Early Prediction of Labour Force Dropout in SLE Patients

Integrating Longitudinal Deep Learning through an LSTM RNN with Random Forests

Javier Mencia Ledo[1], Dennisse Bonilla[2], Zahi Touma, MD, PhD[2], Behdin Nowrouzi-Kia, PhD [1]

[1]ReSTORE Lab
University of Toronto

[2]Toronto Western Lupus Clinic
University Health Network

SUDS Showcase, 2025

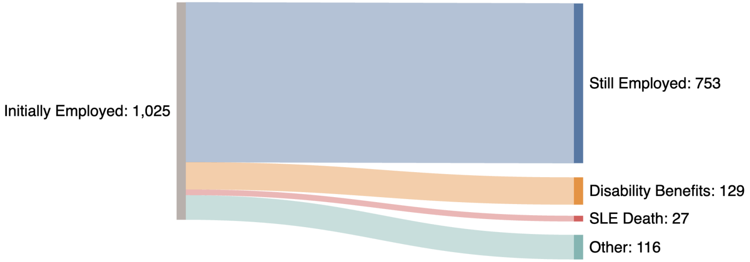
Background & Objectives

- SLE is a chronic autoimmune disease that affects functioning
- **Problem:** 20-40% of SLE patients exit workforce prematurely
- **Impact:**
  - Work gives financial stability and sense of purpose
    $\rightarrow$ Social isolation and mental health decline
  - Socioeconomic burden
- **Solution:**
  - Hybrid RF-LSTM tool to flag at-risk patients *before* departure to allow early clinical intervention

Employment Trajectories

Overview:

- Records at the Toronto Lupus Clinic (UHN) between 1996-2024



Source: Robroek et al., 2015 [1]

- Still employed includes actively working or retired after the retirement age
- Other states: Unemployed, Economically Inactive, Early Retirement

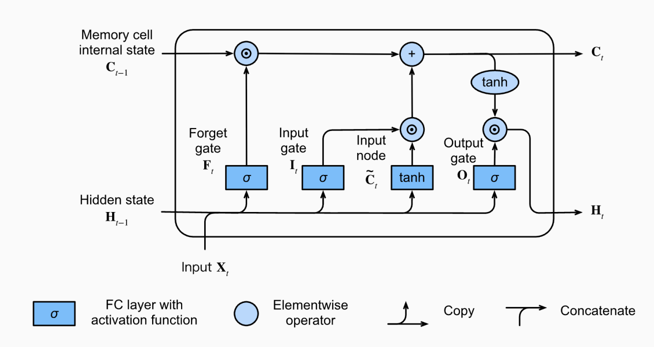
Long-Short Term Memory Model Architecture

Why LSTMs?

- Capture **long-term dependencies** across years of patient visits

Input:
25 clinical features per visit selected w/RF and using class balancing via synthetic minority augmentation: $\mathbf{X} \in \mathbb{R}^{N \times T \times 25}$

Model cell at visit t ($1 \leq t \leq T$):



Source: Hochreiter et al. [2], Zhang et al. [3]

Prediction Pipeline

Hidden state at visit t:

$$\mathbf{H}_t = \text{LSTM}(\mathbf{X}_t, \mathbf{H}_{t-1}, \mathbf{C}_{t-1}) \tag{1}$$

Linear transformation:

$$\mathbf{z_{t,k}} = \mathbf{W}_k \mathbf{H}_t + \mathbf{b}_k \tag{2}$$

1. Class prediction (t = T):

$$p(y = k) = \frac{\exp(z_{t,k})}{\sum_{j=1}^{K} \exp(z_{t,j})} \quad (K = 3 \text{ classes}) \tag{3}$$

$$\hat{y} = \underset{k \in \{\text{employed, disability, mortality}\}}{\arg\max} p(y = k) \tag{4}$$

2. Risk score computation for disability and mortality resp. $(t \leq T)$:

$$\text{risk}_{t,k} = \sigma(\mathbf{z}_{t,k}) = \frac{1}{1 + \exp(-\mathbf{z}_{t,k})} \tag{5}$$

Alert if $\text{risk}_{t,k} \geq \theta_k$ where θ_k is a hyperparameter that optimizes early detection time(Δt)

Model Performance Highlights

Final Employment State Classification
After 5-fold CV:

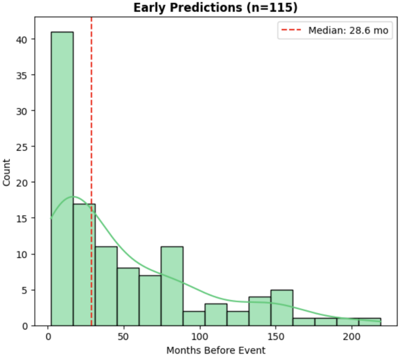
Overall balanced accuracy: **91%**

Early Warning System for Disability

Early warning system achieves **89% case detection**
It has a median lead time of **28.6 Months** (IQR: 14.3–42.1)

Early Intervention Strategies:

- Proactive rehabilitation programs
- Multidisciplinary specialist coordination
- Workplace accommodation planning



Acknowledgements

ReSTORE

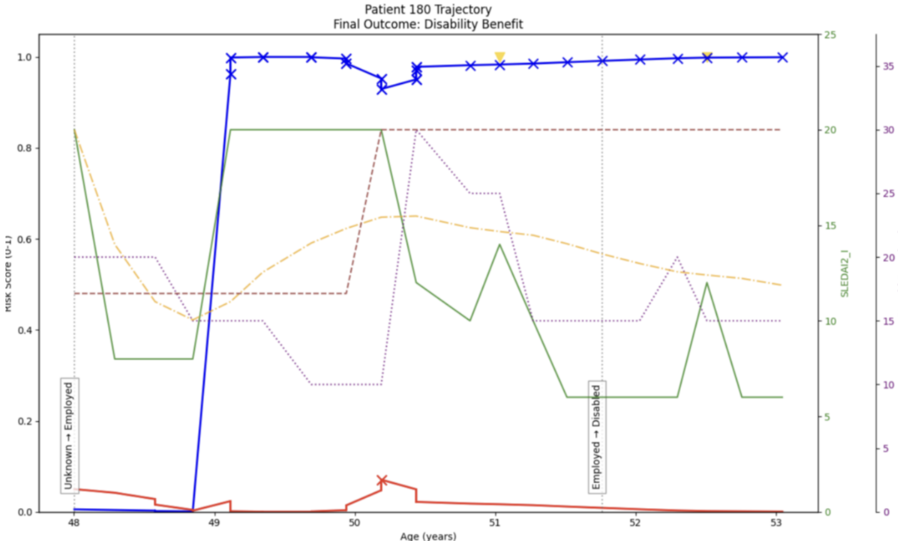
LUPUS ONTARIO
Life Without Lupus

DATA SCIENCES INSTITUTE

UNIVERSITY OF TORONTO

This work was supported by the Data Sciences Institute, University of Toronto

Individual Patient Trajectory



Robroek, S. J., Rongen, A., Arts, C. H., Otten, F. W., Burdorf, A., & Schuring, M. (2015). Educational inequalities in exit from paid employment among Dutch workers: The influence of health, lifestyle and work. *PloS one*, *10*(8), e0134867. https://doi.org/10.1371/journal.pone.0134867

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2021). *Dive into Deep Learning*. Cambridge University Press. https://d2l.ai/chapter_recurrent-modern/lstm.html.