✓
AutoML	✗	✗	✗	✗	✗	✗	✗	✓

Augmented MD AutoML Causal inference COVID-19 Ensemble learning
 Feature selection Interpretability Personalized medicine Reinforcement learning
 Time series analysis Transfer learning





Mihaela van der Schaar

Mihaela van der Schaar is the John Humphrey Plummer Professor of Machine Learning, Artificial Intelligence and Medicine at the University of Cambridge, a Fellow at The Alan Turing Institute in London, and a Chancellor's Professor at UCLA.

Mihaela has received numerous awards, including the Oon Prize on Preventative Medicine from the University of Cambridge (2018), a National Science Foundation CAREER Award (2004), 3 IBM Faculty Awards, the IBM Exploratory Stream Analytics Innovation Award, the Philips Make a Difference Award and several best paper awards, including the IEEE Darlington Award.

In 2019, she was identified by National Endowment for Science, Technology and the Arts as the most-cited female AI researcher in the UK. She was also elected as a 2019 "Star in Computer Networking and Communications" by N²Women. Her research expertise span signal and image processing, communication networks, network science, multimedia, game theory, distributed systems, machine learning and AI.

Mihaela's research focus is on machine learning, AI and operations research for healthcare and medicine.

[VIEW ALL POSTS](#)



Jinsung Yoon

Google Scholar: <https://scholar.google.com/citations?user=kiFd6A8AAAAJ>

[VIEW ALL POSTS](#)





Zhaozhi Qian

After obtaining a MSc in Machine Learning at UCL, Zhaozhi Qian started his career as a data scientist in the largest mobile gaming company in Europe. Three years later, he found it might be more fulfilling to apply AI to cure cancer than to make the gamers hit the purchase button 1% more often.

He thus joined the group in 2019 as a PhD student focusing on robust and interpretable learning for longitudinal data. So far, his work has included inferring latent disease interaction networks from Electronic Health Records, uncovering the causal structure between events that unfold over time, and calibrating the predictive uncertainty under domain shift.

Zhaozhi also worked as a contractor in the NHS during the COVID-19 pandemic contributing his analytical skills to the national response to the pandemic.

[VIEW ALL POSTS](#)



Dan Jarrett

Dan is a second-year Ph.D. student in the machine learning and artificial intelligence research group at the department of mathematics, advised by Professor van der Schaar.

He graduated from Princeton University with a B.A. in economics, and from Oxford with an MSc. in computer science.

He has professional experience in finance, consulting, and technology spaces, and research interests include representation learning and decision-making over time.

[VIEW ALL POSTS](#)





Ioana Bica

Ioana Bica is a second year PhD student at the University of Oxford and at the Alan Turing Institute. She has previously completed a BA and MPhil in Computer Science at the University of Cambridge where she has specialised in machine learning and its applications to biomedicine.

Ioana's PhD research focuses on building machine learning methods for causal inference and individualised treatment effect estimation from observational data. In particular, she has developed methods capable of estimating the heterogeneous effects of time-dependent treatments, thus enabling us to determine when to give treatments to patients and how to select among multiple treatments over time.

Recently, Ioana has started working on methods for understanding and modelling clinical decision making through causality, inverse reinforcement learning and imitation learning.

[VIEW ALL POSTS](#)



[Announcing the NeurIPS 2020 hide-and-seek privacy challenge](#)

[van der Schaar Lab at ICLR 2020: two papers and a keynote](#)



You may also like

NEWS

Spotlight on organ transplantation research projects

In this post, we explain how organ transplantation presents a web of complex problems and introduce our cutting-edge machine learning methods.



Mihaela van der Schaar



James Jordon



Jeroen Berrevoets



Nick Maxfield

📅 4 days ago





van der Schaar Lab at ICML 2021: 4 papers accepted

The van der Schaar Lab's researchers will publish 4 papers at ICML 2021, the leading international academic conference in machine learning.



Nick Maxfield  May 10, 2021

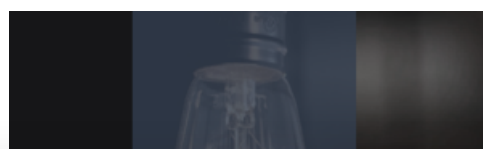
[EVENTS](#) [VIDEO](#)

Revolutionizing Healthcare: second roundtable on interpretability in ML/AI for healthcare

The van der Schaar Lab's seventh...



Nick Maxfield  May 3, 2021





sessions



View posts about...

AISTATS Augmented MD
AutoML Cambridge Adjutorium
Cancer Causal inference
Clinical trials Comorbidities
Competing risks Conference
COVID-19 Cystic Fibrosis
Deep learning
Ensemble learning
Feature selection ICLR
ICME ICML Interpretability
ITE inference Journal
Missing data imputation MLHC
Model NeurIPS Omics





Personalized education

Personalized medicine Pharma

Policy Impact Predictor

Reinforcement learning

Research team Software

Statistics Survival analysis

Synthetic data

Time series analysis

Transfer learning Transplant

Upcoming events

20

MAY

**AI FRONTIERS
SUMMIT KEYNOTE**

21

MAY

**AMBA-SERIES –
COVID-19 & AI:
UNEXPECTED
CHALLENGES AND
LESSONS**

27

MAY

**REVOLUTIONIZING
HEALTHCARE #8**

28

MAY

**INSPIRATION
EXCHANGE #9**

06

JUN

**IEEE DATA SCIENCE
AND LEARNING
WORKSHOP 2021
KEYNOTE**





Copyright ©2021 van der Schaar Lab

