MATCH-Net: Dynamic Prediction in Survival Analysis using Convolutional Neural Networks

DANIEL JARRETT*, JINSUNG YOON†, & MIHAELA VAN DER SCHAAR*,†,‡

*UNIVERSITY OF OXFORD, †UCLA, ‡ALAN TURING INSTITUTE

This work was supported by the Office of Naval Research and the National Science Foundation

INTRODUCTION

Problem: Survival Analysis

- Accurate prediction of **Disease Trajectories** is critical for the **Early Identification** and **Timely Treatment** of patients at risk.

Current Methods

- Statistical methods like **Cox Landmarking** and **Joint Modeling** are often limited by parametric assumptions and computationally constrained.
- Recent **Deep Learning** approaches improve on these limitations, but do not capture potential information in longitudinal covariate histories.

Main Ideas

- Issue **Dynamically Updated** survival predictions via longitudinal sliding-window mechanism.
- Use **Temporal Convolutions** to capture explicit representations of temporal dependencies.
- Accommodate potentially informative patterns of **Missingness** with dual-stream structure.

PROBLEM FORMULATION

Notation

- **Covariate Vector** $\mathbf{x}_{i,t}$ for **Patient** $i \in \{1, ..., N\}$ at **Time** t, where time has discrete resolution δ.
- **Survival Datum** $(t, \mathbf{x}_{i,t}, s_{i,t})$, where $s_{i,t}$ is the binary **Survival Indicator** for event of interest.
- **Time-to-Event** $T_i = \min\{T_{i,\text{surv}}, T_{i,\text{cens}}\}$, where $T_{i,\text{surv}}$ is the random variable for time of **Event Occurrence** and $T_{i,\text{cens}}$ for **Right-Censoring**.

Dynamic Prediction

- **Historical Window** of observations in $(t-w, t]$, where w indicates the width of lookback:

$$\mathbf{X}_{i,t,w} = \langle (t', \mathbf{x}_{i,t'}, s_{i,t'}) \rangle_{t' \in T}$$
$$\text{where } T = \{t' : t - w \leq t' \leq t\}$$

- **Failure Prediction** for forward interval $(t, t+\tau]$, where τ indicates the prediction horizon:

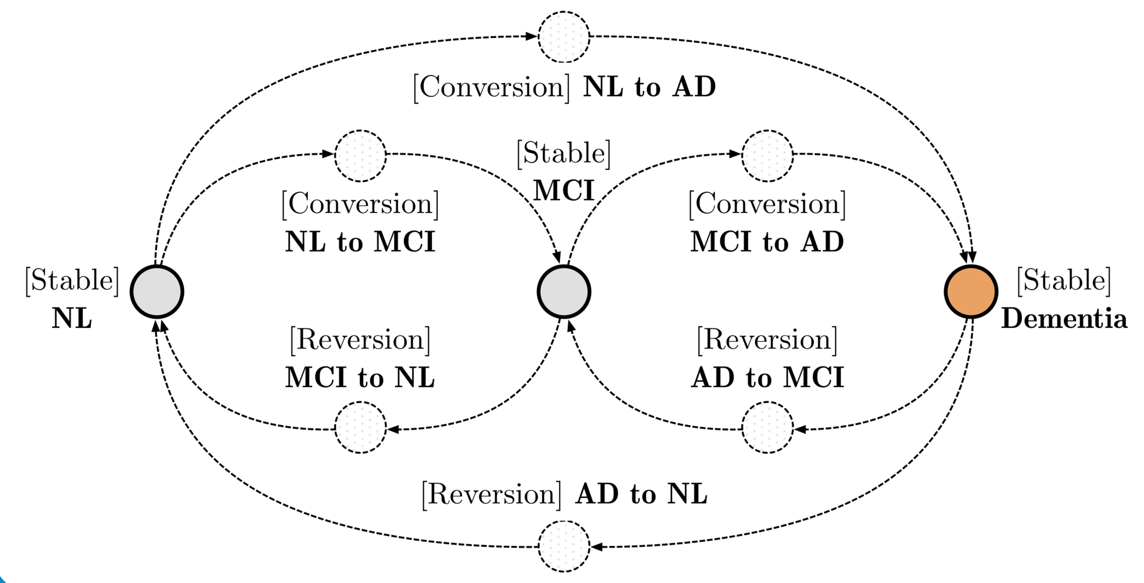
$$F_i(t+\tau|t, w) = \mathbb{P}(T_{i,\text{surv}} \leq t+\tau | T_{i,\text{surv}} > t, \mathbf{X}_{i,t,w})$$

RELATED WORK

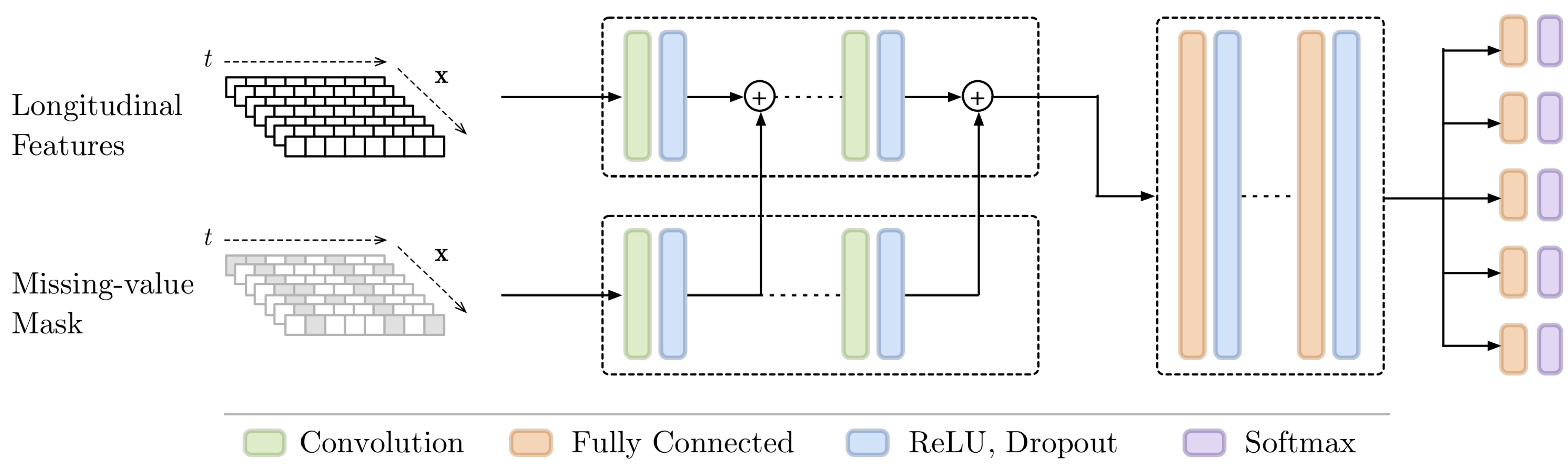
	Non-Linearity	Deep Learning	Direct-to-Probability	Time-Variance	Dynamic Prediction
[1]	✗	N/A	N/A	✗	✗
[2]	✓	✗	✗	✗	✗
[3]	✓	✓	✗	✗	✗
[4]	✓	✓	✓	✗	✗
[5]	✓	✓	✓	✓	✗
[6]	✓	✓	✓	✓	✓

[1] Cox, [2] Faraggi and Simon, [3] Katzman *et al.*, [4] Luck *et al.*, [5] Lee *et al.*, [6] MATCH-Net

ALZHEIMER'S STATE SPACE



MATCH-NET ARCHITECTURE



Input longitudinal features $\mathbf{X}_{i,t,w}$ and missing-value mask $\mathbf{Z}_{i,t,w}$
$\rightarrow$ **Output** failure predictions $\hat{\mathbf{y}}_{i,t} = [\hat{F}_i(t+\delta|t,w), ..., \hat{F}_i(t+\tau_{\max}|t,w)]$

TRAINING MATCH-NET

Minimization of the dissimilarity between empirical and model distributions of survival times:

$$D_{\text{KL}}(\hat{F}_i(t+\tau|t,w)||s_{i,t+\tau}) = -\mathbb{E}_{\mathbf{X}_{i,t,w} \in \mathbf{X}}[\log \hat{F}_i(t+\tau|t,w) - \log s_{i,t+\tau}]$$

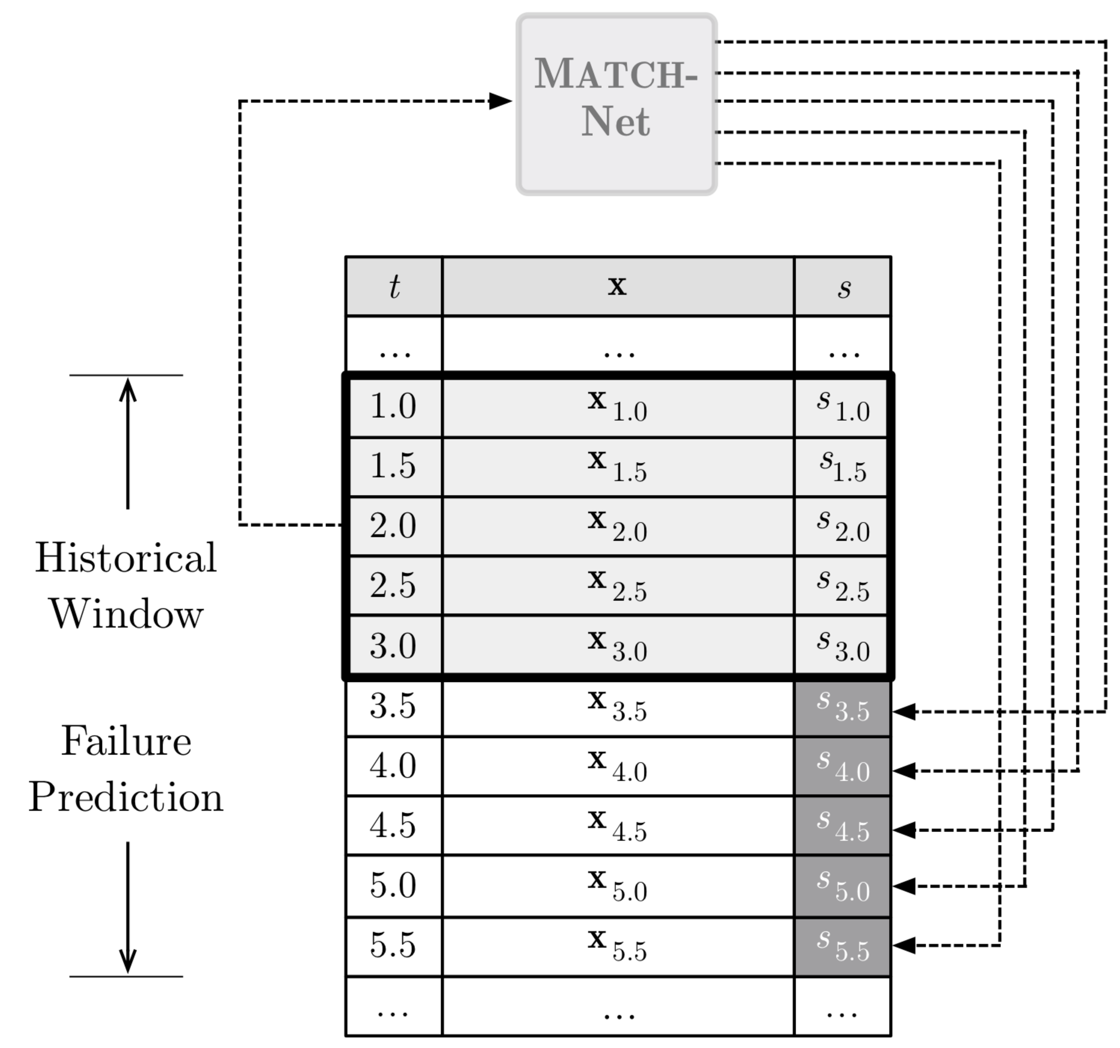
Log-Likelihood of a single empirical result $s_{i,t+\tau}$ and model estimate $\hat{F}_i(t+\tau|t,w)$ of failure:

$$\mathcal{L}_{i,t,\tau}(\boldsymbol{\theta}) = -[s_{i,t+\tau} \log \hat{F}_i(t+\tau|t,w) + (1-s_{i,t+\tau}) \log(1-\hat{F}_i(t+\tau|t,w))]$$

Total Loss Function for all prediction horizons τ, all times t along each trajectory, and all patients:

$$\mathcal{L}(\boldsymbol{\theta}) \propto \sum_i \sum_j \sum_k \mathcal{L}_{i,j\delta,k\delta}$$

SLIDING WINDOWS



DISCRIMINATIVE PERFORMANCE

	τ	MATCH-Net	S-TCN	S-MLP	FCN	D-Atlas	RNN	MLP	JM	LM
AUROC	0.5	**0.962**	0.961	0.959	0.954	0.959	0.949*	0.948*	0.913*	0.909*
	1.0	**0.942**	0.941	0.932	0.930	0.929	0.930	0.930	0.917*	0.914*
	1.5	**0.902**	**0.902**	0.897	0.895	0.892	0.891	0.890	0.881	0.878
	2.0	**0.909**	0.908	0.904	0.903	0.896	0.901	0.895	0.894	0.890
	2.5	**0.886**	0.884	0.881	0.883	0.884	0.883	0.874	0.883	0.878
AUPRC	0.5	**0.594**	0.580	0.500	0.536	0.517	0.464*	0.469*	0.473*	0.469*
	1.0	**0.513**	0.505	0.447	0.453	0.423	0.410*	0.435	0.415*	0.412*
	1.5	**0.373**	0.367	0.354	0.357	0.364	0.340	0.340	0.319	0.325
	2.0	**0.390**	0.380	0.364	0.375	0.352	0.355	0.359	0.362	0.367
	2.5	**0.384**	0.381	0.371	0.365	0.360	0.365	0.356	0.366	0.363

Performance for $\tau_{\max} = 5\delta$, $\delta = 1/2$ years. * indicates statistically significant difference ($p < 0.05$) with MATCH-Net.

USE CASE: PERSONALIZED SCREENING