Stabilizing Sparse Cox Model using Statistic and Semantic Structures in Electronic Medical Records

Shivapratap Gopakumar, Tu Dinh Nguyen, Truyen Tran, Dinh Phung, and Svetha Venkatesh

Center for Pattern Recognition and Data Analytics,
Deakin University, Australia 3216
{sgopakum, ngtu, truyen.tran, dinh.phung, svetha.venkatesh}@deakin.edu.au

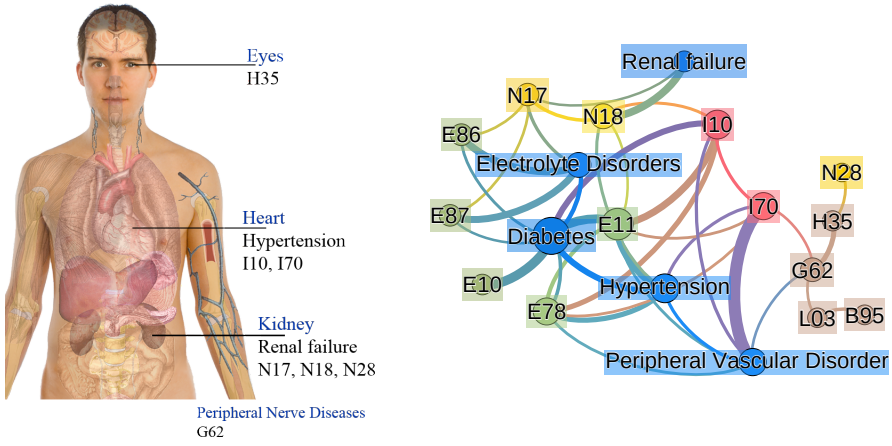
Abstract. Stability in clinical prediction models is crucial for transferability between studies, yet has received little attention. The problem is paramount in high dimensional data, which invites sparse models with feature selection capability. We introduce an effective method to stabilize sparse Cox model of time-to-events using statistical and semantic structures inherent in Electronic Medical Records (EMR). Model estimation is stabilized using three feature graphs built from (i) Jaccard similarity among features (ii) aggregation of Jaccard similarity graph and a recently introduced semantic EMR graph (iii) Jaccard similarity among features transferred from a related cohort. Our experiments are conducted on two real world hospital datasets: a heart failure cohort and a diabetes cohort. On two stability measures – the Consistency index and signal-to-noise ration (SNR) – the use of our proposed methods significantly increased feature stability when compared with the baselines.

1 Introduction

Stability is fundamental to prognosis. Besides good performance, a prognostic model needs to be interpretable and stable to warrant clinical adoption. This translates to a small group of succinct predictors that are consistent in the face of data re-sampling. Hence strong feature selection is key when deriving clinical models.

When data is high dimensional and correlated, automated feature selection causes instability in linear [1] and survival models [2]. These aspects are intrinsic to modern healthcare data. Medical events often co-occur, especially in aged cohorts. Comorbidities or diseases that co-exist with the primary disease in a patient, cause multiple diagnoses that are strongly correlated to each other. For example, Fig. (1.a) shows the common complications in a diabetic cohort. These correlations can be visualized as a network, as in Fig. (1.b). When deriving a prediction model from such data, we have to account for the complex interconnectedness of the features.

Integrating domain knowledge to improve learning has been gaining much attention recently [3]. Biological understanding of gene-disease networks, for example, has enabled discovery of what genes contribute to a disease, and what proteins would bind with a particular chemical compound [4]. However, little has been explored in networks derived from the healthcare processes and their contribution to prediction models. Domain knowledge, represented as networks, should ideally guide the feature selection process in clinical prediction.



(a) (b)

Fig. 1: Extracting disease correlations in diabetic cohort. Common comorbidities and diagnosis codes are shown in (a). A portion of the disease graph constructed using Jaccard similarity between EMR features in a diabetic cohort is shown in (b). The nodes represent EMR features, and links represent interaction strength, measured using Jaccard index. Blue nodes are co-occurring diseases, green nodes are diagnosis codes for diabetes, orange nodes for heart diseases, and yellow nodes for urinary diseases.

In this paper, we address the problem of stabilizing a high dimensional model derived from routinely collected EMR data. We focus on minimizing the variance in feature subsets and model estimation parameters. We construct a feature graph with nodes as EMR features and edges as relationship between the features. We look at three feature relationships: (i) Jaccard score between features (ii) aggregate of Jaccard score and the semantic EMR link used in [5] (iii) Jaccard scores between features transferred from a related cohort. A random walk regularization of the proposed graphs is used to stabilize a sparse Cox model that predicts time to readmission. Our experiments are conducted on 2 real world hospital datasets: a heart failure cohort and a diabetes cohort. We measure feature stability using the Consistency index and model estimation stability using signal-to-noise ratio (SNR). Our proposed method, when compared with elastic net and recently introduced semantic EMR graph regularization [5], confirmed better feature and model stability when validated against both cohorts.

In summary, our main contributions are:

1. Representation of medical domain knowledge as feature graphs that embed (i) statistical correlations between features using Jaccard index (ii) aggregate of statistical and semantic correlation among features (iii) correlations between features transferred from a related cohort.
2. A random walk regularizer based on the proposed feature graphs to stabilize a Cox model as opposed to the traditional Laplacian regularizer. While Laplacian reg-

ularizer focuses on pairwise similarity, the random walk regularizer encourages groupwise similarity.
3. Demonstration of improved feature stability as measured by the Consistency index and improved model stability as measured by signal-to-noise ratio (SNR) for model regularization using proposed feature graphs. The stability measures were compared with lasso, elastic net and a recently introduced Laplacian semantic EMR graph regularization [5] on a cohort of $1,784$ index admissions in heart failure patients and $2,370$ index admissions in diabetic patients admitted to a regional hospital in Australia.
4. Demonstration of improved stability, using transfer learning on related cohorts. Related cohorts like heart failure and diabetes share comorbidities and predictors. Hence, the feature graph constructed from one cohort is used to stabilize the model derived from related cohort. Stability is measured using the Consistency index and SNR.

The significance of our study lies in understanding the importance of incorporating underlying feature relationships into model learning. This promotes stable feature selection in a clinical setting.

2 Related Background

Stabilizing clinical prediction has received little attention, partly because most models are built using a small subset of well-defined predictors, chosen either by domain experts or from prior knowledge. In most high dimensional models, the primary regularizer of choice is the lasso because of its convexity and sparsity inducing property [6]. However, when data is correlated, as in the healthcare domain, lasso regularized models are susceptible to data variations resulting in loss of stability [1]. The inconsistency of Lasso to handle data correlation has been demonstrated for linear models [7] and recently for survival models [2]. When prior knowledge about feature relationships are available, Sandler *et al*. proposed additional regularization using graph networks where the nodes are features and edges represent relationships [3]. Graph regularization ensures statistical weight sharing among related features. In such scenarios, the first challenge is to identify useful prior information.

High-dimensional models in bioinformatics domain often resort to online databases like KEGG, Pathway Commons and BioCarta to extract context specific data to build feature graphs [8]. Recent linear classification models have started using such gene-pathway networks and protein interaction networks to improve prediction accuracy and model interpretability [9]. For genomic data, a recent study employed a quadratic Laplacian regularizer into Cox regression, where the Laplacian graph was derived from prior gene regulatory network information [10]. Another study investigated the significance of including domain knowledge as network information into Cox model for identifying biomarkers in breast cancer [11]. Specifically, the study compared eight network based regularizers and three non-network based regularizers for Cox regression. All methods were validated on five public breast cancer datasets. The study observed no significant advantage for network-based approaches over non-network-based approaches in terms of prediction performance or signature stability.

In the healthcare domain, elastic net regularized Cox model showed superior performance over Lasso for prostate cancer dataset [12]. Though elastic net regularization can handle correlated features, it cannot incorporate structural relationships. Vinzamuri *et al*. introduced a modification to the elastic net involving an RBF kernel for handling feature correlations to improve accuracy and reduce redundancy [13]. In contrast, our work focuses on handling correlations to stabilize the model estimation and top predictors. Feature similarity is captured by building a Jaccard similarity graph for additional regularization. Recent studies use semantic graphs constructed from ICD-10 code relations and temporal relations in medical events to stabilize linear readmission models derived from electronic medical records [5]. We compare our approach with this EMR graph regularization. Further, we investigate the effect of aggregating the semantic EMR graph with the statistical Jaccard graph on model stability.

3 Framework

3.1 Sparse Cox Model

We use Cox regression to model risk of readmission (hazard function) at a future time instance, based on data from EMR. Unlike logistic regression where each patient is assigned a nominal label, Cox regression models the readmission time directly [13]. The proportional hazards assumption in Cox regression assumes a constant relationship between readmission time and EMR-derived explanatory variables. Let $\mathcal{D} = \{\boldsymbol{x}_n, y_n\}_{n=1}^{N}$ be the training dataset with N observations, ordered on increasing y_n, where $\boldsymbol{x}_n \in \mathbb{R}^K$ denotes the feature vector for n^{th} index admission and y_n is the time to next unplanned readmission. When a patient withdraws from the hospital or does not encounter readmission in our data during the follow-up period, the observation is treated as right censored. Let M observations be uncensored and $R(t_n)$ be the remaining events at readmission time t_n.

Since the data $\mathcal{D}$ is high dimensional (possibly $K \gg N$), we apply lasso regularization for sparsity induction [14]. The feature weights $\boldsymbol{w} \in \mathbb{R}^K$ are estimated by maximizing the ℓ_1-penalized partial likelihood:

$$\mathcal{L}_{\text{lasso}} = \frac{1}{N}\mathcal{L}\left(\boldsymbol{w}; \mathcal{D}\right) - \alpha \sum_k |w_k| \tag{1}$$

where $\|\boldsymbol{w}\|_1 = \sum_k |w_k|$, $\alpha > 0$ is the regularizing constant, and $\mathcal{L}\left(\mathbf{w}; \mathcal{D}\right)$ is the log partial likelihood [15] computed as:

$$\mathcal{L}\left(\mathbf{w}; \mathcal{D}\right) = \sum_{m=1}^{M} \left\{ \boldsymbol{w}^\top \boldsymbol{x}_m - \log \left[\sum_{\jmath \in R(t_m)} \exp\left(\boldsymbol{w}^\top \boldsymbol{x}_\jmath\right) \right] \right\}$$

Lasso induces sparsity by driving the weights of weak features towards zero. However, sparsity induction is known to cause instability in feature selection [16]. Instability occurs because Lasso randomly chooses one in two highly correlated features. Each training run with slightly different data could result in a different feature from the correlated pair. The nature of EMR data further aggravates this problem. The EMR data

is, by design, highly correlated and redundant. Also, features in the EMR data maybe weakly predictive for some task, thereby limiting the probability that they are selected. These sum up to lack of reproducibility between model updates or external validations, hindering the method credibility and adoption by clinicians.

A popular solution to instability is elastic net [12]. Elastic net regularization modifies the likelihood function in Eq. (1) as:

$$\mathcal{L}_{\text{elastic.net}} = \frac{1}{N}\mathcal{L}\left(\boldsymbol{w}; \mathcal{D}\right) - \alpha\left(\lambda \sum_k |w_k| + (1-\lambda)\sum_k w_k^2\right) \tag{2}$$

Here, the ridge regression term $\sum_k w_k^2$ tends to give equal weights to correlated features, while the lasso term $\sum_k |w_k|$ introduces sparsity. However, this formulation ignores domain knowledge.

3.2 Stabilization using Feature Graph

Medical events often co-occur, especially in aged cohorts. For example, the presence of comorbidities causes multiple diagnoses at the same time. We capture feature correlation in a knowledge network, with features as nodes and relations between features as edges. Let the adjacency matrix of the feature graph be G, where $G_{ij} = g \in (0, 1)$ represents the weighted similarity score between features i and j. We ensure all features have equal prominence by constraining the out-links of each node to sum to one. The medical events linked together in the feature graph should have similar weights.We introduce a random walk regularizer [3]:

$$\begin{aligned}\Omega(\boldsymbol{w};\, \mathbf{G}) &= \sum_k \left(w_k - \sum_i G_{ki} w_i\right)^2 \\ &= \boldsymbol{w}^\top \left(I - \mathbf{G}\right)^\top \left(I - \mathbf{G}\right) \boldsymbol{w}\end{aligned} \tag{3}$$

where I is the identity matrix. The graph stabilized model likelihood can be written as:

$$\mathcal{L}_{\text{graph}} = \mathcal{L}_{\text{lasso}} - \frac{1}{2}\beta \boldsymbol{w}^\top \left(I - \mathbf{G}\right)^\top \left(I - \mathbf{G}\right) \boldsymbol{w} \tag{4}$$

Here the $\ell_1$regularizer introduces sparsity by pushing weak features towards zero, while the random walk regularizer distributes smoothness equally among correlated features. The gradient of Eq. (4) becomes:

$$\begin{aligned}\frac{\partial \mathcal{L}_{\text{graph}}}{\partial w} = \sum_{m=1}^{M} \left\{ \boldsymbol{x}_m - \frac{\sum\limits_{\jmath \in R(t_m)} \boldsymbol{x}_\jmath \exp(\boldsymbol{w}^\top \boldsymbol{x}_\jmath)}{\sum\limits_{\jmath \in R(t_m)} \exp(\boldsymbol{w}^\top \boldsymbol{x}_\jmath)} \right\} \\ - \alpha\, \text{sign}(w) - \beta \left(I - \mathbf{G}\right)^\top \left(I - \mathbf{G}\right) \boldsymbol{w}\end{aligned} \tag{5}$$

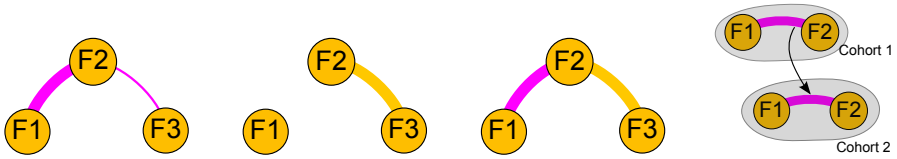
Parameter estimation is done using L-BFGS algorithm[17]. We build and compare different feature graphs to stabilize our model. Each feature graph differs in the construction of its adjacency matrix $\mathbf{G}$. A recent study [5] introduced a semantic EMR graph,

where nodes denoted features and edges denoted a temporal relation or ICD-10 structural relation between features (Fig. 2.(b)). Using this as a baseline, we construct $\mathbf{G}$ using the following methods. First, we represent the edges using the Jaccard index between features, as in Fig. 2.(a). Second, we aggregate the baseline semantic EMR graph and the Jaccard graph. Here, each edge is the maximum of Jaccard and semantic scores between the features (Fig. 2.(c)). Finally, we investigate transferring the adjacency matrix between related cohorts. Specifically, the Jaccard similarity scores between features in one cohort is transferred to a related cohort (Fig. 2.(d)). We detail these methods below.

Jaccard graph. The *Jaccard index* is the measure of the percentage of agreement between components among feature vectors. Given two feature vectors F_i and F_j, the pairwise Jaccard score reads:

$$J_{ij} = \frac{a}{a+b+c} \tag{6}$$

where a is the number of non-zero components in F_i and F_j, b is the number of non-zero components in F_i but not in F_j and c is number of non-zero components in F_j but not in F_i. We construct an undirected graph with nodes as features and edges representing the Jaccard score between features.



(a) Jaccard (b) semantic (c) aggregate (d) TL Jaccard

Fig. 2: Feature correlation captured by constructing feature graph with nodes as features and edges as: (a) statistical correlation measured using Jaccard score (b) semantic relations derived from temporal and ICD-10 structures (c) aggregation of Jaccard and Semantic graphs. (d) transfer of Jaccard similarity between features from a related cohort

Graph aggregation. We investigate the effect of combining the semantic EMR graph with Jaccard graph on model stability and feature stability. The semantic EMR graph captures the general relationship between diagnostic codes based on the ICD-10 structures, while the Jaccard graph is cohort specific. Here, we use a simple aggregation technique to construct the final $\langle \text{EMR}; \text{Jaccard} \rangle$ graph as:

$$G_{\langle \text{EMR;Jaccard} \rangle} = \max(G_{\text{EMR}}, G_{\text{Jaccard}}) \tag{7}$$

Transfer learning. Finally, we examine the capability of our proposed method in transfer learning. Knowledge from one domain can be transferred to a related domain when data is scare or expensive to collect [18]. Getting high quality training data is often difficult, particularly in a medical setting. Cohorts that share comorbidities and diagnoses,

	Heart failure		Diabetes	
	Training set	Testing set	Training set	Testing set
Checkpoint	Sep 2010		Dec 2008	
Number of admissions	1,415	369	1,341	1,029
Unique patients	1,088	317	951	765
Gender				
Male	541 (49.7%)	155 (48.9%)	501 (52.68%)	407 (53.20%)
Female	547 (50.2%)	162 (51.1%)	450 (47.32%)	358 (46.80%)
Mean age (years)	78.3	79.4	57.8	56.4

Table 1: Characteristics of training and validation cohorts.

as in diabetes and cardiovascular diseases, are likely to have similar correlations among features. Accordingly, we propose to stabilize a Cox model derived from one cohort using the Jaccard similarity graph constructed from a related cohort. We denote the transferred graph as: TL-Jaccard graph. Further, we use TL-Jaccard graph to construct the aggregated graph:

$$G_{\langle \text{EMR;TL-Jaccard} \rangle} = \max(G_{\text{EMR}}, G_{\text{TL-Jaccard}})$$

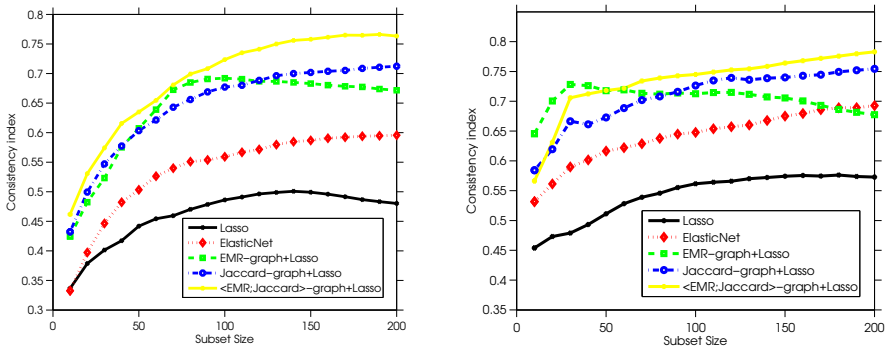
Here, the temporal and hierarchical feature relations in the cohort are captured by the EMR graph. The statistical relations among features, which can be expensive to calculate, are transferred from the related cohort using TL-Jaccard graph.

4 Experiments

In this section, we evaluate feature and model stability of our framework. The results are reported on two cohorts: heart failure (HF) and diabetes (DB), provided by Barwon Health, a regional health service provider which has been serving more than 350,000 residents in Victoria, Australia[1].We collect retrospective data for heart failure and diabetes patients from the hospital EMR database. The heart failure cohort contains all patients with at least one ICD-10 diagnosis code I50, while the diabetes cohort includes all patients with at least one diagnosis code between E10-E14. This resulted in 1,885 heart failure admissions and 2,840 diabetes admissions between January 2007 and December 2011. Patients of all age groups were included whilst inpatient deaths were excluded. We focus our study on emergency attendances and unplanned admissions of patients.

We use the one-sided convolutional filter bank introduced in [19] to extract a large pool of features from EMR databases. The filter bank summarizes event statistics over multiple time periods and granularities. The feature extraction process resulted in 3,338 features for heart failure cohort and 7,641 features for diabetes cohort. The extracted features are used to derive a sparse Cox model. Our proposed feature graphs capture correlations between these features to stabilize model learning.

[1] Ethics approval was obtained from the Hospital and Research Ethics Commitee at Barwon Health (number 12/83) and Deakin University.



(a) Heart failure.

(b) Diabetes.

Fig. 3: Stabilization using statistical and semantic structures. Feature stability measured by the consistency index as functions of the subset size for readmission prediction within 6 months for heart failure (Fig. 3a) and 12 months for diabetes patients (Fig. 3b). Larger indices imply more stability.

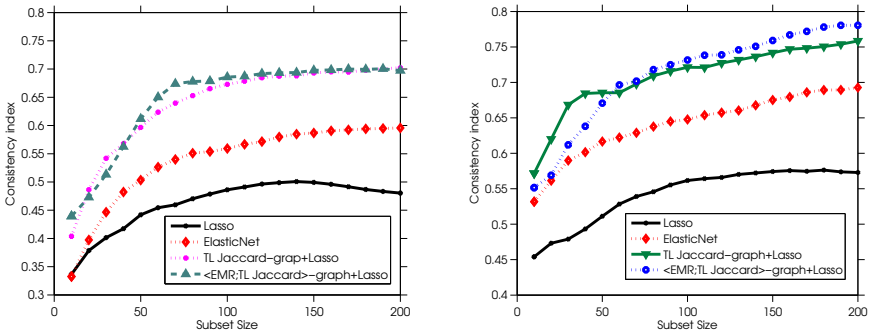
4.1 Evaluation protocol

The baseline regularization methods for Cox regression are chosen to be (i) lasso (ii) elastic net (iii) semantic EMR graph (as in [5]). Based on the construction of the feature graph, we arrive at four different models: (i) Jaccard graph regularized model: feature graph is the Jaccard similarity graph among features in the given cohort (ii) EMRJaccard regularized model: feature graph is the aggregation of Jaccard graph with semantic EMR graph, as in Eq. (7) in the given cohort (iii) TL Jaccard regularized model: feature graph is the Jaccard similarity graph transferred from a related cohort (iv) EMR; TL Jaccard regularized model: feature graph is the aggregation of semantic EMR graph from the given cohort and Jaccard graph transferred from a related cohort.

Temporal validation. We ensure that the training and testing sets are completely separated in time . This validation strategy is chosen because it better reflects the common practice of training the model in the past and using it in the future.We gather admissions which have discharge dates before September 2010 for heart failure and before 2009 for diabetes patients to form the training set and after that for testing. Next we specify the set of unique patients in the training set. We then remove all admissions of such patients in the testing set to guarantee no overlap between two sets. The statistical characteristics of two cohorts are summarized in Table 1.

Our model is then learned using training data and evaluated on testing data. Model performance is evaluated using measures of AUC (area under the ROC curve) with confidence intervals based on Mann-Whitney statistic. The AUC is computed basing upon the ranking of hazard rates of the patient readmissions.

Measuring Model Stability. We use the consistency index [20] to measure stability of feature selection process. The consistency index (CI) supports feature selection in ob-



(a) Heart failure.

(b) Diabetes.

Fig. 4: Stabilization using transfer of Jaccard graph (TL Jaccard). Stabilization using statistical and semantic structures. Feature stability measured by the consistency index as functions of the subset size for readmission prediction within 6 months for heart failure (Fig. 4a) and 12 months for diabetes patients (Fig. 4b). Larger indices imply more stability.

taining several desirable properties, i.e., monotonicity, limits and correction for chance. To simulate data variations due to sampling, we create B data bootstraps of original size n. For each bootstrap, a model is trained and a subset of top k features is selected. Features are ranked according to their importance, which are product of feature weight and standard deviation. Finally, we obtain a list of feature subsets $S = \{S_1, S_2, ..., S_B\}$ where $|S_b| = k$.

The *consistency index* corrects the overlapping due to chance. Considering a pair of subsets S_i and S_j, the pairwise consistency index I_C is defined as:

$$CI(S_a, S_b) = \frac{rd - k^2}{k(d-k)} \tag{8}$$

in which $|S_a \cap S_b| = r$ and d is the number of features. The stability for the set $S = \{S_1, S_2, ..., S_B\}$ is calculated as average across all pairwise $CI(S_a, S_b)$. The consistency index is bounded in $[-1, +1]$.

We further our investigations on the model stability. The model estimation stability is defined as variance in parameters. A measure is the signal-to-noise ratio (SNR): $\text{SNR}\,(i) = \bar{w}_i/\sigma_i$ in which $\bar{w}_i$ is the mean feature weight across bootstraps for feature i, and σ_i is its standard deviation. We take the average of the 20 highest SNR values. Higher score indicates better stability.

4.2 Results

Our models are designed using two hyper-parameters: lasso regularization parameter α and graph regularization parameter β. We empirically tune these parameters to improve feature stability without hurting model discrimination. Overall, feature stability depended more on graph parameter β, while model discrimination was more sensitive to

α. A good tradeoff was achieved at $\alpha = 0.003$ and $\beta = 0.8$. All models are externally validated against (i) heart failure cohort with a 6-month horizon (ii) diabetes cohort with a 12-month horizon. Table 2 reports the AUC scores with confidence intervals for the different models. The predictive performance of our proposed graph stabilization models and transfer learning techniques are comparable with the baselines.

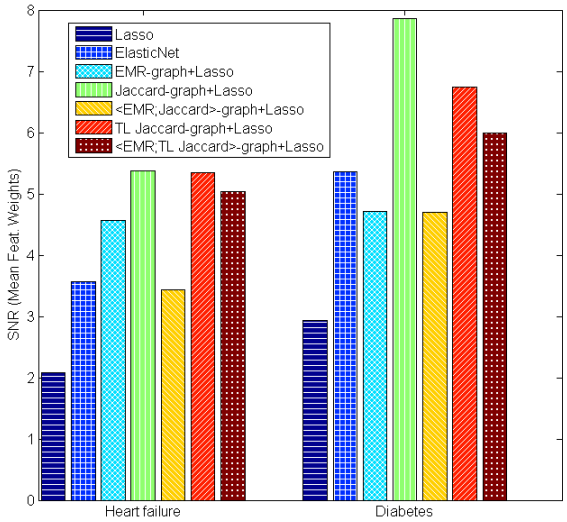


Fig. 5: Model estimation stability measured by signal-to-noise ratios (SNR) of feature weights. High value of SNR indicates more stability. *"TL Jaccard"* means the Jaccard-graph used in transfer learning settings: heart failure Jaccard graph for stabilizing diabetes data and diabetes Jaccard graph for stabilizing heart failure cohort.

Stabilization using Statistical and Semantic Graphs Graph regularized models consistently produced more stable features than lasso and elastic net regularized models. When comparing different graph regularizations, we found the semantic EMR graph to be more effective for small feature subsets (see Fig. 3). For increasing feature subset sizes (>100), Jaccard graphs proved effective. The temporal and structural relations of diagnosis codes have stronger effect for small set of features, while Jaccard index was effectual on larger sets. This behavior suggests aggregating statistical and semantic structures. For the top 100 predictors, EMRJaccard graph stabilization demonstrated the highest feature stability in both cohorts (see Fig. 3).

Next, we compare variance in parameter weights using SNR measures. In Fig. 5, each model is represented by average of its 20 highest SNR values. The Jaccard graph regularized model proved to be most robust in both cohorts. Interestingly, model stability using EMRJaccard graph was similar to elastic net and was not able to improve upon semantic EMR graph or Jaccard graph.

	Heart failure	Diabetes
Lasso	0.60 [0.55; 0.66]	0.74 [0.70; 0.77]
Elastic net	0.61 [0.55; 0.67]	0.75 [0.72; 0.79]
EMR-graph+Lasso	0.61 [0.56; 0.67]	0.76 [0.72; 0.79]
Jaccard-graph+Lasso	0.62 [0.56; 0.67]	0.76 [0.73; 0.79]
⟨EMR; Jaccard⟩-graph+Lasso	0.62 [0.56; 0.67]	0.76 [0.73; 0.79]
Stabilization using Transfer Learning		
TL_Jaccard-graph+Lasso	0.62 [0.56; 0.67]	—
⟨HF_EMR; TL_Jaccard⟩-graph+Lasso	0.62 [0.57; 0.68]	—
TL_Jaccard-graph+Lasso	—	0.76 [0.73, 0.79]
⟨DB_EMR; TL_Jaccard⟩-graph+Lasso	—	0.75 [0.72, 0.79]

Table 2: AUC scores with confidence intervals for readmission prediction within 6 months for heart failure and 12 months for diabetes patients. Model performance on individual cohorts and on cohorts with Jaccard graph transferred from the other cohort is shown in separate sections.

Stabilization using transfer learning. We investigate transfer of feature graphs between related cohorts. For the heart failure cohort, TL Jaccard graph represents the Jaccard scores transferred from diabetes cohort, while EMR;TL Jaccard graph is the aggregation of the semantic EMR graph of heart failure cohort and Jaccard graph transferred from diabetes cohort. The same technique is applied to stabilize diabetes cohort, where the Jaccard scores are transferred from heart failure cohort. We compare the transferred graph stabilizations with lasso and elastic net. Figs. (4;5) show that the cross-domain graphs also help the stabilities of feature selections and model estimation.

5 Discussion and Conclusion

Stability facilitates reproducibility between model updates and generalization across medical studies. Stable predictors inspire confidence in prognosis, as they are often subjected to further examinations. In this paper, we utilize statistical and semantic relations in EMR data to stabilize a sparse Cox model for predicting readmission. Unlike recent work that concentrate on getting better features during model learning [21,22], we focus on stabilizing the features selected by the model. The model is validated on two different retrospective cohorts. The random walk regularization of the aggregated feature graph promotes group level selection and rare-but-important features. On two stability measures, the proposed method has demonstrated to largely improved stability. In related cohorts, when collecting data becomes expensive, transferring domain knowledge using TL-Jaccard graph was also found to improve stability. Also, our proposed model is derived entirely from commonly available data in medical databases. All these factors suggest that our model could be easily integrated into the clinical pathway to serve as a fast and inexpensive screening tool in selecting features and patients for further investigation. Future work includes applying the same technique for a variety of cohorts and investigating other latent correlations in EMR to enhance feature stability.

References

1. Austin, P.C., Tu, J.V.: Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. Journal of clinical epidemiology **57** (2004) 1138–1146
2. Lin, W., Lv, J.: High-dimensional sparse additive hazards regression. Journal of the American Statistical Association **108** (2013) 247–264
3. Sandler, T., Blitzer, J., Talukdar, P.P., Ungar, L.H.: Regularized learning with networks of features. In: Advances in Neural Information Processing Systems 21. Curran Associates, Inc. (2009) 1401–1408
4. Barabási, A.L., Gulbahce, N., Loscalzo, J.: Network medicine: a network-based approach to human disease. Nature Reviews Genetics **12** (2011) 56–68
5. Tran, T., Phung, D., Luo, W., Venkatesh, S.: Stabilized sparse ordinal regression for medical risk stratification. Knowledge and Information Systems (2014) 1–28
6. Ye, J., Liu, J.: Sparse methods for biomedical data. ACM SIGKDD Explorations Newsletter **14** (2012) 4–15
7. Zhao, P., Yu, B.: On model selection consistency of lasso. The Journal of Machine Learning Research **7** (2006) 2541–2563
8. Cun, Y., Fröhlich, H.: Biomarker gene signature discovery integrating network knowledge. Biology **1** (2012) 5–17
9. Dao, P., Wang, K., Collins, C., Ester, M., Lapuk, A., Sahinalp, S.C.: Optimally discriminative subnetwork markers predict response to chemotherapy. Bioinformatics **27** (2011) i205–i213
10. Sun, H., Lin, W., Feng, R., Li, H.: Network-regularized high-dimensional cox regression for analysis of genomic data. Statistica Sinica **24** (2014) 1433–1459
11. Fröhlich, H.: Including network knowledge into cox regression models for biomarker signature discovery. Biometrical Journal **56** (2014) 287–306
12. Simon, N., Friedman, J., Hastie, T., Tibshirani, R.: Regularization paths for coxŠs proportional hazards model via coordinate descent. Journal of Statistical Software **39** (2011) 1–13
13. Vinzamuri, B., Reddy, C.: Cox regression with correlation based regularization for electronic health records. In: Data Mining (ICDM), 2013 IEEE 13th International Conference on. (2013) 757–766
14. Tibshirani, R., et al.: The lasso method for variable selection in the cox model. Statistics in medicine **16** (1997) 385–395
15. Cox, D.R.: Partial likelihood. Biometrika **62** (1975) 269–276
16. Xu, H., Caramanis, C., Mannor, S.: Sparse algorithms are not stable: A no-free-lunch theorem. Pattern Analysis and Machine Intelligence, IEEE Transactions on **34** (2012) 187–193
17. Liu, D.C., Nocedal, J.: On the limited memory bfgs method for large scale optimization. Mathematical programming **45** (1989) 503–528
18. Pan, S.J., Yang, Q.: A survey on transfer learning. Knowledge and Data Engineering, IEEE Transactions on **22** (2010) 1345–1359
19. Tran, T., Phung, D.Q., Luo, W., Harvey, R., Berk, M., Venkatesh, S.: An integrated framework for suicide risk prediction. In: KDD. (2013) 1410–1418
20. Kuncheva, L.I.: A stability index for feature selection. In: Artificial Intelligence and Applications. (2007) 421–427
21. Vinzamuri, B., Li, Y., Reddy, C.K.: Active learning based survival regression for censored data. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. CIKM '14, New York, NY, USA, ACM (2014) 241–250
22. Bilal, E., Dutkowski, J., Guinney, J., Jang, I.S., Logsdon, B.A., Pandey, G., Sauerwine, B.A., Shimoni, Y., Vollan, H.K.M., Mecham, B.H., et al.: Improving breast cancer survival analysis through competition-based multidimensional modeling. PLoS computational biology **9** (2013) e1003047