Disease-Atlas: Navigating Disease Trajectories using Deep Learning

Bryan Lim & Mihaela van der Schaar
(2018)

Presented by Yang Qu & Qingyang Yu

Agenda

- Background
- Related Work & Problem
- Model
- Dataset
- Evaluation & Results
- Discussion

Background

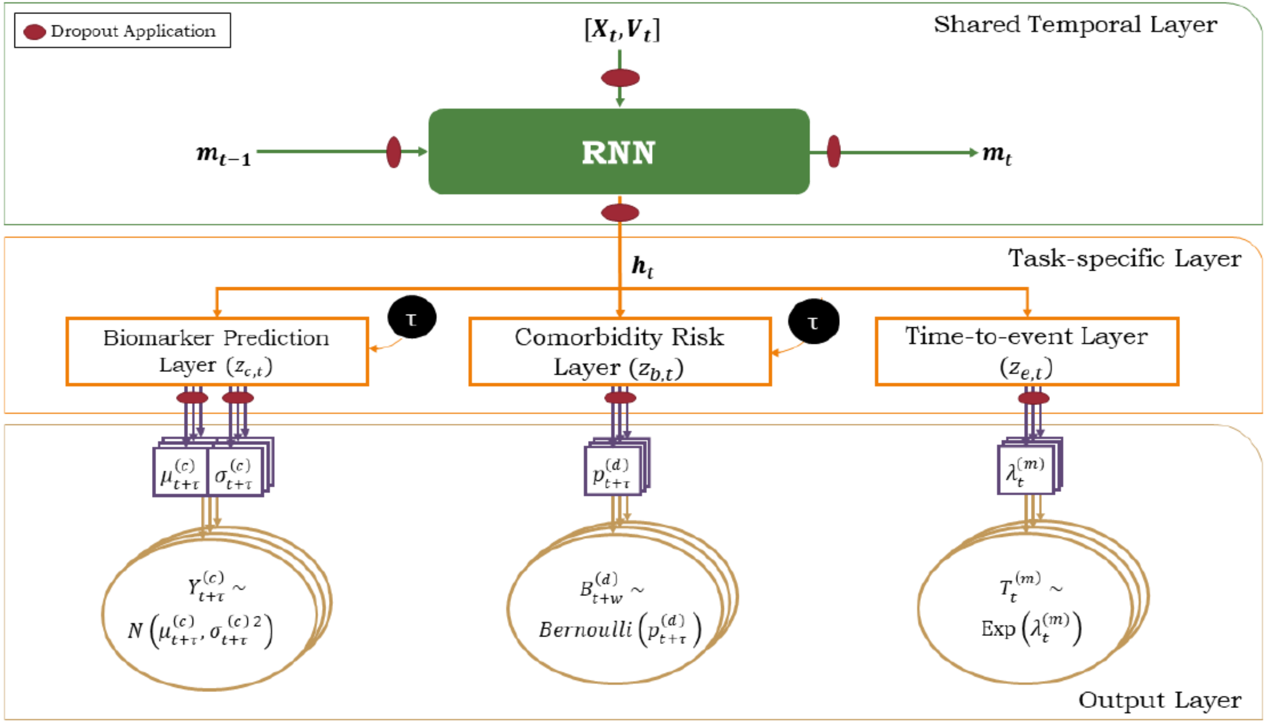
- Rich literature in Machine Learning models focusing on short-term predictions
- E.g., use data collected from in-hospital patients to predict the ICU admission
- Patient with chronic diseases are followed up over the span of years
- Additional comorbidities can in turn affect key biomarkers
- Increasing demand for jointly forecasting biomarker trajectories, comorbidity, and survival probability

Related Work + Problem

- Joint models in longitudinal studies
  - Standard joint model deals with high dimensional dataset
  - Gaussian process (GP) - incorporation with patient covariates is too simple

- Deep learning in traditional survival analysis
  - Fail to have dynamic prediction over time
  - Lack of uncertainty estimate

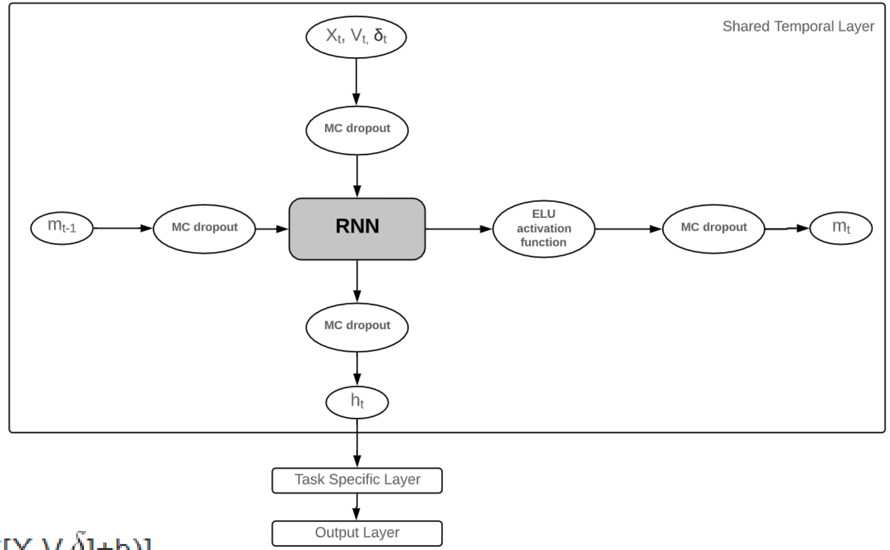
Disease-Atlas Network Architecture

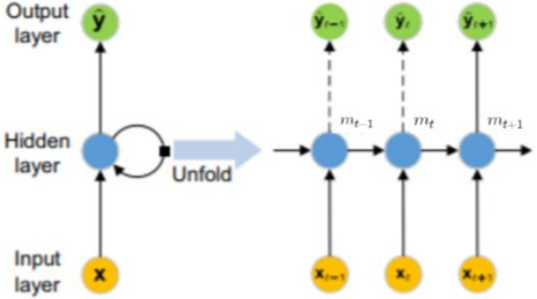


X_t : external covariates at time t
V_t : longitudinal measuremnt at time t
δ_t : event occurrences at time t
m_{t-1} : memory state at time t-1
m_t : memory state at time t
h_t : output at time t

Shared Temporal Layer

- Goal: Learn correlation between variables
- Input: Longitudinal data, Event Occurrence; Covariates; Memory state
- Activation Function: Exponential Linear Unit (ELU)
- Monte Carlo (MC) Dropout: Regularization & Uncertainty Prediction



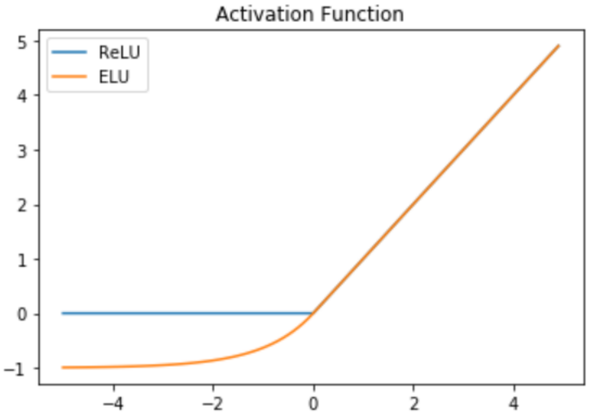


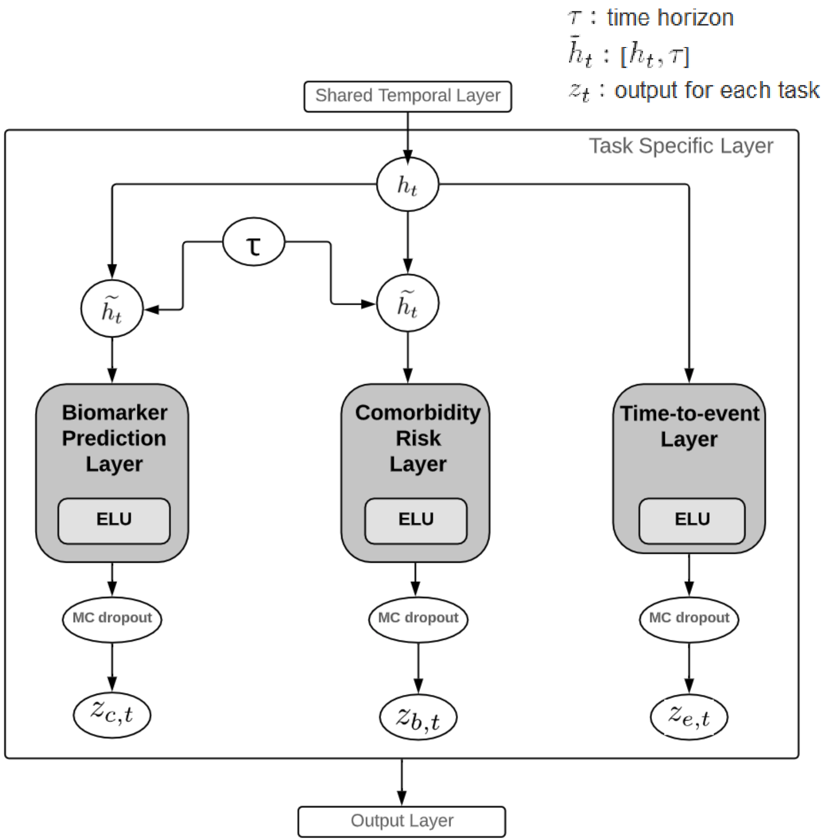
$$m_t = \text{ELU}[(W*m_{t-1}+b)+(W*[X,V,\delta]+b)]$$

W: weight matrix, b: bias matrix

Task-specific Layer

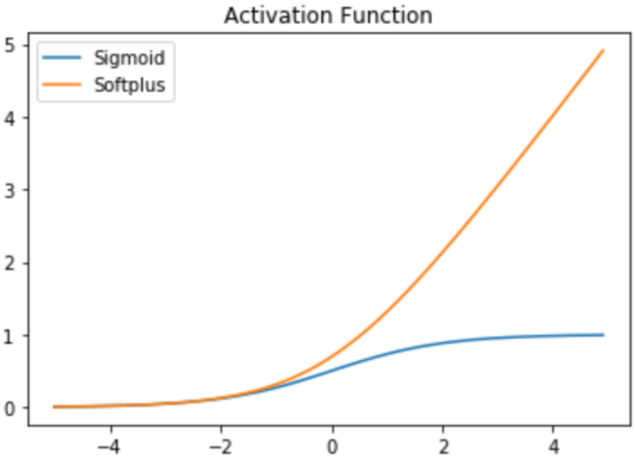
- Goal: Learn shared representations between related trajectories





Output Layer

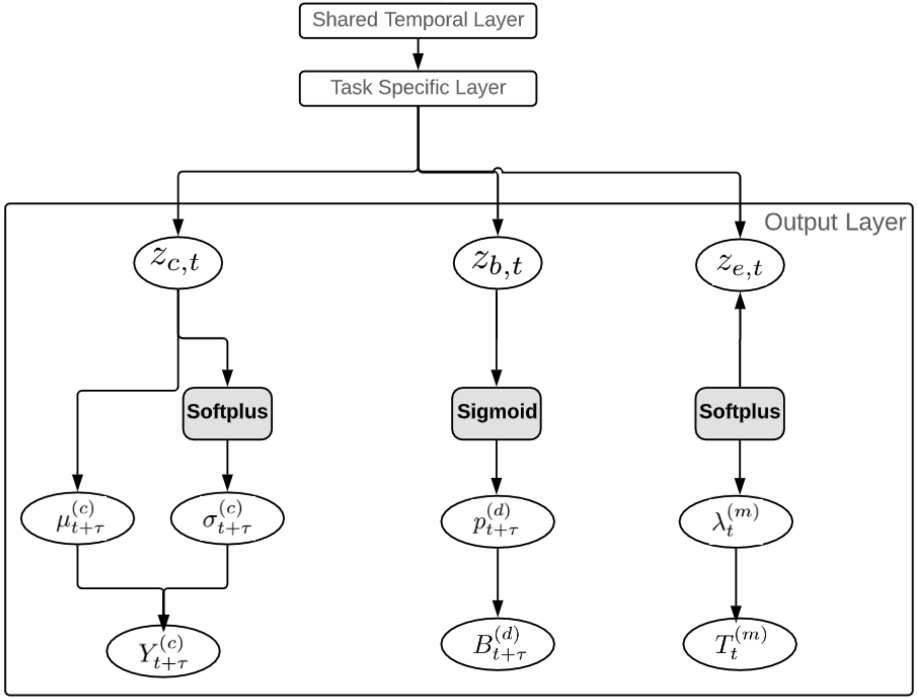
- Goal: Compute parameters for predictive sub-model distributions



$Y_{t+\tau}^{(c)}$: cth continuous longitudinal data at time t+ τ

$B_{t+\tau}^{(d)}$: dth binary longitudinal data at time t+ τ

$T_t^{(m)}$: time of mth event



$Y_{t+\tau}^{(c)} \sim N\left(\mu_{t+\tau}^{(c)}, \sigma_{t+\tau}^{(c)2}\right)$ $\quad B_{t+\tau}^{(d)} \sim \text{Bernoulli}\left(p_{t+\tau}^{(d)}\right)$ $\quad T_t^{(m)} \sim \text{Exponential}\left(\lambda_t^{(m)}\right)$

Multitask Learning

- Better Survival Representations

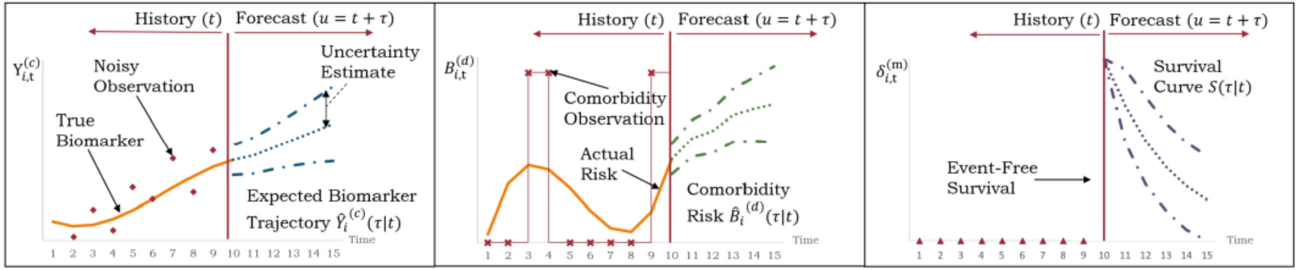
$$L(\mathbf{W}) = -\underbrace{\alpha_c \sum^{i,t,w,c} \log f_c\left(Y_{t+\tau}^{(c)}|\mathbf{W}\right)}_{\text{Continuous Longitudinal Loss } l_c} - \underbrace{\alpha_b \sum^{i,t,w,d} \log f_b\left(B_{t+\tau}^{(d)}|\mathbf{W}\right)}_{\text{Binary Longitudinal Loss } l_b}$$
$$- \underbrace{\alpha_T \sum^{i,t,m} \log f_T\left(T_t^{(m)}|\mathbf{W}\right)}_{\text{Time-to-event Loss } l_T}$$

- Handling Irregularly Sampled Data :
- Definition: Some data collected in consistent frequency, others not
- With Multitask: Improving the prediction accuracy without relying too much on the choice of imputation

Forecasting Disease Trajectories - Dynamic Prediction

- Estimation of the expected values of longitudinal variables and survival probabilities

- Uncertainty estimates with Monte-Carlo dropout approach



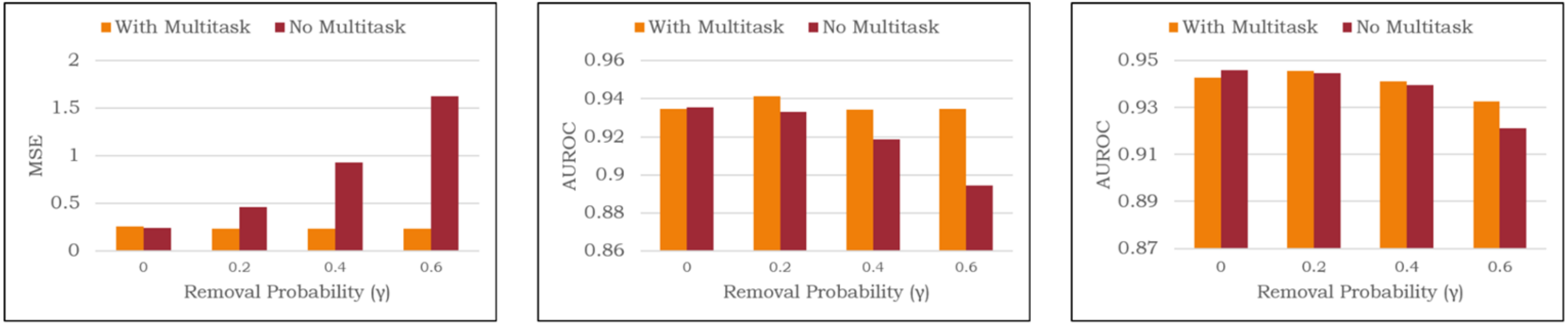
UK Cystic Fibrosis (CF) Registry Dataset

- 10,980 CF patients
- annual follow ups between 2008-2015
- a total of 87 variables associated with each patient across all years
- Interests lies in:
- 2 continuous lung function scores (FEV1 and Predicted FEV1)
- 20 binary longitudinal variables of comorbidity and infection
- death as the event of interest
- Training/Validation/Test split: 60%-20%-20%

		Type	% Patients
Event	Death	Binary (Event)	4.70%
Biomarkers	FEV1	Continuous	100.00%
	Predicted FEV1	Continuous	100.00%
Comorbidities	Liver Disease	Binary	20.80%
	Asthma	Binary	22.96%
	Arthropathy	Binary	9.50%
	Bone fracture	Binary	1.94%
	Raised Liver Enzymes	Binary	23.91%
	Osteopenia	Binary	20.37%
	Osteoporosis	Binary	9.58%
	Hypertension	Binary	3.30%
	Diabetes	Binary	24.56%
Bacterial Infections	Burkholderia Cepacia	Binary	5.59%
	Pseudomonas Aeruginosa	Binary	65.18%
	Haemophilus Influenza	Binary	30.55%
	Aspergillus	Binary	29.29%
	NTM	Binary	6.38%
	Ecoli	Binary	5.32%
	Klebsiella Pneumoniae	Binary	4.93%
	Gram-Negative	Binary	3.78%
	Xanthomonas	Binary	13.18%
	Staphylococcus Aureus	Binary	52.59%
	ALCA	Binary	5.06%

Multi-task Learning with Irregular Sampling

- The removal probability, gamma, is the probability that all data points are removed across each tasks at one time step.



(a) FEV1 MSE (b) Comorbidity AUROC (c) Mortality AUROC

Evaluation - Mortality Prediction

	τ	DA-LSTM	DA-NN	LSTM	MLP	L	JM
AUROC	1	**0.944(± 0.0004)**	0.943(± 0.0003)	0.943(± 0.0007)	0.941(± 0.0003)	0.824	0.870
	2	**0.924(± 0.0008)**	0.923(± 0.0005)	0.923(± 0.0005)	0.919(± 0.0003)	0.812	0.870
	3	**0.910(± 0.0003)**	0.905(± 0.0002)	0.908(± 0.0002)	0.907(± 0.0002)	0.825	0.851
	4	**0.905(± 0.0003)**	0.902(± 0.0008)	0.904(± 0.0003)	0.904(± 0.0006)	0.776	0.828
	5	**0.895(± 0.0003)**	0.892(± 0.0005)	0.894(± 0.0005)	0.888(± 0.0007)	0.765	0.806
AUPRC	1	**0.278 (± 0.0037)**	0.238 (± 0.0040)	0.230 (± 0.0020)	0.219 (± 0.0036)	0.161	0.119
	2	**0.193 (± 0.0014)**	0.169 (± 0.0033)	0.165 (± 0.0017)	0.186 (± 0.0036)	0.082	0.092
	3	0.103 (± 0.0005)	0.092 (± 0.0007)	0.099 (± 0.0028)	**0.105 (± 0.0001)**	0.085	0.089
	4	**0.109 (± 0.0007)**	0.101 (± 0.0014)	0.095 (± 0.0010)	0.102 (± 0.0006)	0.062	0.068
	5	**0.101 (± 0.0007)**	0.091 (± 0.0008)	0.093 (± 0.0017)	0.100 (± 0.0017)	0.058	0.059

$$TPR = \frac{TP}{TP+FN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$FPR = \frac{FP}{TN+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

Evaluation - Longitudinal Variables Prediction

	τ	MSE FEV1	MSE Pred. FEV1	AUROC [Mean ± SD] Comorbidities	AUROC [Mean ± SD] Infections	AUPRC [Mean ±SD] Comorbidities	AUPRC [Mean ±SD] Infections
DA-LSTM	1	**0.182**	**121.3**	**0.957 (± 0.025)**	**0.888 (± 0.056)**	**0.680 (±0.261)**	**0.416 (±0.247)**
	2	**0.191**	**139.4**	**0.926 (±0.047)**	**0.850 (± 0.044)**	**0.648 (± 0.244)**	**0.337 (± 0.261)**
	3	**0.275**	**191.3**	**0.882 (±0.048)**	**0.798 (± 0.057)**	**0.555 (± 0.213)**	**0.337 (± 0.261)**
	4	**0.374**	**254.4**	**0.817 (±0.085)**	**0.723 (± 0.068)**	**0.459 (± 0.184)**	**0.309 (± 0.252)**
	5	**0.461**	**308.1**	**0.790 (±0.067)**	**0.669 (± 0.126)**	**0.388 (± 0.169)**	**0.269 (± 0.247)**
JM	1	0.553	368.6	0.699 (± 0.148)	0.673 (± 0.069)	0.176 (± 0.088)	0.161 (± 0.176)
	2	0.593	411.1	0.694 (± 0.139)	0.651 (± 0.060)	0.180 (± 0.089)	0.157 (± 0.181)
	3	0.641	451.8	0.685 (± 0.140)	0.631 (± 0.072)	0.185 (± 0.090)	0.160 (± 0.186)
	4	0.695	490.1	0.681 (± 0.132)	0.607 (± 0.077)	0.187 (± 0.091)	0.159 (± 0.188)
	5	0.750	519.7	0.673 (± 0.130)	0.580 (± 0.082)	0.188 (± 0.093)	0.155 (± 0.186)

Discussion

Strengths:

- Handle high dimensional data
- Complex interaction between variables
- Uncertainty estimates
- Robustness to Irregular Sampling via Multitask Learning

Limitations:

- Imbalanced data: Resamping; Changing weight of loss function
- Many hyperparameters to tune (3600):
- Grid Search vs Random Search: Tradeoff between accuracy & computational efficiency
- Choice of activation function

Questions