MACHINE LEARNING FOR COMPREHENSIVE FORECASTING OF ALZHEIMER'S DISEASE PROGRESSION

PAPER PRESENTATION

OCTOBER 22, 2021

WEIMING REN, TIANSHU ZHU

UNIVERSITY OF TORONTO

AGENDA

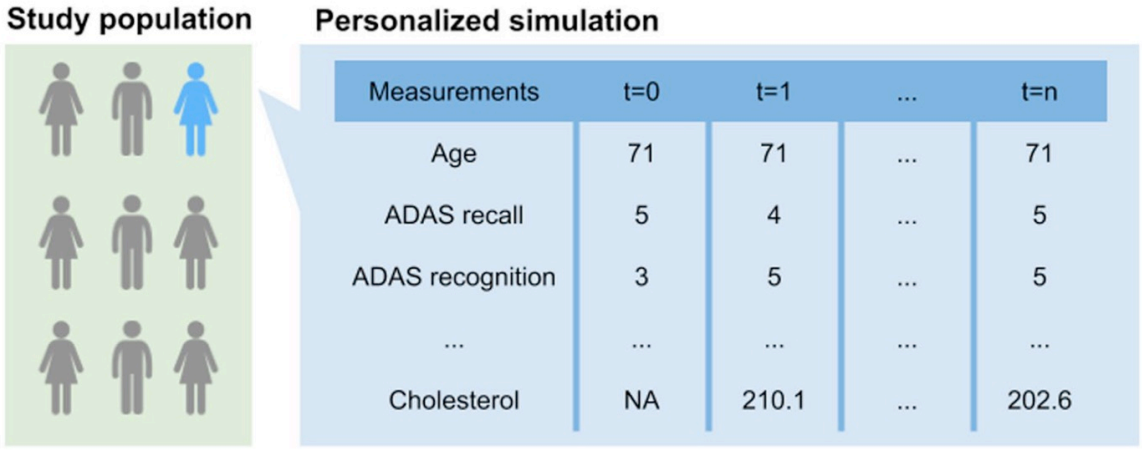
- Background & Motivation
- Dataset
- The Model: Conditional Restricted Boltzmann Machine
- Experiments & Evaluations
- Discussion

BACKGROUND & MOTIVATION

- Alzheimer's Disease (AD) and Mild Cognitive Impairment (MCI) are complex neurodegenerative diseases with multiple cognitive and behavioral symptoms.
  - Difficult to diagnose, manage and treat
- Existing supervised learning methods: predicting a single endpoint
  - E.g. change in ADAS-Cog score from baseline
- This paper proposed an unsupervised model to simulate the progression of entire patient profiles
  - Conditional Restricted Boltzmann Machine (CRBM)
- The main goals of the model:
  - Generate completely synthetic patient profiles
  - Predict patient trajectories based on real baseline data

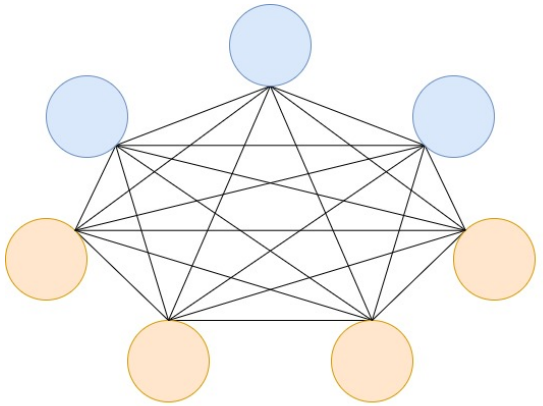
DATASET

- Coalition Against Major Diseases (CAMD) Online Data Repository for AD
- **18-month** longitudinal trajectories of **1909 patients** with **MCI or AD**
- **44 *multimodal* variables**, divided into **3-month intervals** (7 time points)
  - Alzheimer's Disease Assessment Scale (ADAS)
  - Mini Mental State Exam (MMSE)
  - Laboratory tests
  - Clinical characteristics
  - Background information



THE MODEL: RESTRICTED BOLTZMANN MACHINE

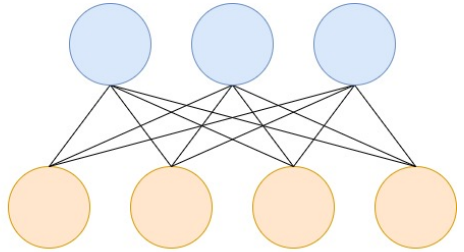
- **Boltzmann machine**: fully connected **Markov Random Field**
  - **Undirected**, connected graph
  - Each node represents a **random variable**
  - Two types of nodes: **visible** and **hidden** nodes
  - The joint probability distribution fulfills the **(local) Markov property** w.r.t. the graph



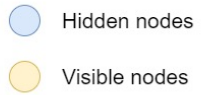
Boltzmann Machine

- **Restricted** Boltzmann machine:
  - Only contains edges between visible and hidden nodes
  - Models the joint probability distribution of the visible and hidden nodes

$$p(\mathbf{v}, \mathbf{h}) = Z^{-1} \exp\left(\underbrace{\sum_j a_j(v_j)}_{\text{Visible node bias}} + \underbrace{\sum_\mu b_\mu(h_\mu)}_{\text{Hidden node bias}} + \underbrace{\sum_{j\mu} W_{j\mu} \frac{v_j}{\sigma_j^2} \frac{h_\mu}{\epsilon_\mu^2}}_{\text{Connection between visible and hidden layer}} \right)$$



Restricted Boltzmann Machine



- Generative model: generate new samples from the joint distribution

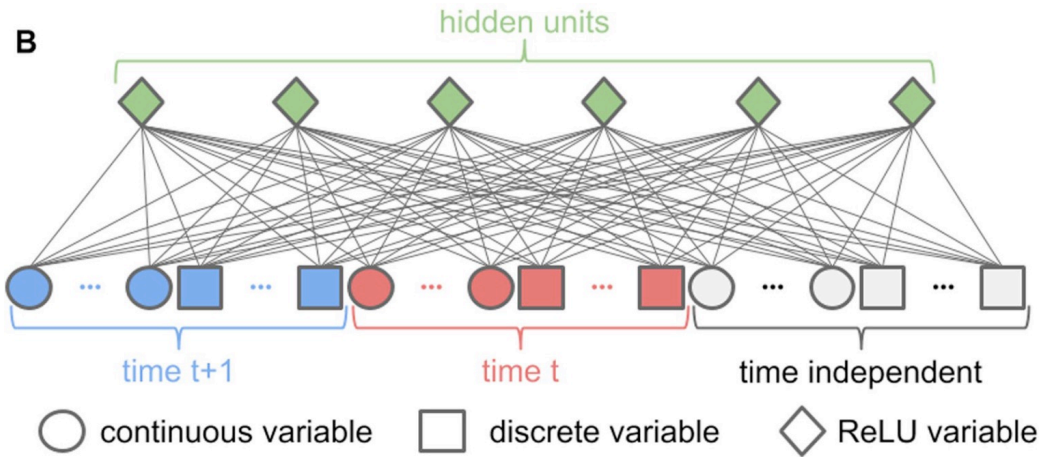
MODEL ARCHITECTURE DETAILS

- Conditional RBM: modeling interdependence between time points

- Training input: two types of covariates
  - Static covariates $\boldsymbol{x}_i^{static}(t=0)$
  - Dynamic covariates $\boldsymbol{x}_i^{dynamic}(t)$

- $\boldsymbol{v}_i(t) = \{\boldsymbol{x}_i^{dynamic}(t+1), \boldsymbol{x}_i^{dynamic}(t), \boldsymbol{x}_i^{static}(t=0)\}$

- 50 hidden nodes with ReLU activations



TRAINING CRBM

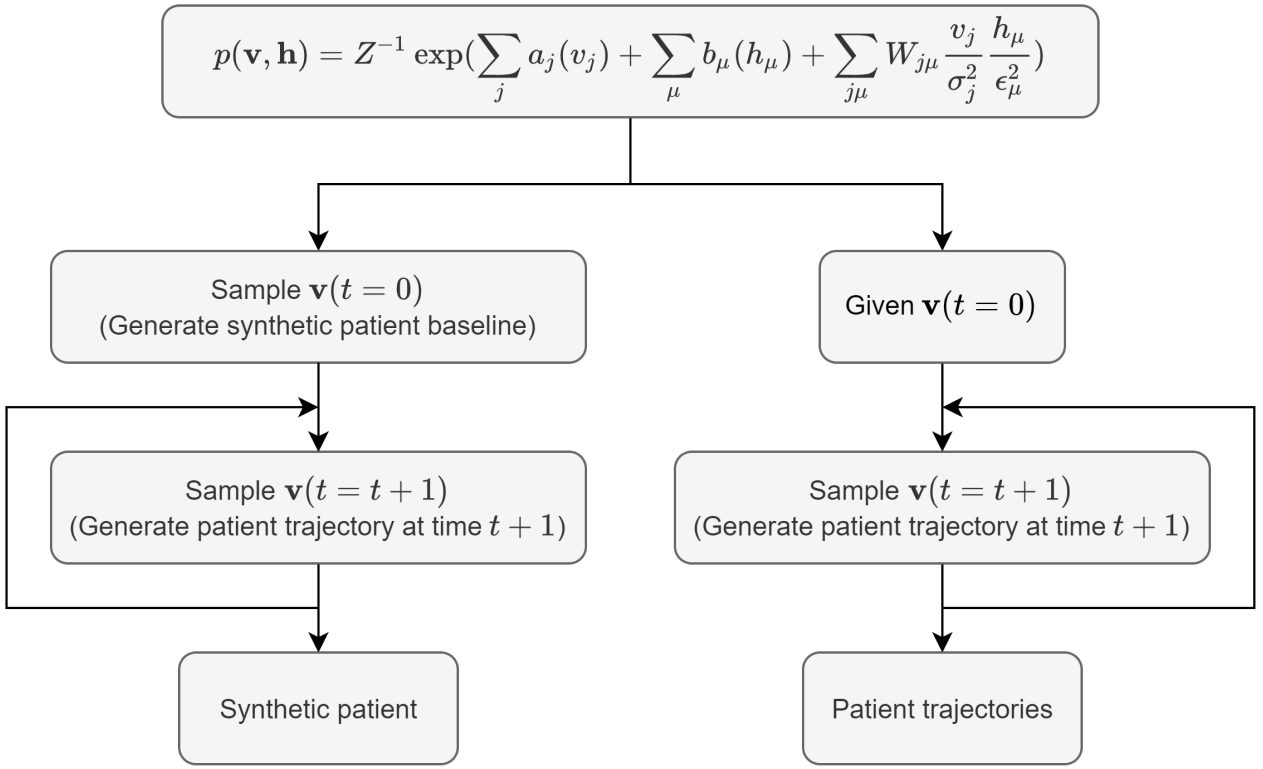
- Unsupervised training: Maximum likelihood estimation + Adversarial training

- Maximum likelihood estimation
  - Maximize the **log likelihood** of the data under the model distribution
  - equivalent to minimizing the **Kullback–Leibler (KL) divergence** between the unknown data distribution and the distribution parameterized by the model

- Adversarial training
  - Measures how easy it is to distinguish patient profiles generated from the **statistical model** from **real patient profiles**.
  - A random forest classifier is used to distinguish real and synthetic patient profiles

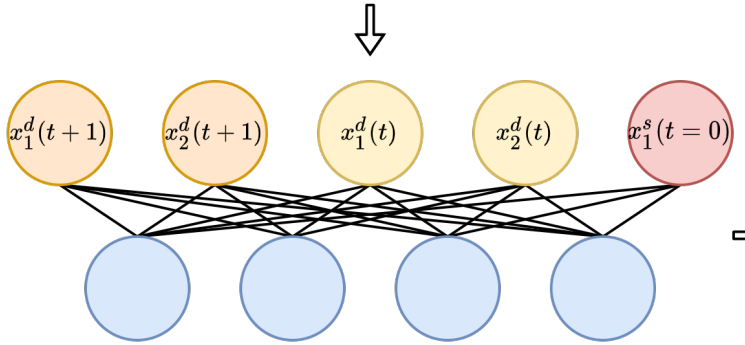
GENERATING SYNTHETIC PATIENT & PATIENT TRAJECTORIES

- Directly sampling from the joint distribution is computationally infeasible.
- Sampling is performed using Markov Chain Monte Carlo (MCMC) methods.



OVERALL TRAINING & INFERENCING

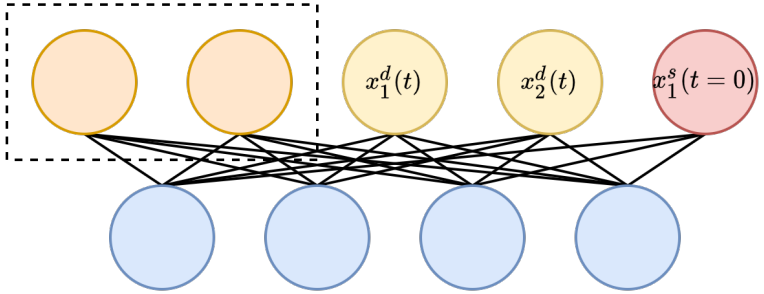
$$\mathbf{v}_i(t) = \{\mathbf{x}_i^{dynamic}(t+1), \mathbf{x}_i^{dynamic}(t), \mathbf{x}_i^{static}(t=0)\}$$



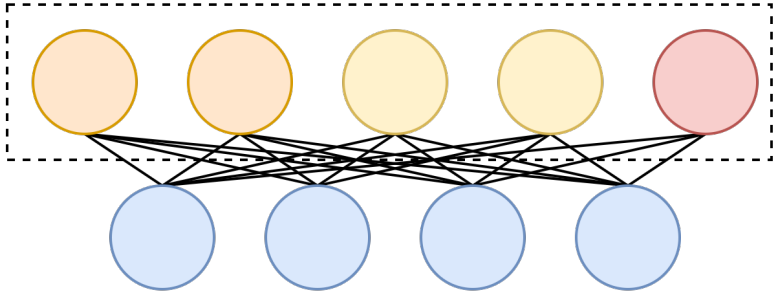
Unsupervised learning (Maximum likelihood + adversarial training)

$$p(\mathbf{v}, \mathbf{h}) = Z^{-1} \exp\left(\sum_j a_j(v_j) + \sum_\mu b_\mu(h_\mu) + \sum_{j\mu} W_{j\mu} \frac{v_j}{\sigma_j^2} \frac{h_\mu}{\epsilon_\mu^2}\right)$$

Sample next time point



Sample synthetic baseline



GOODNESS-OF-FIT OF THE MODEL

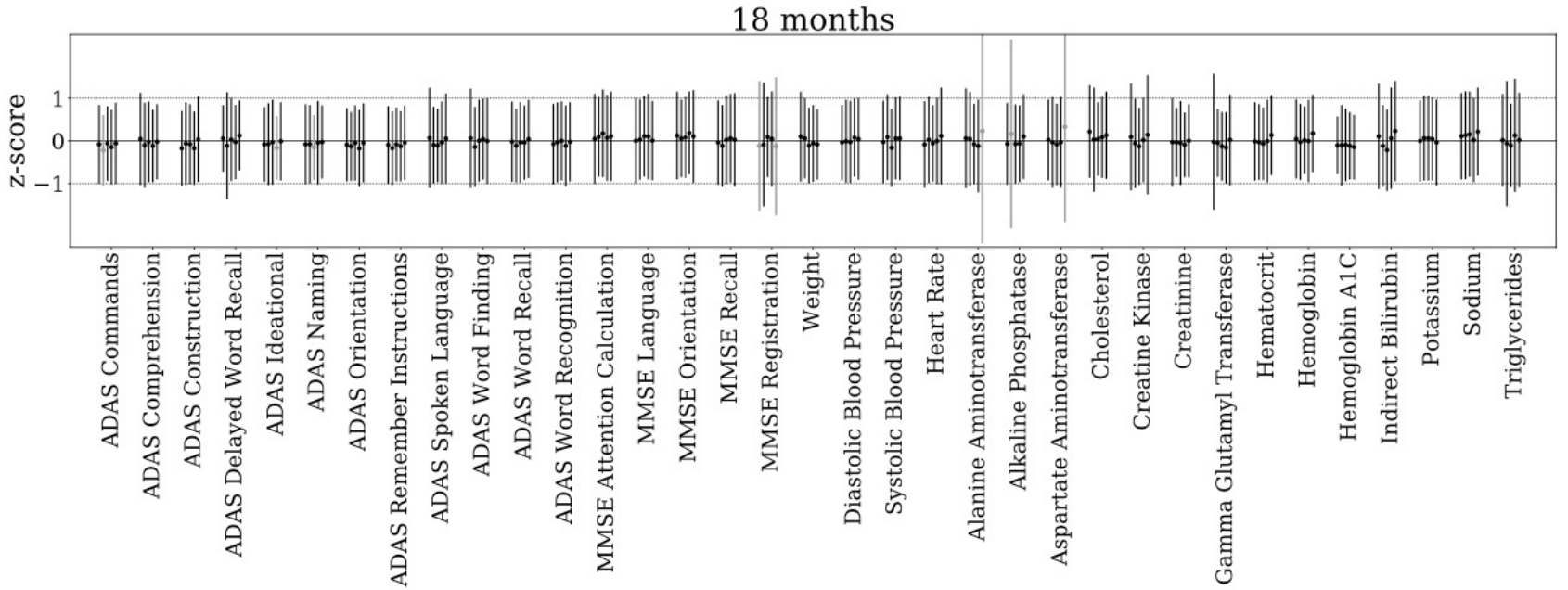
- **Fundamental assumption**: Each patient's time-dependent variable is ***stochastic*** *(sampled from a distribution).*
- Compare the **distribution predicted by CRBM** to the **distribution of real data** using a variety of metrics
  - Means, standard deviations, and p-values
  - Correlations and autocorrelations
  - Differentiability between real and predicted data

ASSESS MEANS AND STANDARD DEVIATIONS

- $\boldsymbol{x_{ij}(t)}$: value from **real** data of variable ***j***, *f*or patient ***i***, at time ***t***.
- $\mathbf{E}[x_{ij}(t)|\mathbf{x}_i(0)]$: **predicted** conditional mean.
- $\mathbf{Var}[x_{ij}(t)|\mathbf{x}_i(0)]$: **predicted** conditional variance.
- $z_{ij}(t) = \frac{x_{ij}(t) - \mathbf{E}[x_{ij}(t)|\mathbf{x}_i(0)]}{\sqrt{\mathbf{Var}[x_{ij}(t)|\mathbf{x}_i(0)]}}$: z-score
- Aggregate across patients to get predicted distributions of z-score.
- If predicted distributions are **consistent** with real data, then z-score will have **0 mean and 1 standard deviation**.

ASSESS MEANS AND STANDARD DEVIATIONS

- **Good-fit** if z-scores of **0 mean and 1 standard deviation**.

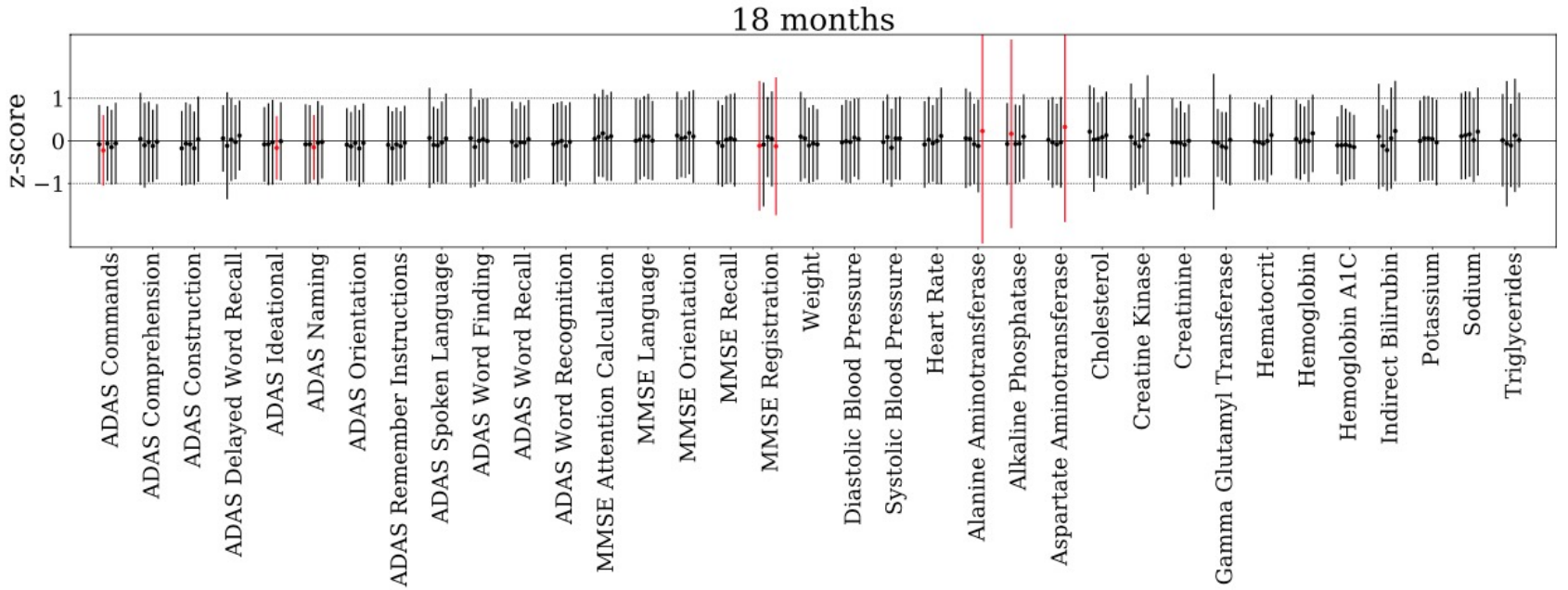


ASSESS P-VALUES

- **Simplifying assumption**: $x_{ij}(t)$ ~ normal distribution. If predicted distributions are consistent with real data, then z-score ~ standard normal distribution.
- **P-value：** assess closeness between predicted distribution of z-score and standard normal distribution.
  - computed using the Kolmogorov-Smirnov test statistic and survive a Bonferroni multiple-testing correction.
- If predicted distributions are **consistent** with real data, then **Non-significant p-value** (significant at $p < 0.05$).

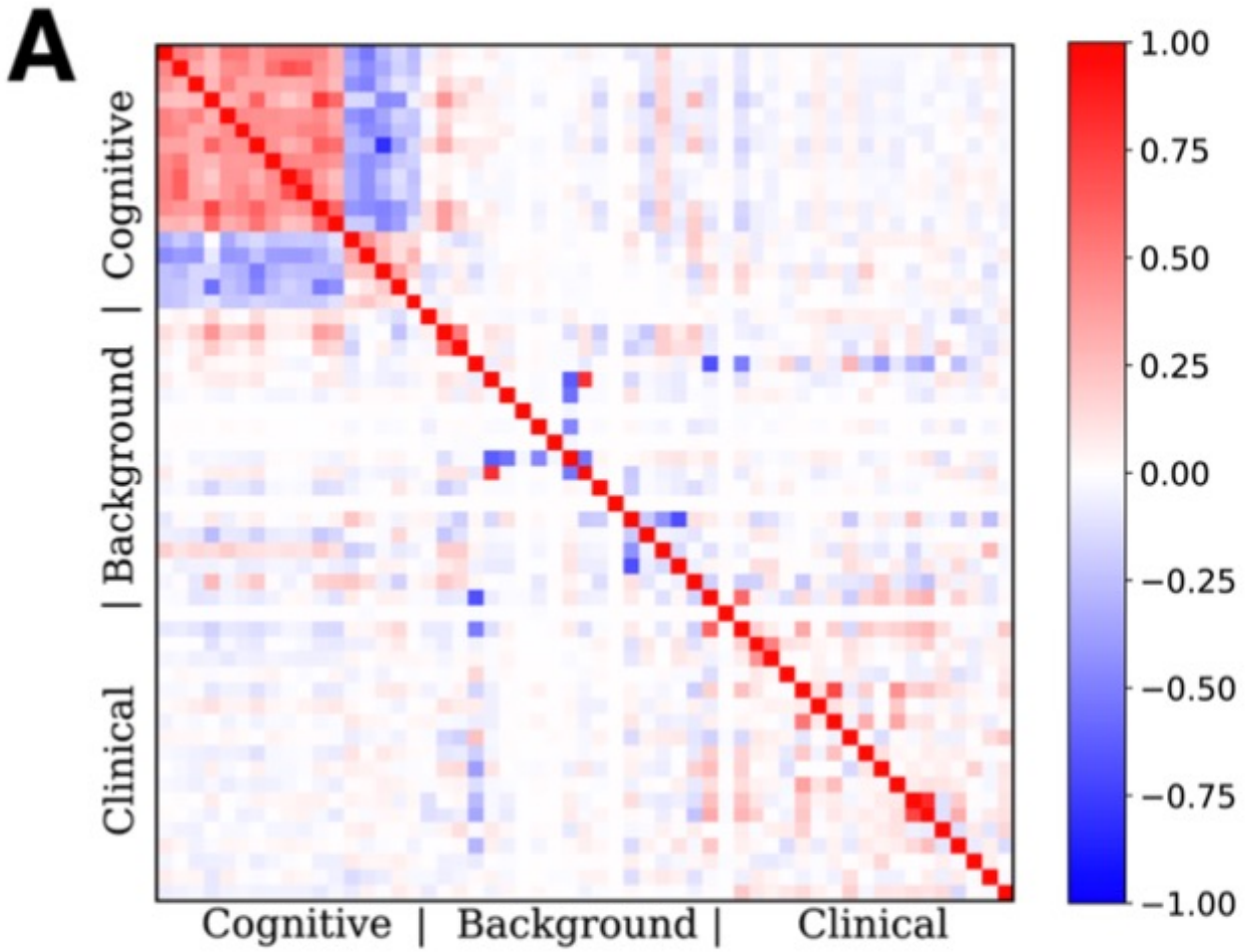
ASSESS P-VALUES

- **Good-fit** if z-scores of **0 mean** and **1 standard deviation**.
- **Good-fit** if non-significant p-value (**not marked red**).



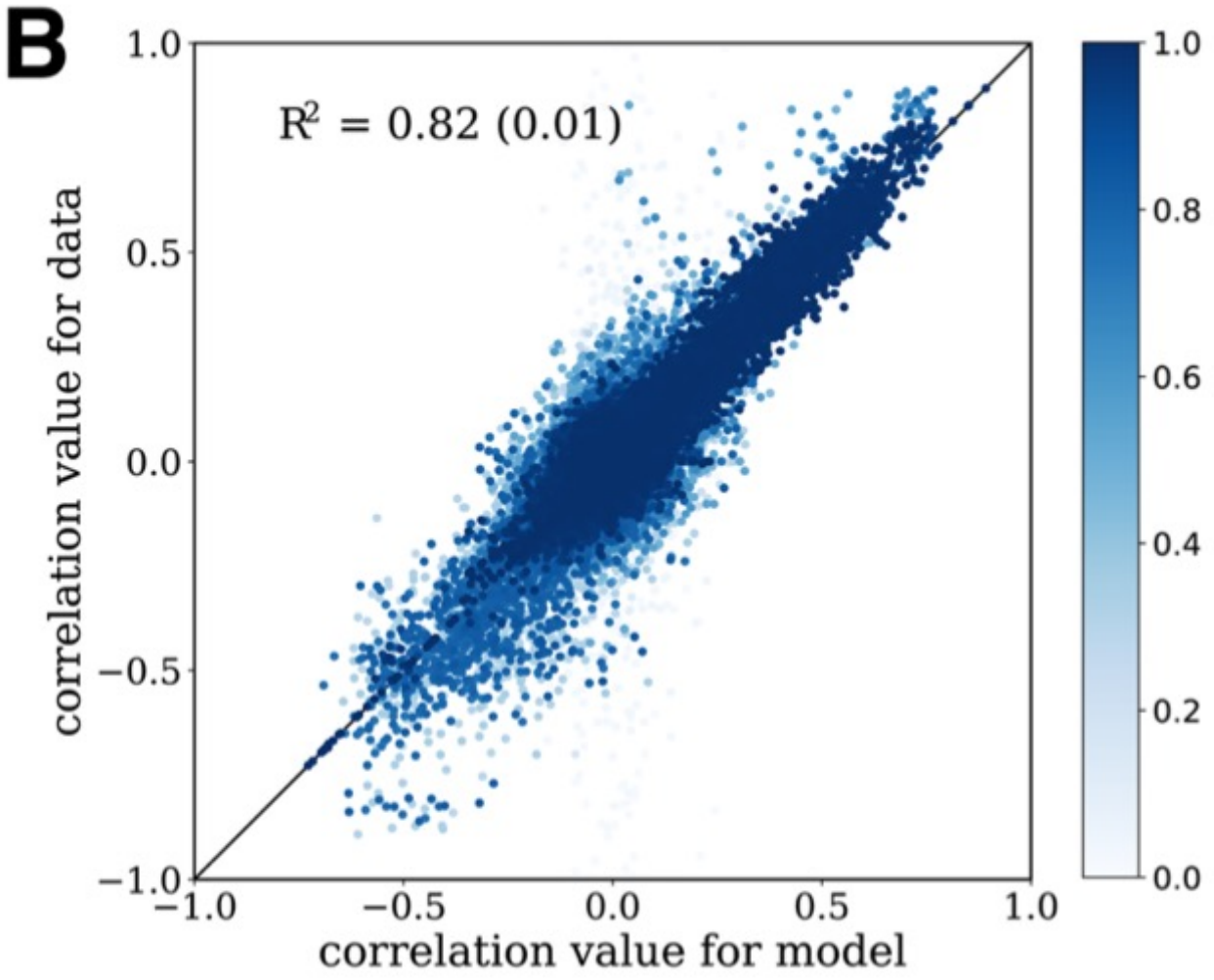
ASSESS CORRELATIONS

- Correlations between pairs of predicted variables (below diagonal) and pairs of real variables (above diagonal) .
- **Good-fit** if **symmetry** between below diagonal and above diagonal.



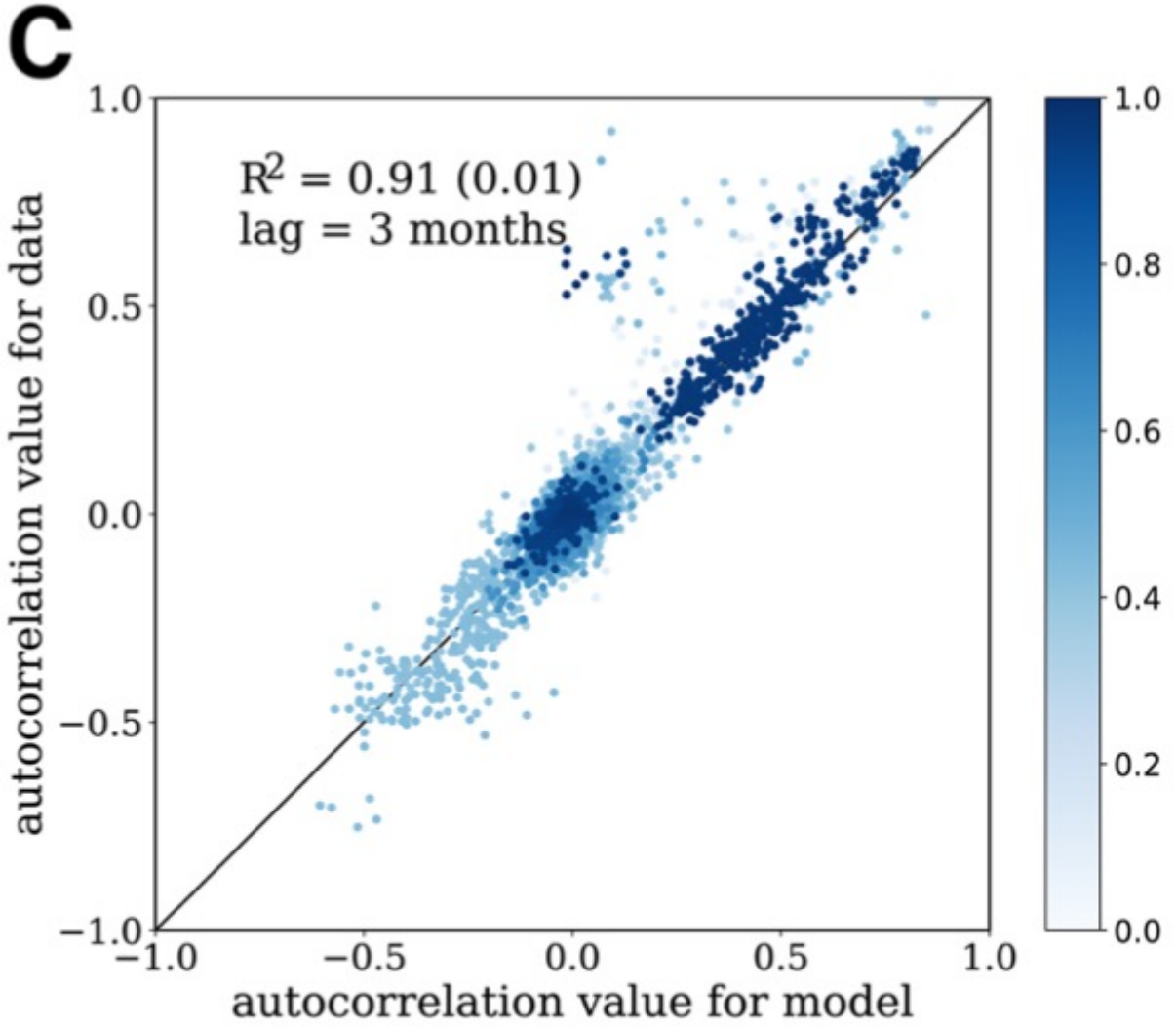
ASSESS CORRELATIONS

- Correlations between pairs of predicted variables (x-axis) and pairs of real variables (y-axis).
  - lighter colors: more missing data.
- **Good-fit** if points are **close to diagonal**.



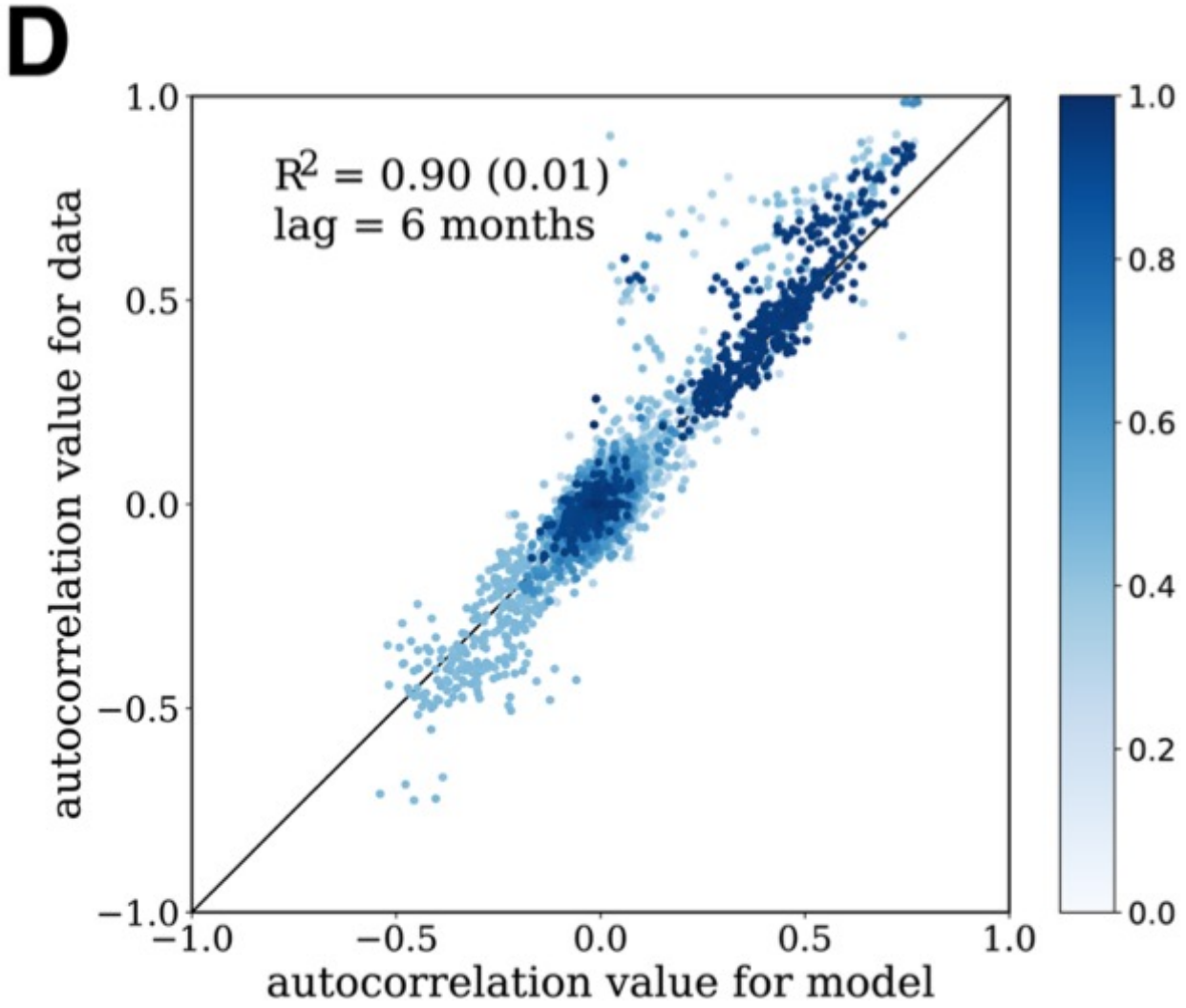
ASSESS AUTOCORRELATIONS

- Autocorrelations between same predicted variables (x-axis) and real variables (y-axis) for 3 month time lag.
  - lighter colors: more missing data.
- **Good-fit** if points are **close to diagonal**.



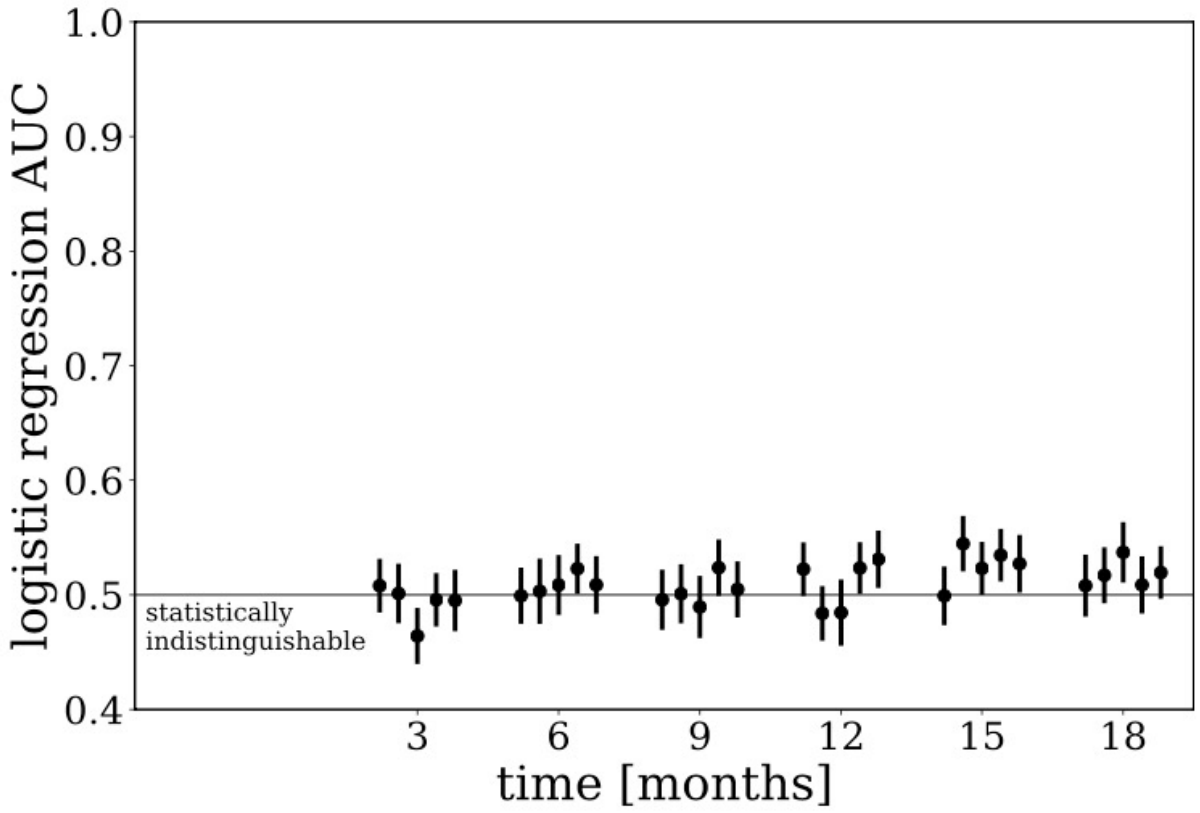
ASSESS AUTOCORRELATIONS

- Autocorrelations between same predicted variables (x-axis) and real variables (y-axis) for 6 month time lag.
  - lighter colors: more missing data.
- **Good-fit** if points are **close to diagonal**.



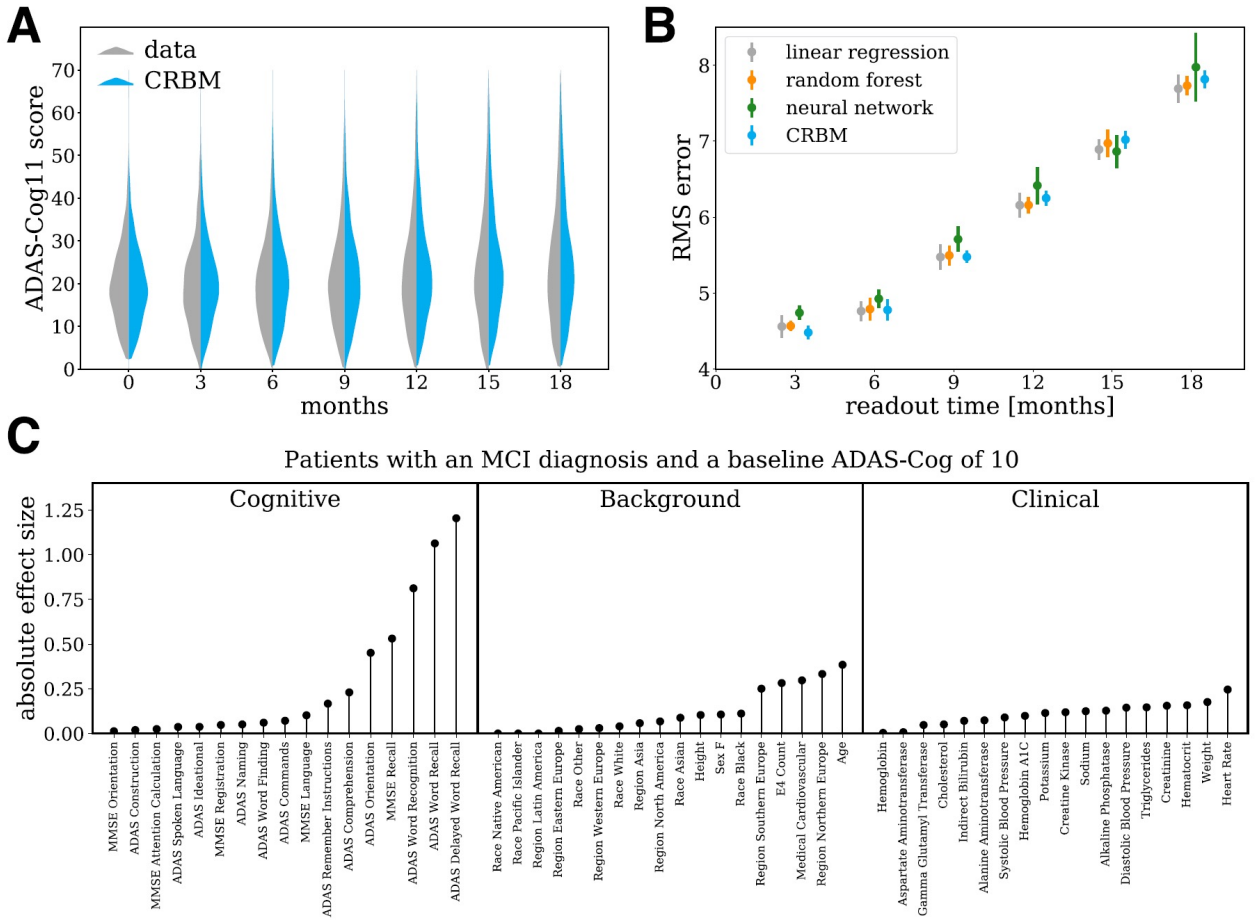
ASSESS DIFFERENTIABILITY - PREDICTED & REAL DATA

- At each time point, trained logistic regression model to differentiate predicted and real data.
- Estimated AUC metric. AUC close to 0.5 if model cannot reliably differentiate.
- **Good-fit** if AUC close to **0.5**.



FORECASTING AND INTERPRETING DISEASE PROGRESSION

- Convert patient profile trajectories back to ADAS-Cog11 scores
- ADAS-Cog11 score distribution for data and model prediction are very similar
- CRBM shows similar performance to other supervised learning methods
- Patients with poor performance on the recall and word recognition test tend to progress more rapidly



DISCUSSION

- Simulation of stochastic disease progression of individual patients enables personalized data-driven medicine
- CRBM directly integrates multimodal data with both continuous and discrete variables, time series and static variables within a single model
- This unsupervised disease progression model can be easily extended to other diseases

- Limitation: dataset size is relatively small
  - Only 44 variables are included in the model
  - Each time series variable only contains 7 time points
  - The small dataset thereby limits the choice of the model used in the project

THANK YOU FOR YOUR ATTENTION!

UNIVERSITY OF
TORONTO