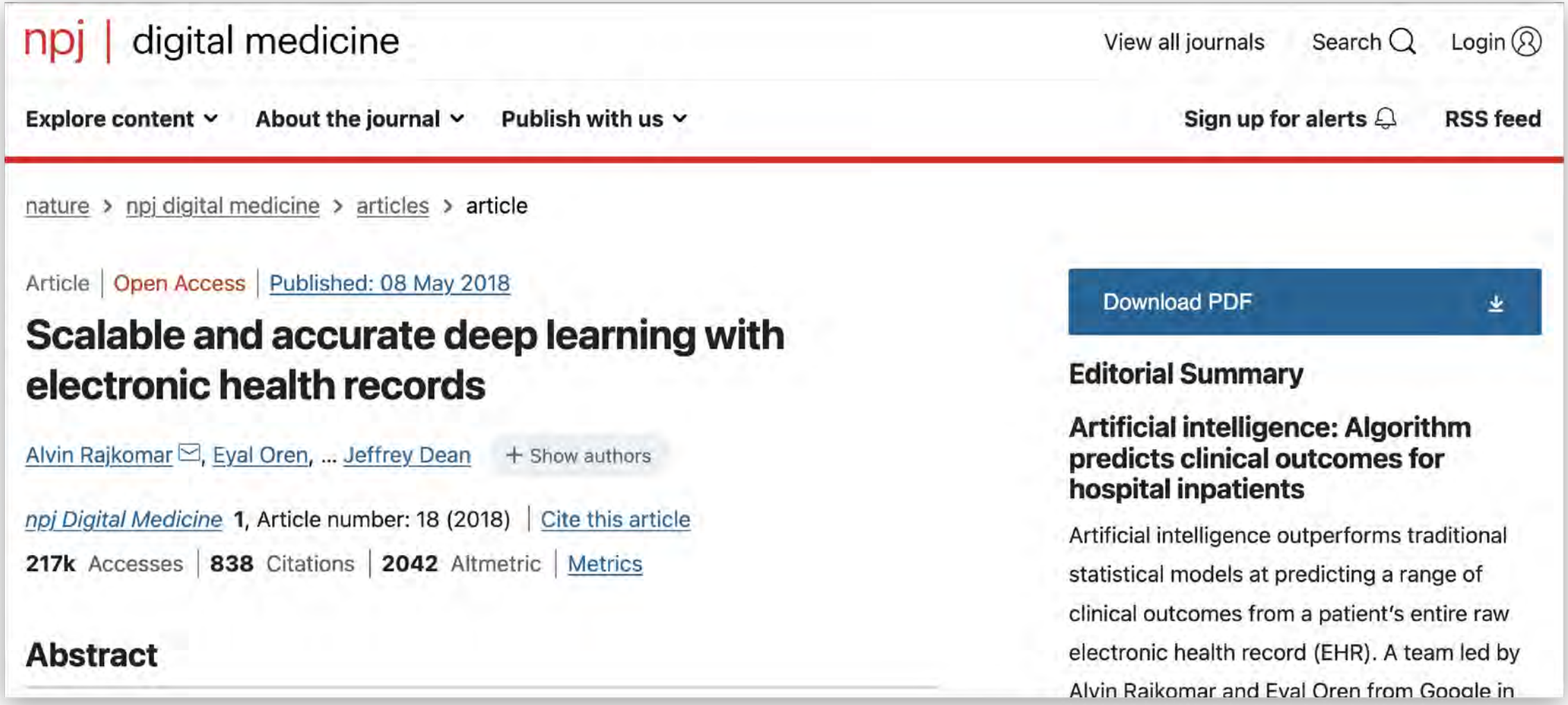
# BST-209: Collaborative Data Science in Healthcare

## Summer Program in Clinical Effectiveness

31 July 2023


# Become familiar with machine learning concepts



npj | digital medicine

View all journals Search Login

Explore content About the journal Publish with us Sign up for alerts RSS feed

nature > npj digital medicine > articles > article

Article | Open Access | Published: 08 May 2018

## Scalable and accurate deep learning with electronic health records

Alvin Rajkomar, Eyal Oren, ... Jeffrey Dean + Show authors

*npj Digital Medicine* **1**, Article number: 18 (2018) | Cite this article

**217k** Accesses | **838** Citations | **2042** Altmetric | Metrics

### Abstract

Download PDF

### Editorial Summary

**Artificial intelligence: Algorithm predicts clinical outcomes for hospital inpatients**

Artificial intelligence outperforms traditional statistical models at predicting a range of clinical outcomes from a patient's entire raw electronic health record (EHR). A team led by Alvin Rajkomar and Eyal Oren from Google in

# Get hands-on coding experience

```
---
title: "Predicting outcome of patients in the ICU"
output: pdf_document
date: "1/1/2023"
---

```{r load_data, include=FALSE}
sql_query <- "SELECT i.subject_id, i.hadm_id, i.los
              FROM `physionet-data.mimiciii_demo.icustays` i;"
data <- run_query(sql_query)
head(data)
```

This document shows how RMarkdown can be used to create a reproducible analysis using
MIMIC-III (version 1.4). Let's calculate the median length of stay in the ICU and
then include this value in our document.

```{r calculate_mean_los, include=FALSE}
avg_los <- median(data$los, na.rm=TRUE)
rounded_avg_los <- round(avg_los, digits = 2)
```

So the median length of stay in the ICU is `r avg_los` days. Rounded to two decimal
places, this is `r rounded_avg_los` days. We can plot the distribution of length of
```

# Build collaborations

Photo: 49th Annual Conference of the Society of Critical Care Medicine (SCCM 2020)

# Schedule

# Overview

- First two weeks, focus on methods for learning from data
    - Responsible machine learning
    - Data wrangling
    - Building and evaluating models
    - Communication of results
- Final week of the course:
    - Explore bias in oxygen saturation measurements.
    - Plan and present a team project.

# Week 1

| | Workshop (1 - 2pm) | Talk (2 - 2.30pm) |
|---|---|---|
| Mon 31 Jul | Course introduction | Leo Celi (MIT) |
| Tue 1 Aug | Responsible ML | Vinith Suriyakumar (MIT) |
| Wed 2 Aug | Responsible ML | Ahmed Abdelfattah (Harvard) |
| Thu 3 Aug | Introduction to ML | Tristan Naumann (Microsoft) |
| Fri 4 Aug | Introduction to ML | Danielle Bitterman (Harvard) |

# Week 2

| | Workshop (1 - 2pm) | Talk (2 - 2.30pm) |
|---|---|---|
| Mon 7 Aug | Introduction to ML | Fábio Duarte (MIT) |
| Tue 8 Aug | Tree models | Vesela Kovacheva (Harvard) |
| Wed 9 Aug | Tree models | Weiwei Pan (Harvard) |
| Thu 10 Aug | Process mining | Suzy McKinney (Harvard) |
| Fri 11 Aug | Generative AI | Eugenio Zuccarelli (CVS) |

# Week 3

| | Workshop (1 - 2.30pm) |
|---|---|
| Mon 14 Aug | Project (pulse oximetry) |
| Tue 15 Aug | Project (pulse oximetry) |
| Wed 16 Aug | Project (pulse oximetry) |
| Thu 17 Aug | Prepare group presentations |
| Fri 18 Aug | Group presentations |

# Presentation (Fri 18 Aug)

- Propose a project
- 6 minute talk (group, slides)
  - Introduction
  - Goals
  - Data
  - Methods

# Groups

**Group 1**

João Matos

Yusuke Takeda

Naira Link

Hui Miao

**Group 2**

Renata Proa

Chrystinne Fernandes

David Gritsch

Crystal McLellan

**Group 3**

Niklas Adams

Fredrik Willumsen Haug

Pui Ning Pauline Yeung

Asimina Lazaridou

**Group 4**

Lasse Hansen

Zara Sheikh

Hiten Naik

Kieun Seok

**Group 5**

Nikolaj Munch

Sarah Loh

Anvesh Narimiti

Heena Manglani

Sung Hae Chang

**Group 6**

Tristan Struja

Khushboo Teotia

Rachel Rosen

Ana Cecilia Farfan Ruiz

Hiroki Mizuno

Lisa Gudenkauf

**Group 7**

David Restrepo

Ohad Oren

Ardel Romero Pabon

Kevin An

**Group 8**

Eptehal Nashnoush

Krishnaveni Parvataneni

Christopher Callahan

Margaret Ong

Kimberly Mills

**Group 9**

Kevin Ma

Hugh Kim

Yung Lee

Christopher Dall

**Group 10**

Po-Chih Kuo

Chao-Ju (Luna) Chen

James Stone

Tina Shiang

Rodrigo Rosa Gameiro

Sotonye Imadojemu

**Group 11**

Jack Gallifant

Adrien Carrel

Zihan Quian

Hong Xiong

Heng Cai

# Evaluation

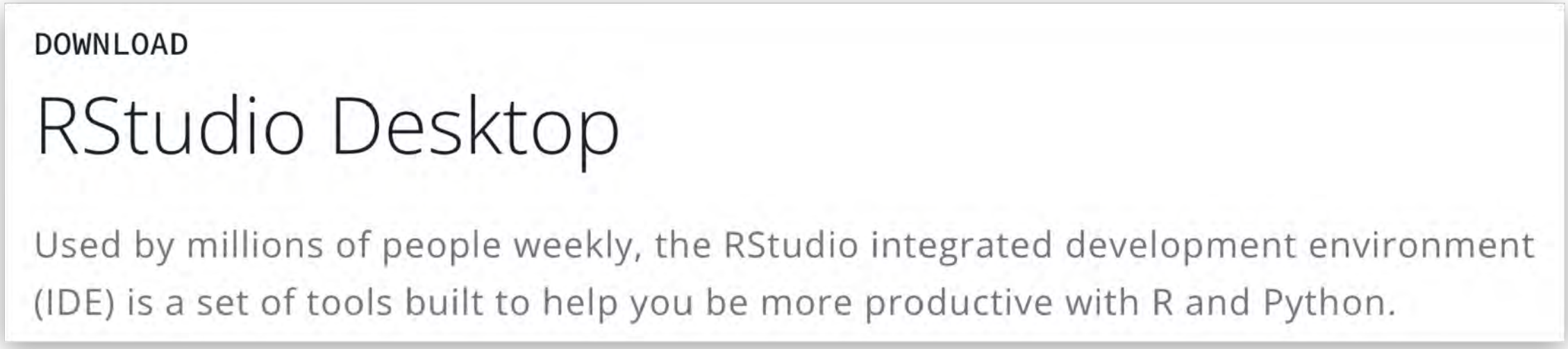
# Final grades based on:

- Attendance and participation: 60%
- Group presentation 40%

# Any questions?

# Getting set up

# RStudio




https://posit.co/download/rstudio-desktop/

# Course materials

https://github.com/mit-lcp/bst209