EIVIA 2025: Deep Learning for Time Series and Applications to Healthcare

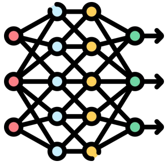
Gonzalo Uribarri
KTH Royal Institute of Technology & SciLifeLab

@gonzauri

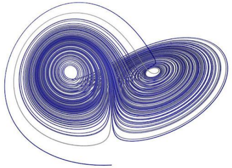
About me

Argentina (Phd)

Gabo Mindlin



Machine Learning

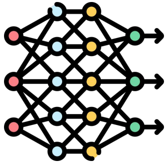


Dynamical Systems

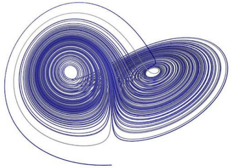
About me

Argentina (Phd)

Gabo Mindlin



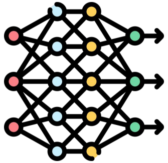
Machine Learning



Dynamical Systems

Sweden (Postdoc)

Erik Fransén



Machine Learning

Biomedical Data

About the course

About the course

About the course

The goal of the course is for you to learn:

- ❏ An Overview of Time Series Classification Methods
- ❏ Opportunities and Challenges of Applications in Healthcare

About the course

The goal of the course is for you to learn:

- ❑ An Overview of Time Series Classification Methods
- ❑ Opportunities and Challenges of Applications in Healthcare

Things we will **not cover** in the course:

- ❑ An Introduction to ML Fundamentals
- ❑ Details about the Technical Implementation of the Models

Course Plan

1/ INTRODUCTION

- Challenges of ML for healthcare
- Introduction to ML for time series

2/ TIME SERIES:
Standard Algorithms

- Basic ML for TS Classification
- Deep Learning for TS Classification

3/ TIMES SERIES:
State-of -the-art

- SOTA TS Classification models
- ROCKET and InceptionTime

4/ APPLICATIONS

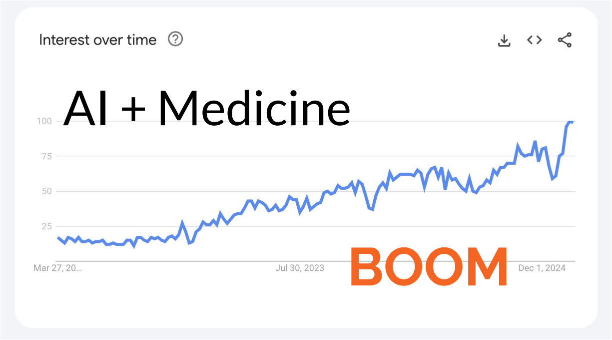
- Case studies in healthcare: Eye-tracking and EEG data
- Proper Evaluation

5/ LARGE MODELS

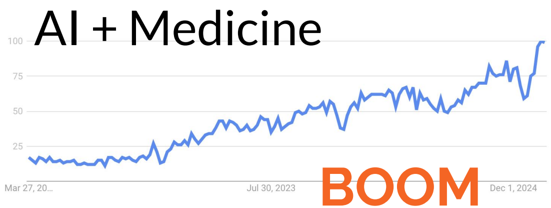
- TS models for large datasets
- Transfer learning: Foundational models for TS?

ML for Healthcare

ML for Healthcare



ML for Healthcare



scientific reports

Explore content · About the journal · Publish with us

nature > scientific reports > articles > article

Article | Open access | Published: 13 February 2024

Deep learning predicts prevalent and incident Parkinson's disease from UK Biobank fundus imaging

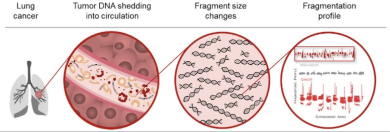
Charlie Tran, Kai Shen, Kang Liu, Akshay Ashok, Adolfo Ramirez-Zamora, Jinghua Chen, Yulin Li & Ruogu Fang

Scientific Reports 14, Article number: 3637 (2024) | Cite this article

5525 Accesses | 10 Citations | 39 Altmetric | Metrics

Artificial Intelligence Blood Test Provides a Reliable Way to Identify Lung Cancer

06/06/2024



npj | digital medicine

Explore content · About the journal · Publish with us

nature > npj digital medicine > news & views > article

News & Views | Open access | Published: 25 November 2024

Artificial Intelligence awarded two Nobel Prizes for innovations that will shape the future of medicine

Ben Li & Stephen Gilbert

npj Digital Medicine 7, Article number: 336 (2024) | Cite this article

12k Accesses | 26 Altmetric | Metrics

nature

Explore content · About the journal · Publish with us

nature > articles > article

Article | Published: 04 September 2024

A pathology foundation model for cancer diagnosis and prognosis prediction

Xiyue Wang, Junhan Zhao, Eliana Marostica, Wei Yuan, Jietian Jin, Jiayu Zhang, Ruijiang Li, Hongping Tang, Kanran Wang, Yu Li, Fang Wang, Yulong Peng, Junyou Zhu, Jing Zhang, Christopher R. Jackson, Jun Zhang, Deborah Dillon, Nancy U. Lin, Lynette Sholl, Thomas Denize, David Meredith, Keith L. Ligon, Sabina Signoretti, Shuji Ogino, ... Kun-Hsing Yu + Show authors

Nature 634, 970–978 (2024) | Cite this article

38k Accesses | 365 Altmetric | Metrics

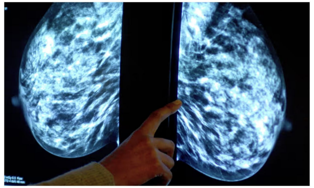
The Guardian

ion · Sport · Culture · Lifestyle

edia · Society · Law · Scotland · Wales · Northern Ireland

NHS to launch world's biggest trial of AI breast cancer diagnosis

If successful, the scheme could speed up testing and reduce radiologists' workload by around half



At present, the NHS uses a 'second reader' system where two radiologists check mammograms for signs of breast cancer. Photograph: Rui Vieira/PA

Most viewed

- Five shot in attack i… Swedish city of Öreb…
- Live · Sweden shooting: fi… people shot at educ… centre in Örebro by a… police believe is amo… injured - latest upda…
- Donald Trump Jr ac… killing protected bir… Venice lagoon
- 'My most sincere ap… to all the people wh… have been offended'… Oscar campaigns im…
- Live · Robert F Kennedy Jr… health secretary nomination advanc… Senate vote - live

ML for Healthcare

Why is there a boom?

- Data Availability
- Better Models
- Hype

ML for Healthcare

Why is there a boom?

- Data Availability
- Better Models
- Hype

But… CHALLENGES

- Data Privacy
- Interpretability / Explainability
- Generalization (Bias, Fairness)
- Deployment to Clinics

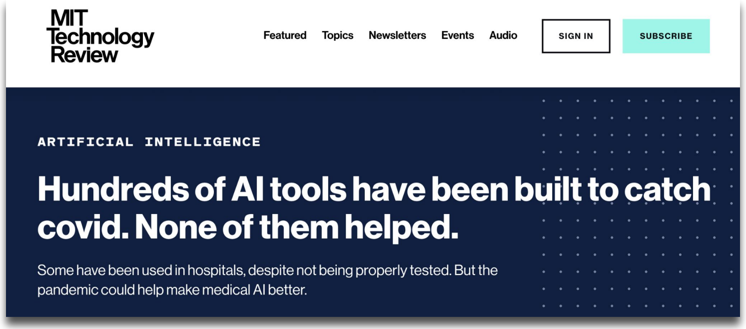
ML for Healthcare

Why is there a boom?

- Data Availability
- Better Models
- Hype

But… CHALLENGES

- Data Privacy
- Interpretability / Explainability
- Generalization (Bias, Fairness)
- Deployment to Clinics


MIT Technology Review

Featured Topics Newsletters Events Audio SIGN IN SUBSCRIBE

ARTIFICIAL INTELLIGENCE

Hundreds of AI tools have been built to catch covid. None of them helped.

Some have been used in hospitals, despite not being properly tested. But the pandemic could help make medical AI better.

ML for Healthcare

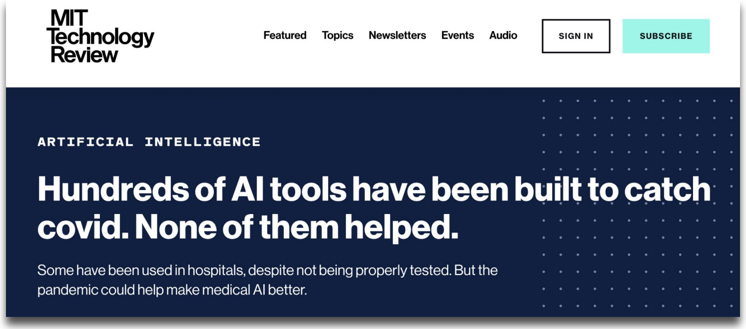
Why is there a boom?

- Data Availability
- Better Models
- Hype

But… CHALLENGES

- Data Privacy
- Interpretability / Explainability
- Generalization (Bias, Fairness)
- Deployment to Clinics

Ethical / Technical


MIT Technology Review

Featured Topics Newsletters Events Audio SIGN IN SUBSCRIBE

ARTIFICIAL INTELLIGENCE

Hundreds of AI tools have been built to catch covid. None of them helped.

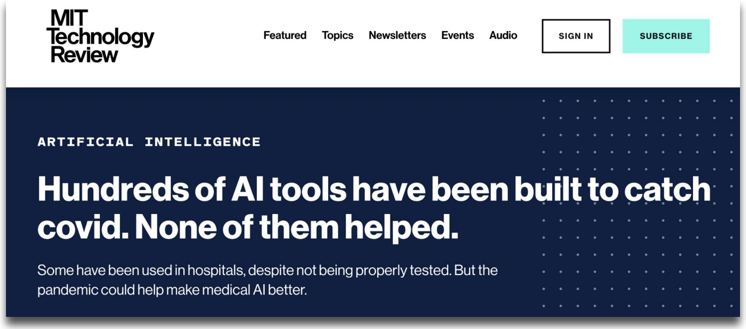
Some have been used in hospitals, despite not being properly tested. But the pandemic could help make medical AI better.

ML for Healthcare

Why is there a boom?

- Data Availability
- Better Models
- Hype

But… CHALLENGES



- Data Privacy
- Interpretability / Explainability
- Generalization (Bias, Fairness)
- Deployment to Clinics

Ethical / $\boxed{\text{Technical}}$

Challenge in Clinical Datasets

Ideal Classification Setting

Subjects

Condition (Label)

A B

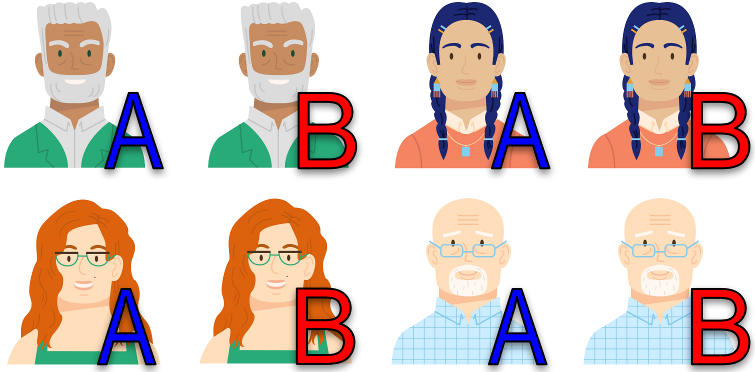
Ideal Classification Setting

Subjects

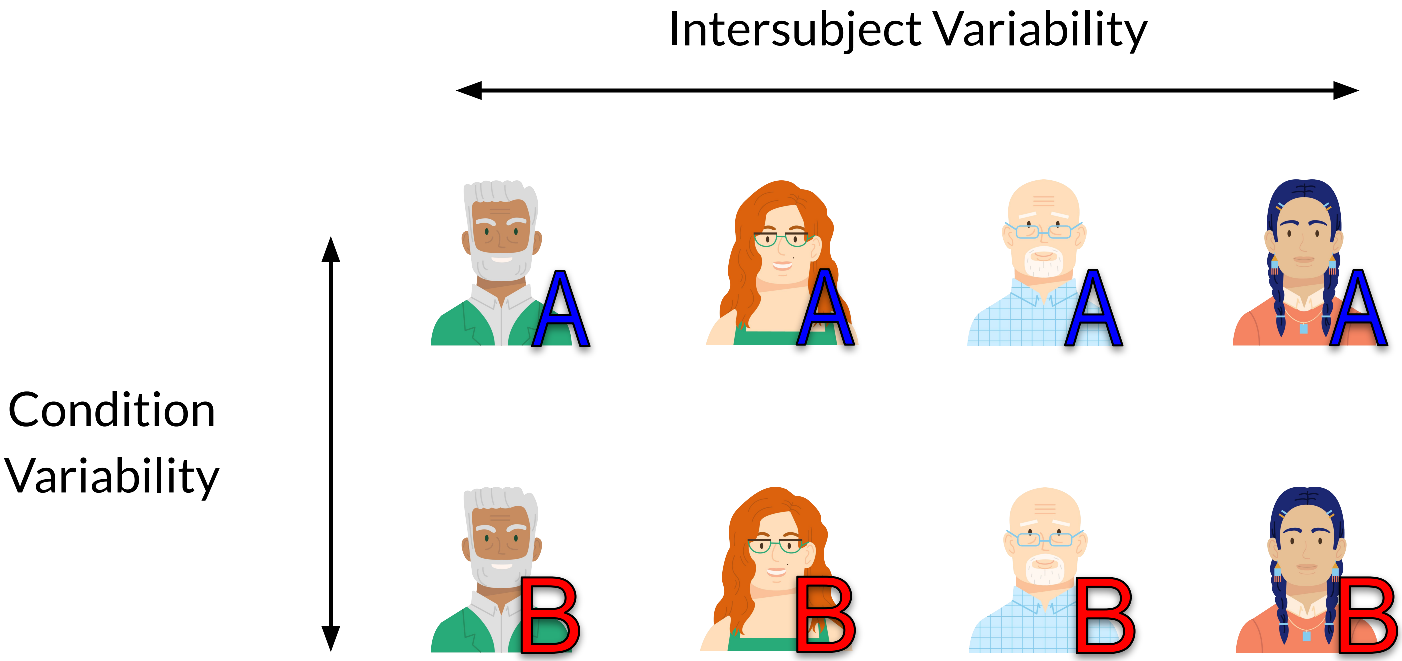
Condition (Label)

A B

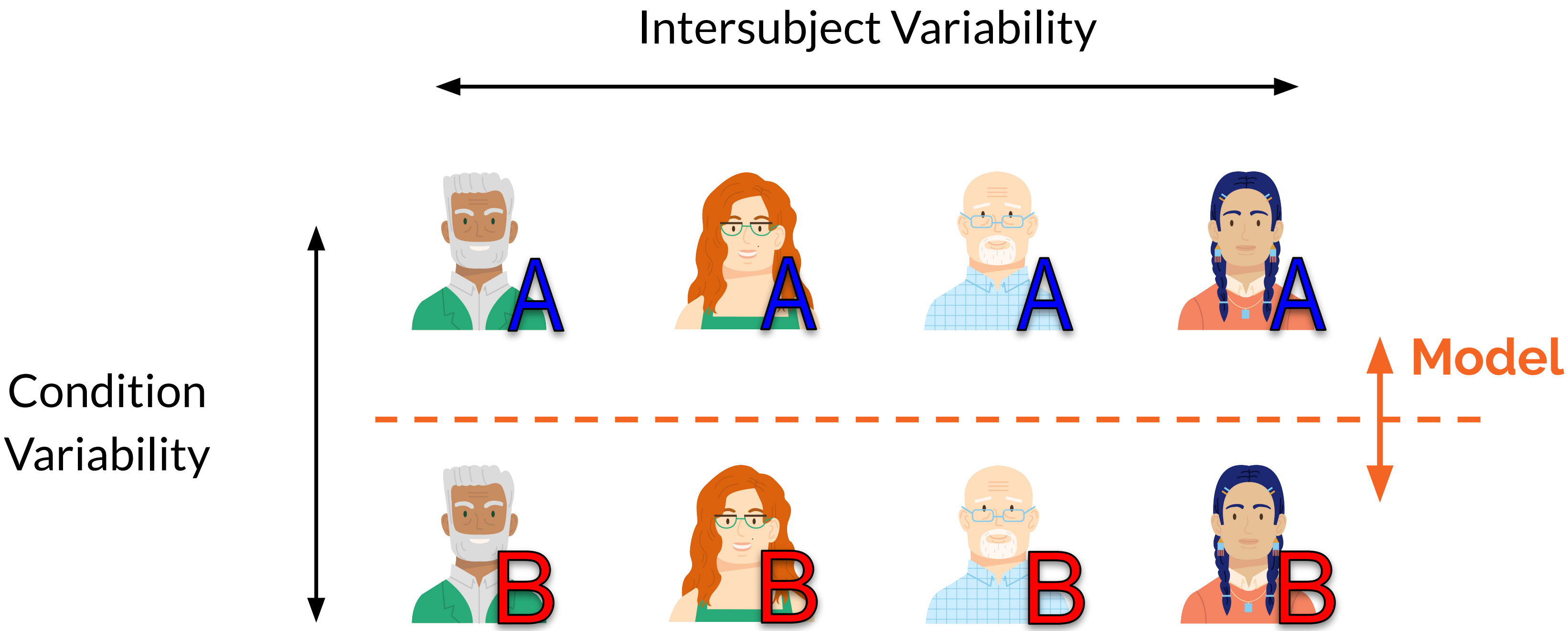
Ideal dataset (8 trials)



Ideal Classification Setting



Ideal Classification Setting



Challenging Classification Setting

Subjects

Condition (Label)

A B

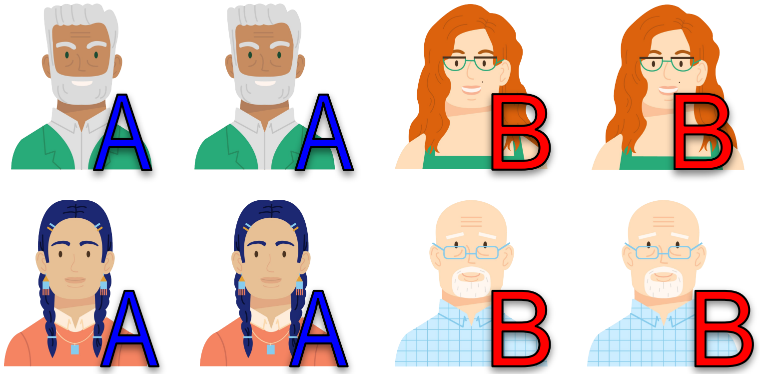
Challenging Classification Setting

Subjects

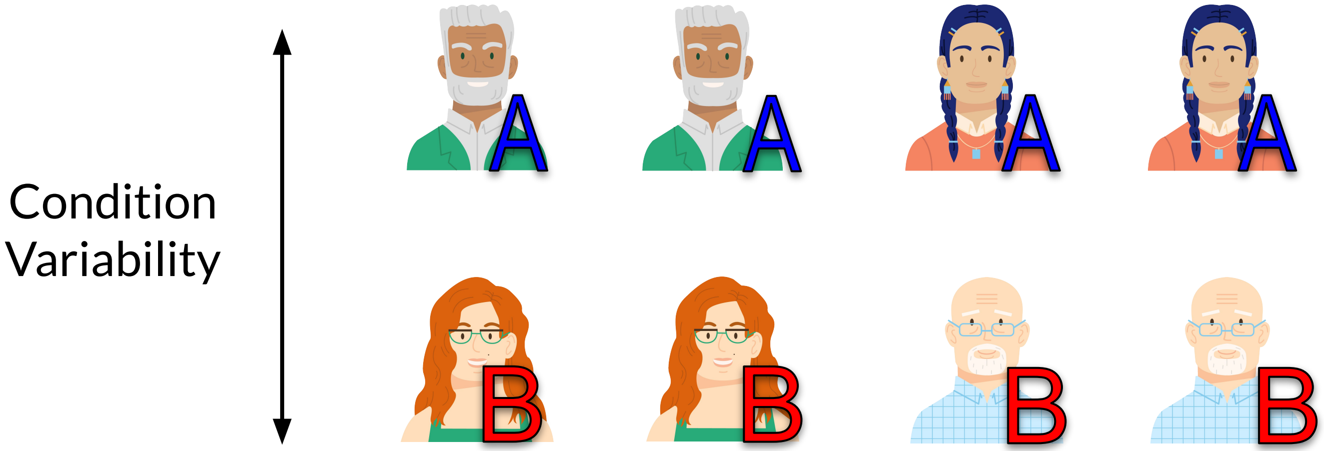
Condition (Label)

A B

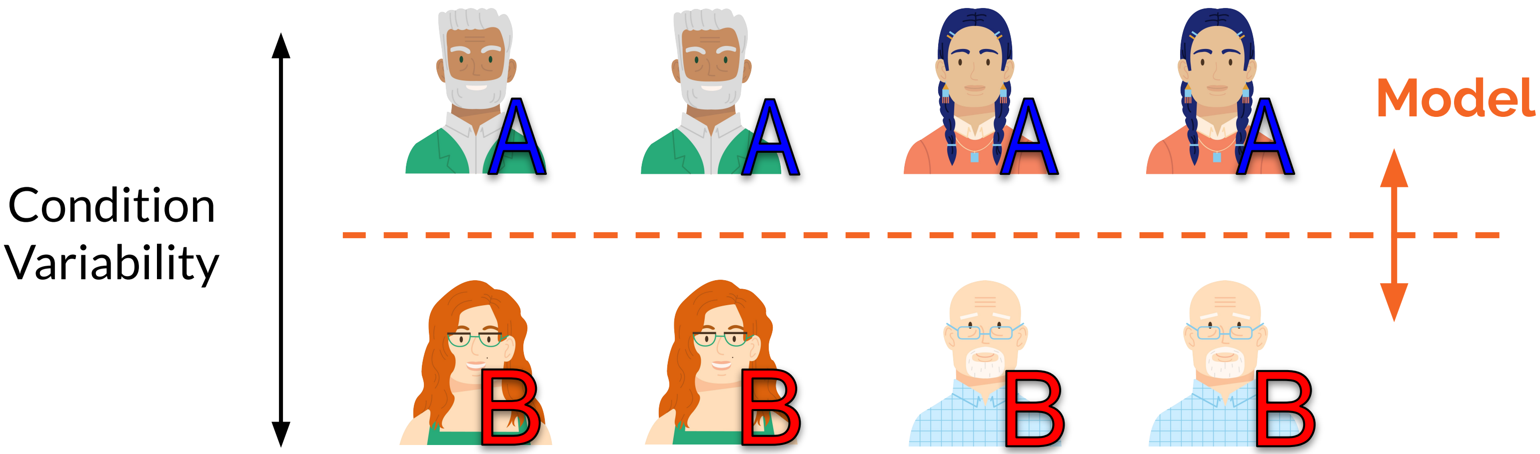
Neurodegenerative disease Dataset



Challenging Classification Setting

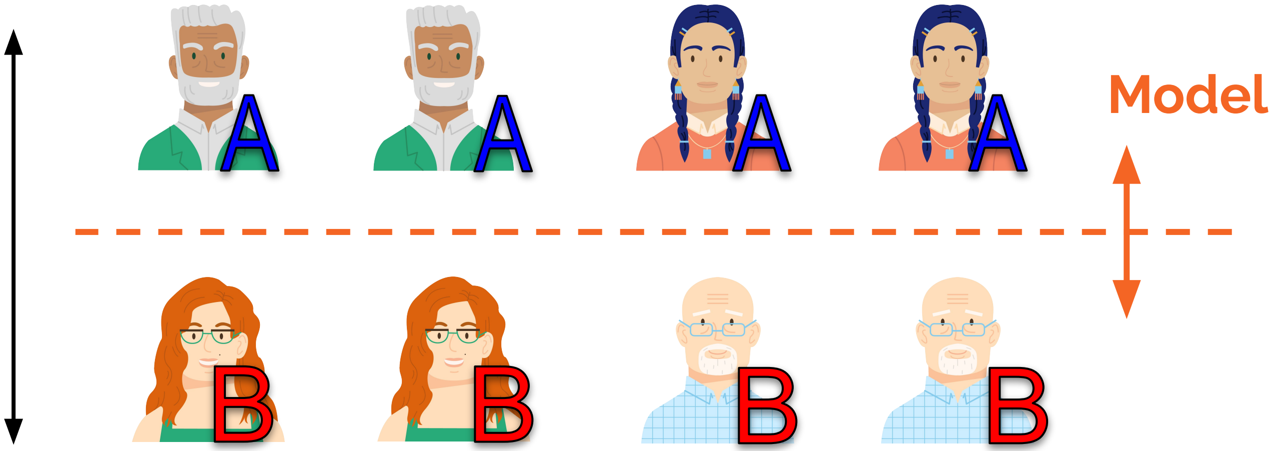


Challenging Classification Setting



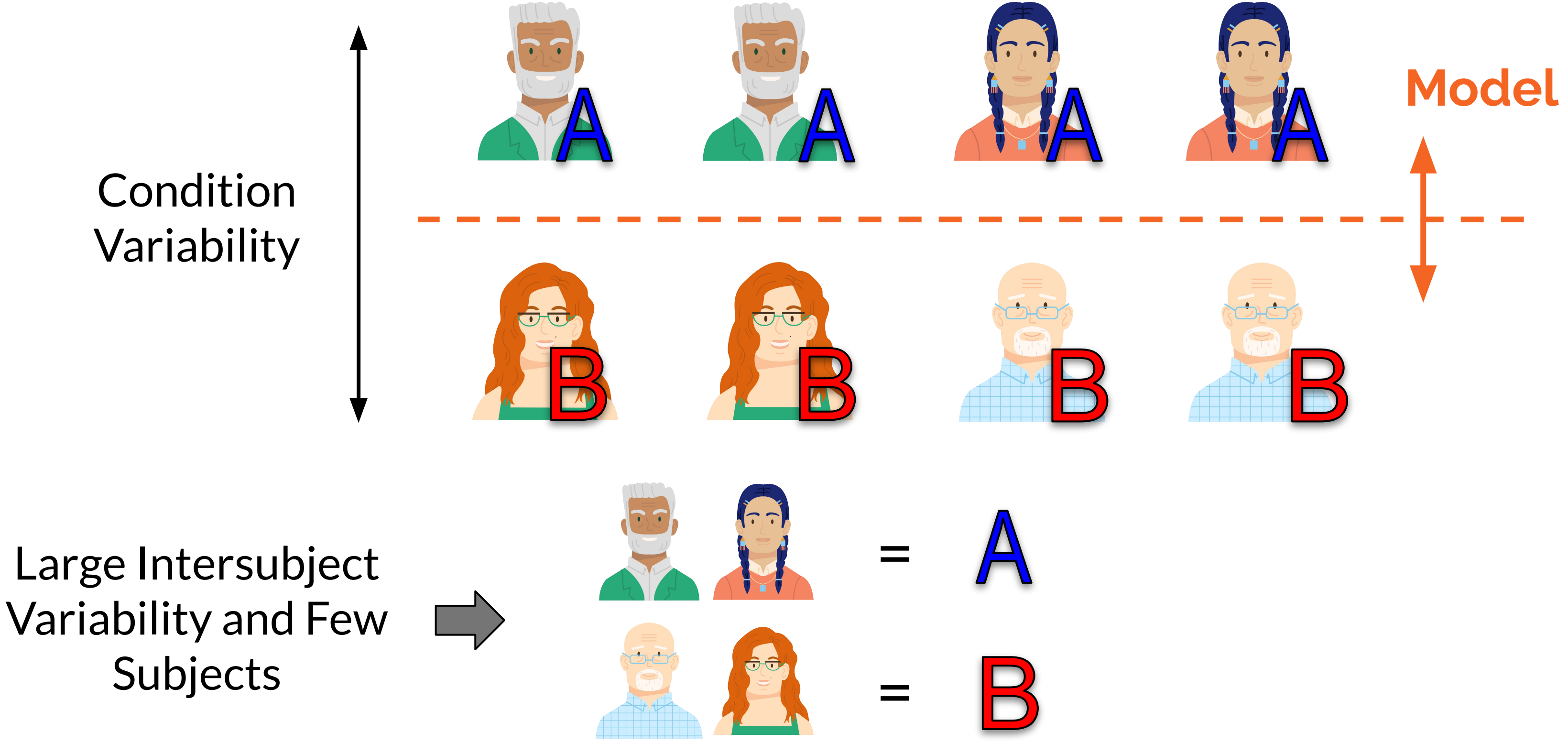
Challenging Classification Setting

Condition Variability



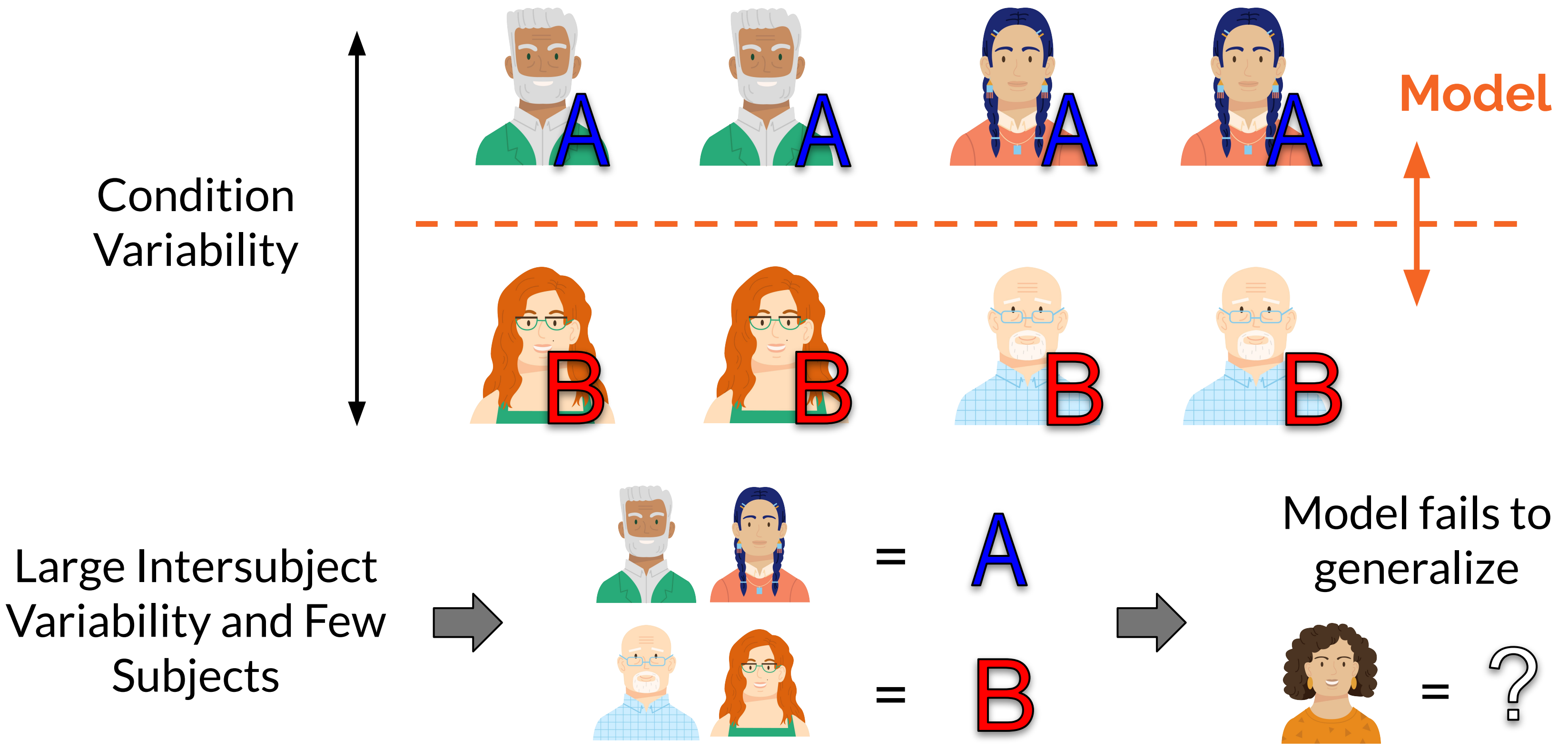
Large Intersubject Variability and Few Subjects

Challenging Classification Setting



Comment on individual variability (useful up to some point)

Challenging Classification Setting



Overcoming the Generalization Problem

Possible solutions:

1) **Increase Number of Participants**

Overcoming the Generalization Problem

Possible solutions:

1) **Increase Number of Participants.**
2) **Longitudinal Study:**

Overcoming the Generalization Problem

Possible solutions:

1) **Increase Number of Participants.**
2) **Longitudinal Study**:

3) **Data Specificity:** Select an experimental task and data modality where the condition variability is large.

Overcoming the Generalization Problem

Possible solutions:

1) **Increase Number of Participants.**
2) **Longitudinal Study**:

3) **Data Specificity:** Select an experimental task and data modality where the condition variability is large.
4) **Machine Learning Framework Design:** Do feature engineering and/or impose a particular inductive bias on the ML model.

Overcoming the Generalization Problem

Possible solutions:

Experimental Design

1) **Increase Number of Participants.**
2) **Longitudinal Study**:

3) **Data Specificity:** Select an experimental task and data modality where the condition variability is large.
4) **Machine Learning Framework Design:** Do feature engineering and/or impose a particular inductive bias on the ML model.

Overcoming the Generalization Problem

Possible solutions:

Experimental Design

1) **Increase Number of Participants.**
2) **Longitudinal Study**:

3) **Data Specificity:** Select an experimental task and data modality where the condition variability is large.

This is where we work

4) **Machine Learning Framework Design:** Do feature engineering and/or impose a particular inductive bias on the ML model.

Real World Example (preview)

The Dataset

- The dataset consists of **84** subjects in total:
  - **54** patients with Parkinson's disease
  - **30** Healthy controls

The Dataset

- ❑ The dataset consists of **84** subjects in total:
  - ❑ **54** patients with Parkinson's disease
  - ❑ **30** Healthy controls
- ❑ Around **100** trials per subject.
- ❑ Two different data modalities are recorded:
  - ❑ MEG (Magnetoencephalography)
  - ❑ Eye-Tracking

The Dataset

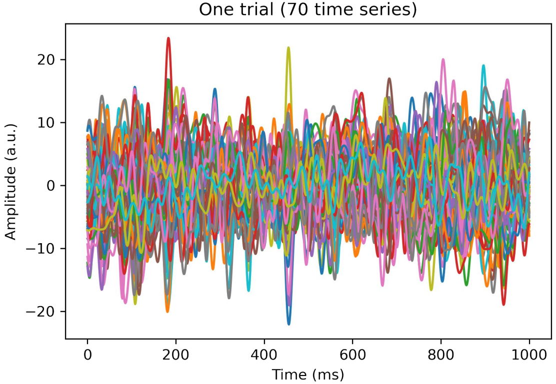
- ❑ The dataset consists of **84** subjects in total:
  - ❑ **54** patients with Parkinson's disease
  - ❑ **30** Healthy controls
- ❑ Around **100** trials per subject.
- ❑ Two different data modalities are recorded:

Time Series
  - ❑ MEG (Magnetoencephalography)
  - ❑ Eye-Tracking

Task

Brain activity recording (1 trial)

Multichannel Data
(High-dimensional)



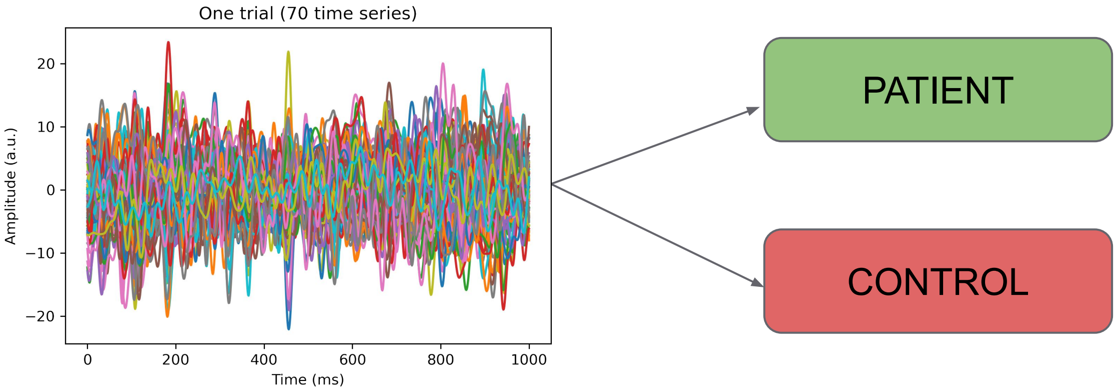
1s of recording

Task

Brain activity recording (1 trial)

Can we classify?
(Diagnosis task)

Multichannel Data
(High-dimensional)



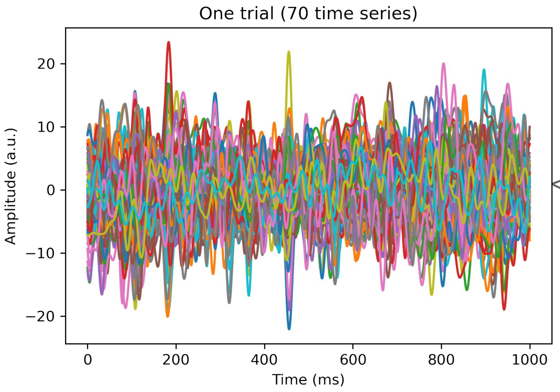
1s of recording

Task

Brain activity recording (1 trial)

Can we classify?
(Diagnosis task)

Multichannel Data
(High-dimensional)



1s of recording

PATIENT

CONTROL

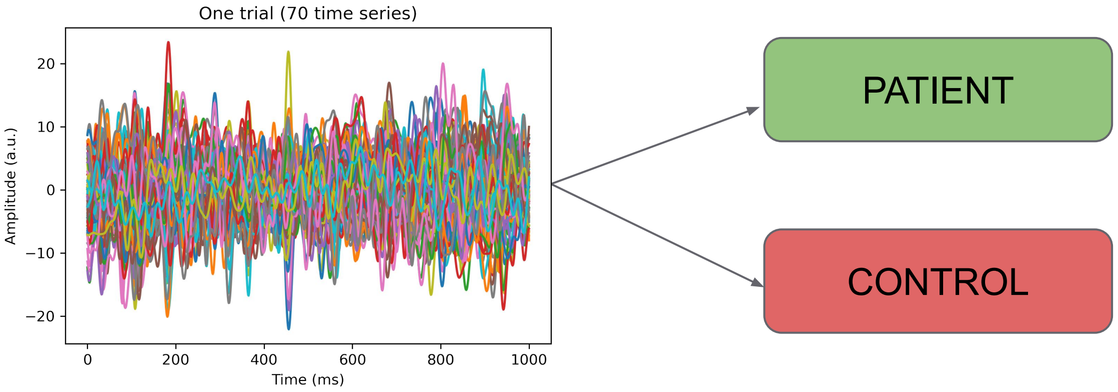
We have **8400** trials…
but only **84** subjects.

Task

Brain activity recording (1 trial)

Can we classify?
(Diagnosis task)

Multichannel Data
(High-dimensional)

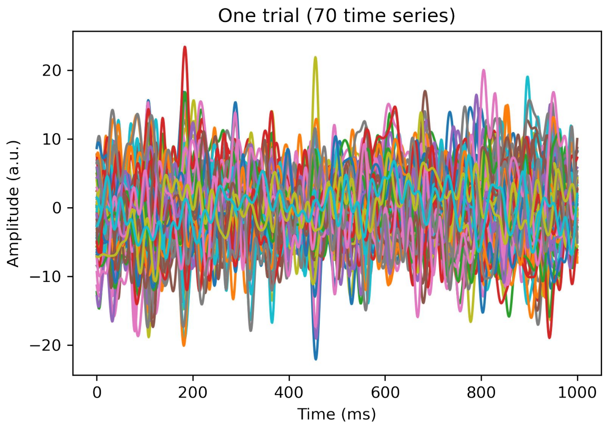


1s of recording

We have **8400** trials...
but only **84** subjects.

It is **really challenging** to generalize tu new subjects

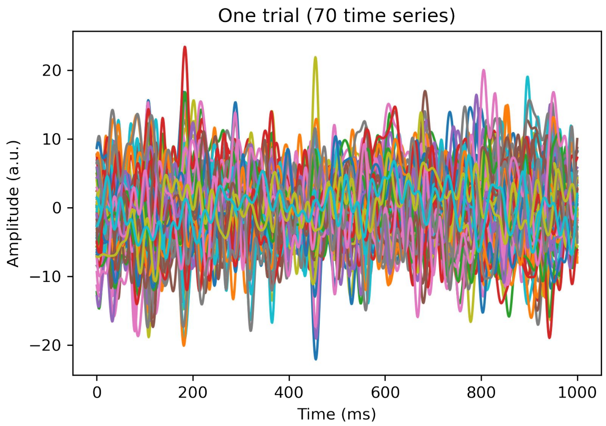
Our Goal



We have **8400** high-dimensional trials and only **84** subjects.

We need to **carefully design** ML models that can deal with this **high-dimensional data** and **generalize**.

Our Goal

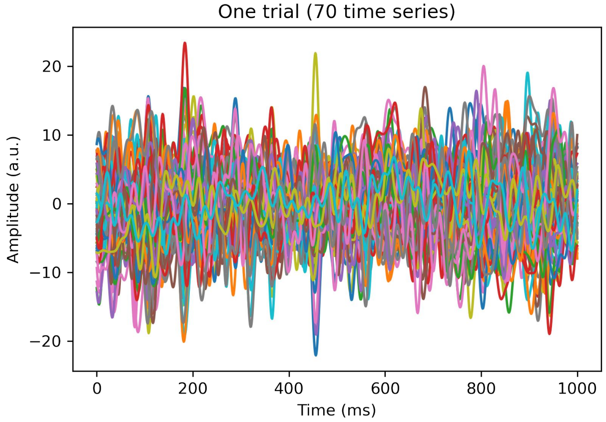


We have **8400** high-dimensional trials and only **84** subjects.

We need to **carefully design** ML models that can deal with this **high-dimensional data** and **generalize**.

We have to exploit **properties** of the data modality.

Our Goal



We have **8400** high-dimensional trials and only **84** subjects.

We need to **carefully design** ML models that can deal with this **high-dimensional data** and **generalize**.

We have to exploit **properties** of the data modality.

We are going to explore how can we do this for Time Series Data

Machine Learning for Time Series

Sequential data

Sequential data is data arranged in sequences where order matters. Data points are conditioned on other data points in the sequence.

Sequential Data

Sequential data

Sequential data is data arranged in sequences where order matters. Data points are conditioned on other data points in the sequence.

Sequential Data

Discrete Sequences	Time Series
Series of discrete tokens where order matters, but there is no explicit time variable	Quasi-continuous numerical values that evolve as a function of time.
E.g., Text, DNA, List of events	**E.g., ECG, Sound, Temperature**

Sequential data

Sequential data is data arranged in sequences where order matters. Data points are conditioned on other data points in the sequence.

Sequential Data

Transformers

Discrete Sequences	Time Series
Series of discrete tokens where order matters, but there is no explicit time variable	Quasi-continuous numerical values that evolve as a function of time.
E.g., Text, DNA, List of events	**E.g., ECG, Sound, Temperature**

Linear RNNs: xLSTM, Mamba, RWKV

Sequential data

Sequential data is data arranged in sequences where order matters. Data points are conditioned on other data points in the sequence.

Sequential Data

Transformers

Discrete Sequences	Time Series
Series of discrete tokens where order matters, but there is no explicit time variable	Quasi-continuous numerical values that evolve as a function of time.
E.g., Text, DNA, List of events	**E.g., ECG, Sound, Temperature**

Linear RNNs: xLSTM, Mamba, RWKV

Focus

Tasks for Time Series Data

Machine Learning for Time Series

Supervised

- Classification
- Extrinsic Regression
- Forecasting

Unsupervised

- Clustering
- Compression
- Generative
- Outlier Detection

Tasks for Time Series Data

Machine Learning for Time Series

Supervised

Focus
- Classification
- Extrinsic Regression
- Forecasting

Unsupervised

- Clustering
- Compression
- Generative
- Outlier Detection

ML for Time Series Classification

There are two standard archives for benchmarking:

- ❏ UCR

  142 Univariate Time Series Datasets

- ❏ UEA

  30 Multivariate Time Series Datasets

ML for Time Series Classification

There are two standard archives for benchmarking:

- UCR

  142 Univariate Time Series Datasets

- UEA

  30 Multivariate Time Series Datasets

Great reviews:

SPRINGER NATURE Link

Find a journal | Publish with us | Track your research | Search

Home > Data Mining and Knowledge Discovery > Article

Bake off redux: a review and experimental evaluation of recent time series classification algorithms

Open access | Published: 19 April 2024

Volume 38, pages 1958–2031, (2024) Cite this article

SPRINGER NATURE Link

Find a journal | Publish with us | Track your research | Search

Home > Data Mining and Knowledge Discovery > Article

The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances

Open access | Published: 18 December 2020

Volume 35, pages 401–449, (2021) Cite this article

Hands-on Time:
Notebook 1