PREDIÇÃO DINÂMICA DE RISCO DE MORTALIDADE EM UTI USANDO ADAPTAÇÃO DE DOMÍNIO

TIAGO ALVES

PREDIÇÃO DINÂMICA DE RISCO DE MORTALIDADE EM UTI USANDO ADAPTAÇÃO DE DOMÍNIO

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Orientador: Alberto H. F. Laender

Belo Horizonte

Outubro de 2017

TIAGO ALVES

DYNAMICALLY PREDICTION OF ICU MORTALITY RISK USING DOMAIN ADAPTATION

Dissertation presented to the Graduate Program in Computer Science of the Universidade Federal de Minas Gerais in partial fulfillment of the requirements for the degree of Master in Computer Science.

ADVISOR: ALBERTO H. F. LAENDER

Belo Horizonte

October 2017

Alves, Tiago
M1234x Dynamically Prediction of ICU Mortality Risk Using Domain Adaptation / Tiago Alves. — Belo Horizonte, 2017
xxvi, 31 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal de Minas Gerais
Orientador: Alberto H. F. Laender

1. Machine Learning. 2. Deep Learning. 3. ICU. 4. Mortality Prediction. 5. Domain Adaptation. I. Title.

CDU 100.0*01.10

[Folha de Aprovação]

Quando a secretaria do Curso fornecer esta folha,
ela deve ser digitalizada e armazenada no disco em formato gráfico.

Se você estiver usando o `pdflatex`,
armazene o arquivo preferencialmente em formato PNG
(o formato JPEG é pior neste caso).

Se você estiver usando o `latex` (não o `pdflatex`),
terá que converter o arquivo gráfico para o formato EPS.

Em seguida, acrescente a opção `approval={`*nome do arquivo*`}`
ao comando `\ppgccufmg`.

Se a imagem da folha de aprovação precisar ser ajustada, use:
`approval=[`*ajuste*`][`*escala*`]{`*nome do arquivo*`}`
onde *ajuste* é uma distância para deslocar a imagem para baixo
e *escala* é um fator de escala para a imagem. Por exemplo:
`approval=[-2cm][0.9]{`*nome do arquivo*`}`
desloca a imagem 2cm para cima e a escala em 90%.

Dedico este trabalho aos bravos futuros cientistas que irão consultá-lo, que eles possam se lembrar sempre que ninguém nasce sabendo.

Acknowledgments

Agradeço aos meus alunos que me inspiraram a organizar este texto.

Agradeço ao Vilar da Camara Neto por montar e manter o modelo do PPGCC[1].

Dica I: [2] ***Geralmente, se agradece a pessoas/organizações importantes no processo da pós-graduação: família, orientadores, professores, colegas do laboratório, funcionários, CNPq/CAPES/FAPEMIG ou outra agência de fomento da qual tenha obtido bolsa ou que tenha financiado projeto do qual participou.***

[1] Classe do modelo utilizado pelo PPGCC, veja referência [da Camara Neto, 2011].

[2] Neste documento são apresentadas várias dicas relacionadas ao texto. Tais dicas são apresentadas em azul e itálico, numeradas e geralmente são comentários sobre o texto, dicas de estilo ou observações que enfatizam exemplos apresentados.

"Abandon all hope, ye who enter here"
(Dante Alighieri)

Resumo

A Universidade, antes de mais nada, é um ambiente de ensino e aprendizagem, onde professores *tentam* ensinar e estudantes *tentam* aprender da melhor maneira possível. Como parte integrante desse processo, é necessário que estudantes escrevam sobre os problemas e as soluções encontradas para os mesmos. Nesse contexto, o objetivo deste trabalho é *tentar* ensinar estudantes a organizar suas monografias, sejam elas trabalhos de conclusão de curso, dissertações de mestrado ou teses de doutorado. Como efeito colateral, espera-se que o tempo dos professores seja poupado. Em vez de ter de explicar para cada estudante como se escreve um trabalho, o professor poderá utilizar este texto como fonte para simplificar tal tarefa, que é certamente árdua. Neste texto, cada um dos seus capítulos e seções autoaborda o seu conteúdo. Ou seja, o resumo apresenta como o resumo deve ser escrito, a introdução como a mesma deve ser organizada e assim por diante. Espera-se que ao final, o estudante tenha aprendido o básico para começar a escrever seu trabalho.

Este parágrafo inicial é um bom exemplo de resumo. Ele define o contexto (primeira frase), o problema (segunda frase), o objetivo (terceira frase), as consequências, as contribuições e o que se pretende alcançar ao final deste trabalho. Seguem dois conjuntos de dicas para melhorar ainda mais o seu resumo.

Dica II: O conteúdo específico do seu trabalho começa com o resumo. Lembre-se que o resumo especifica melhor o contexto no qual as palavras-chave do título são trabalhadas. Além disso, considere os seguintes aspectos na hora de escrever o resumo: (1) é composto de um ou mais parágrafos ocupando no máximo uma página/uma página e meia; (2) é uma propaganda do texto; (3) sempre menciona informações ou conclusões que estão no texto; (4) não apresenta referências bibliográficas (exceto em ocasiões raras, como modificações a um método publicado previamente); e (5) pode ser o último a ser escrito, porque ao final do trabalho é quando se tem ideia

melhor do todo e, como o trabalho já está completo, é mais fácil de resumir suas ideias principais.

Dica III: Geralmente o resumo é um parágrafo único, eventualmente pode ser dividido em dois. Muito importante: resumo não contém referência bibliográficas por definição. Para apresentar sugestão de resumos de monografias, seria necessário ocupar várias páginas. Seguem então duas sugestões práticas sobre o conteúdo do resumo: (1) escopo do trabalho, principais objetivos e principal resultado ou conclusão; e (2) contexto geral e específico, questão/problema sendo investigado (propósito do trabalho), estado-da-arte (por que precisa de uma solução nova/melhor), solução (nome da proposta, metodologia básica sem detalhes, quais características respondem as questões iniciais), e interpretação dos resultados, conclusões. A seção de Abstract (a seguir) possui um exemplo de resumo tirado de um artigo de conferência com essas partes identificadas em vermelho.

Palavras-chave: Modelo de texto, PPGCC/UFMG, Latex.

Abstract

Early recognition of risky trajectories during an Intensive Care Unit (ICU) stay is one of the key steps towards improving patient survival. Learning such trajectories from epidemiological and physiological parameters that are continuously measured during an ICU stay requires learning time-series features that are robust and discriminative across diverse patient populations. Patients within different ICU populations (or domains) may vary by age, conditions and interventions, and models built using patient data from a particular ICU domain perform poorly in other domains because the features used to train such models have different distributions across the groups. In this paper, we propose a deep model to capture and transfer complex spatial and temporal features from multivariate time-series ICU data. Features are captured in a way that the state of the patient in a certain time depends on the previous state. This enables dynamically predictions and creates a mortality risk space, allowing to easily describe the risk of the patient at a particular time. A comprehensive cross-ICU experiment with diverse domains reveals that our model outperforms all considered baselines. Gains in terms of AUC range from 4% to 8% for early predictions, when compared with a recent state-of-the-art representative for ICU mortality prediction. Our experiments also show the importance of learning models that are specific for each ICU domain. In particular, models for the Cardiac domain achieve AUC numbers as high as 0.87, showing excellent clinical utility for early mortality prediction.

Resumo Estendido

Geralmente, a monografia escrita em inglês (seja dissertação de mestrado ou tese de doutorado) contém um resumo estendido em português. Este resumo é **recomendado**, ou seja, opcional. Nesse caso, o resumo pode ocupar de duas a oito páginas (dependendo do conteúdo) e apresenta uma visão mais detalhada sobre as contribuições do texto. Por exemplo, pode-se escrever um ou dois parágrafos para cada um dos seguintes itens: contexto, problemas em aberto, visão geral da solução, descrição da contribuição1, contribuição2, contribuição3, discussão sobre a validação realizada (provas, experimentos, etc), conclusão e trabalhos futuros.

Palavras-chave: PalavraChave1, PalavraChave2, PalavraChave3.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

The Intensive Care Unit (ICU) is a department of a hospital in which patients who are dangerously ill are kept under constant observation. Usually, those units have a single specialization, such as cardiac surgery or pediatric diseases, and deal with patients who have high mortality risk, and therefore need to be constantly monitorated, by applying equipments that can generate patient status on real time (e.g., a heart beat monitor), or by exam results requested by ICU doctors, also called intensivists.

According to **?**, the estimated ICU length of stay on the United States is of 3.8 days, and the leading causes of death in the ICU are multi-organ failure, cardiovascular failure, and sepsis **?**. Multi-organ failure has a mortality rate of up to 15-28%, and severe respiratory failure has a mortality rate ranging from 20% to 50%, while sepsis, has a mortality rate of up to 45%. Overall, mortality rates in patients admitted to adult ICUs average 10% to 29%, depending on age and severity of illness.

Data from patients in the Intensive Care Unit are extensive, complex, and often produced at a rate far greater than intensivists can absorb. As a consequence, monitoring ICU patients is becoming increasingly complicated, and systems that learn from ICU data in order to alert clinicians to the current and future risks of a patient are playing a significant role in the decision making process [**?**]. However, one of the main barriers in the deployment of these learning systems is the lack of generalization of results, i.e., the learning performance achieved in controlled environments often drops when the models are tested with different patient populations and conditions [Alemayehu and Warner, 2004; Seshamani and Gray, 2004].

This behaviour could be explained by the difference between patient data and the reasons that eventually leads to death, observed not only in different hospital domains, but also inside the same hospital, from an ICU to another, and even inside a single ICU. Each patient is different, and although there might be some similarity between them,

other factors contribute to the outcome variance, such as the designated professional staff, treatments applied and the ICU enviroment.

In this work, we explore domain adaptation to improve the performance of systems evaluated with mismatched training and testing conditions. We propose deep models that extract the domain-shared and the domain-specific latent features. This enables us to learn multiple models that are specific to each ICU domain, improving prediction accuracy over diverse patient populations. For this, we discuss several domain adaptation approaches that differ in terms of the choice of which layers to freeze or tune.

The proposed models are composed of convolutional and recurrent components. They capture local physiological interactions (e.g., heart rate, creatinine, systolic blood pressure) at the lower level using a Convolutional Neural Network (CNN) and extracts the long range dependencies based on convoluted physiological signals at the higher level using a Long Short-Term Memory network (LSTM). Thus, our model exploits spatial and temporal information within vital signals and laboratorial findings to dynamically predict patient outcomes, i.e., the CNN component extracts spatial features of varying abstract levels and the LSTM component ingests a sequence of spatial features to generate temporally dynamic predictions for patient mortality. As a result, our models perform predictions that are based on information continuously collected over time and that can be updated (dynamically) as soon as new information becomes available.

We also propose a novel neural network layer, which we called Swich. This layer is able to create internal dense representations of the patient's features, and then use those representations to modify the features themselves. With this modifications, our layer is able to find different distributions along the dataset, identify which distribution the patient belongs to, and use that information to improve the prediction.

As a consequence, the learned representations along with the predictions for a specific patient during the ICU stay form the corresponding patient trajectory, and thus a mortality risk space can be obtained from a set of past patient trajectories. The fundamental benefit of analyzing future patient trajectories in the mortality risk space is the focus on dynamics, emphasizing the proximity to risky regions of the space and the speed in which the patient condition changes. Therefore, the mortality risk space enables clinicians to track risky trends and to gain more insight into their treatment decisions or interventions.

The data used to validate our hypothesis was drawn from the PhysioNet 2012 dataset [Silva et al., 2012], an open competition that aimed to create new methods for patient-specific prediction of in-hospital mortality. The dataset is made by the records

of 4000 patients who have stayed at least 48 hours on one of four ICU, being those Coronary Care Unit, Cardiac Surgery Recovery Unit, Medical ICU and Surgical ICU.

In this work we elucidate the extent to which ICU mortality prediction may benefit from domain adaptation. In summary, our main contributions are:

- While the combination of convolutional and recurrent structures has been investigated in prior work other than mortality prediction [**?**], this architecture is a proper choice here because it offers a complementary spatial-temporal perspective of the patient condition. As a result, predictions based on information that are continuously collected over time can be dynamically updated as soon as new information becomes available.

- We propose deep models for ICU mortality prediction. Our models are composed of convolutional and recurrent layers, thus offering a complementary spatial-temporal perspective of the patient condition. As a result, our models perform predictions that are based on information continuously collected over time and that can be updated (dynamically) as soon as new information becomes available.

- We propose a novel type of layer that not only improved the results of mortality prediction on ICU, but can also be used in many other domains, since it fits on any neural network architecture.

- We show that patients within different ICU domains form sub-populations with different marginal distributions over their feature spaces. Therefore, we propose to learn specific models for different ICU domains that are trained using different feature transference approaches, instead of learning a single model for different ICU domains. We show that the effectiveness of different feature transference approaches varies greatly depending on the factors that define the target domain.

- We conducted rigorous experiments using the PhysioNet 2012 dataset which comprises data from four different ICU domains. We show that multi-domain ICU data used for adaptation can significantly improve the effectiveness of the final model. Gains in terms of AUC range from 4% to 8% for early predictions, i.e., predictions based on data acquired during the first $5 - 20$ hours after admission. Gains range from 2% to 4% for predictions within the first 48 hours after admission.

- We show that the patient representations along with the predictions provided by our models are meaningful in the sense that they form trajectories in a mortality risk space. Dynamics within this space can be very discriminative, enabling

clinicians to track risky trends and to gain more insight into their treatment decisions or interventions.

Furthermore, this work presents a range of published works about ICU mortality prediction techniques and neural networks on Chapter 2, a detailed description of our proceedings to develop the architecture, layers and visualization, and any other relevant experiments on Chapter 3, and all results obtained on Chapter 4. Finally, Chapter 5 brings our discussion of the subject and final considerations.

Chapter 2

Related Work

In this chapter we bring some of the most relevant researches that guided our work, exposing the methodology used by the authors and how it is correlated to ours. Research on predicting ICU mortality is of great academic interest in medicine [Cai et al., 2016; Tabak et al., 2014; Wu et al., 2017] and in clinical machine learning [Ghassemi et al., 2014; Johnson et al., 2016; Luo et al., 2016; Nori et al., 2017], since a good model can help doctors to save lives. A number of researchers have investigated how to correlate ICU data with patient outcomes. In one of the first studies [Patel et al., 2009], a group of computer scientists, chemists, geneticists, and philosophers of science was brought together to develop a model that could identify parameters in patient data that correlate with its outcome.

2.1 Mortality Prediction

The PhysioNet ICU Mortality Challenge 2012 [Silva et al., 2012] provided benchmark data that incorporate evolving clinical data for ICU mortality prediction. As Johnson et al. [2014] reported, this benchmark data fostered the development of new approaches, leading to up to 170% improvement over traditional risk scoring systems that do not incorporate such clinical data currently used in ICUs [Gall et al., 1993]. In what follows, we discuss previous work in contrast with ours.

Most of the current work uses the PhysioNet ICU Mortality Challenge 2012 data. The most effective approaches are based on learning discriminative classifiers for specific sub-populations.

Citi and Barbieri [2012] proposed a robust SVM classifier,

while Bera and Nayak [2012]; Hamilton and Hamilton [2012] proposed a logistic regression classifier.

Vairavan et al. [2012] also employed logistic regression classifiers, but coupled them with Hidden Markov Models in order to model time-series data.

Shallow neural networks were evaluated in [Xia et al., 2012],

while a tree-based Bayesian ensemble classifier was evaluated in [Johnson et al., 2012].

Krajnak et al. [2012] employed fuzzy rule-based systems for mortality prediction,

and McMillan et al. [2012] proposed an approach that identifies and integrates information in motifs that are statistically over- or under-represented in ICU time series of patients.

More recently, Lee and Horvitz [2017] proposed a Markov model that accumulates mortality probabilities. Likewise, Barajas and Akella [2015] proposed an approach that models the mortality probability as a latent state that evolves over time. Gong et al. [2015] proposed an approach to address the problem of small data using transfer learning in the context of developing risk models for cardiac surgeries. They explored ways to build surgery-specific and hospital-specific models using information from other kinds of surgeries and other hospitals. Their approach is based on weighting examples according to their similarity to the target task training examples. The three aforementioned works are considered as baselines and compared with our approach.

Following Gong et al. [2015], in this work we use feature transference, but in a quite different way, as follows: (i) instead of applying instance weighting, we employed a deep model that transfers domain-shared features; (ii) we studied a broader scenario that includes diverse ICU domains; and (iii) our models employ temporal feature extraction, being able to predict patient outcomes dynamically.

2.2 ICU Domains and Sub-Populations

Imbalanced data [Bhattacharya et al., 2017], sub-populations of patients with different marginal distributions over their feature spaces [Nori et al., 2017], and sparse data acquired from heterogeneous sources [Ghassemi et al., 2015; Huddar et al., 2016] are issues that pose significant challenges for ICU mortality prediction.

Gong et al. [2017] discussed problems due to the lack of consistency in how semantically equivalent information is encoded in different ICU databases. Bhattacharya et al. [2017] discussed the problem of imbalanced ICU data, which occurs when one of the possible patient outcomes is significantly under-represented in the data. Further, since features are often imbalanced, some ICU domains have a significantly larger number of observations than others (e.g., respiratory failure in adults vs. children). In a

recent work, Bonomi and Jiang [2017] proposed a mortality study based on the notion of burstiness, where high values of burstiness in time-series ICU data may relate to possible complications in the patient's medical condition and hence provide indications on the mortality.

While most studies on mortality prediction for ICU patients have assumed that one common risk model could be developed and applied to all the patients, Nori et al. [2017] advocated that this might fail to capture the diversity of ICU patients. As shown by Alemayehu and Warner [2004], as well as by Seshamani and Gray [2004], models built using patient data from particular age groups perform poorly on other age groups because the features used to train the models have different distributions across the groups.

2.3 Distribution-Aware Neural Network

2.4 Present Work

None of the aforementioned approaches attempted to perform ICU domain adaptation, which is a core focus of our work. There is often a mismatch between different ICU domains or patient sub-populations, and domain adaptation seems to be a natural solution for learning more robust models, as different ICU domains share features that exhibit different distributions. While data in different ICU domains may vary, there are potentially shared or local invariant features that shape patients in different ICU domains.

Other focus of our work is to capture spatial and temporal features from time-series ICU data. Features are captured in a way that the state of the patient in a certain time depends on the previous state. This forms a mortality risk space, and trajectories in this space allow to easily describe the state of the patient at a particular time, helping intensivists to estimate the patient progress from the current patient state.

Chapter 3

Methodology

In this chapter we will discuss the methods and guidelines used to create our predictive models and their applications. One can define the task of predicting patient outcomes over time from ICU data is defined as follows. Each ICU patient can be represented by their physiological observations at a given time, such as heart rate, temperature, blood pressure, and others. Since the patient is continuously observed, his representation is an ordered set of multiple discrete time observations.

We then have as input the *training set*, which consists of a sequence of observations of the form $< A_t, o >$, where A_t is a vector of values corresponding to physiological parameters associated with a patient at time t, and o is the outcome for the patient (i.e., whether or not the patient survived the hospitalization). The training set is used to construct a model that relates features within the sequence of observations to the patient outcome. The *test set* consists of a sequence of observations $< A_t, ? >$ for which only the physiological parameters for the patient until time t are available, while the corresponding patient outcome is unknown. The model learned from the training set is used to produce predictions of the outcome for patients in the test set.

The full set of data is split into 5 equally large stratified fold, used to perform a 5-fold cross validation. Each fold is divided in training and test set. Early stopping [**?**] was also applied, so the training set is divided itself in the actual training set, which is used to build the model, and a validation set, used to prevent the neural network from overfitting.

The task of predicting patients outcomes in the ICU has two important requirements:

- It is a domain-specific problem, i.e., a prediction model learned from a subpopulation (or ICU domain) is likely to fail when tested against data from other

population [Seshamani and Gray, 2004]. Feature transferability is thus an appealing way to provide robustness to prediction models.

- It is a time-sensitive problem, i.e., accurately predicting patient outcomes as early as possible may lead to earlier diagnosis and more effective therapy.

We divide this chapter in two sections. The first shall explicit our methods to achieve the best mortality prediction as possible through a model, which is built from multi-domain ICU time-series data and is designed to provide dynamically-updated estimates of patient mortality, while the second section will show some practical applications of those models.

3.1 Mortality Prediction

Here we discuss how to manipulate the patient data into a format that can be dynamically consumed by a model, along with the main blocks that constructs this model. We also present our novel neural layer and explain how to apply domain adaptation to help solve the multi-distribution and generalization problem.

Our goal is to analyze patient data at each moment and evaluate the probability of this patient not surviving the treatment, simulating a real time medical expert with full attention to each patient. In order to do so, we need a well defined data structure that consists of fixed time steps and a invariable set of patient's signs at each time step.

3.1.1 Data and Domains

We use the publicly available dataset of multivariate clinical time-series of 4,000 patients from the PhysioNet 2012 challenge [Silva et al., 2012]. The data for each patient includes age, gender, height, weight and 37 time-stamped physiological parameters measured during the first 48 hours of ICU stay. All those parameters are listed in Table 3.1. Patient outcomes, including mortality, are available. Note that some of those features are measured a lot more frequently than others, as each feature as its own measurement difficulty. For instance, it is quite simple to measure someone's heart rate or temperature, but a lot harder and costly to measure his Cholesterol.

In order to make the data equally formated for each patient, we first propagate measurements forward (or backward) in time to fill gaps, so observations that are less frequent are considered constant until new measurement. We then resample the time series on an hourly basis, averaging the values observed on each hour for each patient

feature, so now our patient can be represented by the mean value for each physiological observation on each hour during its ICU stay. Finally, we scale each variable to fall into the $[0, 1]$ interval. All patients are 16 years or older and had ICU stays of at least 48 hours. In contrast to Bhattacharya et al. [2017], we did not perform feature selection and, thus used the entire feature-set in all experiments.

Table 3.1 shows the average physiological data for patients in each ICU domain. The dataset also specifies the ICU domain to which the patient has been admitted: Cardiac Surgery, Coronary Care Unit, Medical and Surgical. It is possible to conclude that physiological data differ greatly between patients admitted to different ICU domains, but some features also have a common range across one or more ICU, thus reinforcing our main hypothesis that transfer learning can indeed be applied to improve mortality prediction.

Figure 3.1 shows the relative frequency in which physiological parameters are measured within each ICU domain. As can be seen, some ICU domains have a significantly larger number of observations than others (e.g., $PaCO_2$ and PaO_2 are much more frequently measured in the Cardiac ICU, while TroponinT is much more frequently measured in the Coronary ICU).

3.1.2 Network Architecture

In this section we introduce the deep model architectures we evaluated to perform mortality prediction, eventually selecting those with best results. We compared several architectures, from using only a Convolutional Neural Network [Krizhevsky et al., 2012] or recurrent layer, to combining both, and adding intermediate layers, such as Dropout layers. Convolutional and recurrent components offer a complementary perspective of the patient condition, as follows: the convolutional layer emphasizes the local interaction between physiological parameters, while the recurrent layer is designed to capture long range information and forget unimportant local information.

Our first model was a single recurrent layer, more specifically a LSTM layer that sought to capture the tendencies between the patient states each time. Long-Short Term Memories are largely used in time-dependent problems, because of its great ability to deal with series data, so its a natural choice in this case. As our patient can be understood as a series of points moving in a high dimensional space, the LSTM will be able to create a representation based on this movement, which is then used to perform prediction, although it may overlook the feature codependency in a single time step.

We also tried a Convolution-only model, that captured the relationship between

Table 3.1. Average patient physiological data. Mean, first and third quartiles within each physiological parameter. Mortality rate is concentrated in the Medical ICU (49.6% of all the deaths).

	Cardiac	Coronary	Medical	Surgical
N	874	577	1,481	1,067
Age	67.91 (56−79)	69.22 (59−81)	62.83 (51−78)	60.50 (48−76)
Male	530 (60.6%)	333 (57.7%)	753 (50.8%)	630 (59.0%)
Mortality Rate	4.9% (7.8%)	14.0% (14.6%)	18.6% (49.6%)	14.5% (28.0%)
Albumin (g/dL)	2.92 (2.4−3.5)	3.31 (2.9−3.6)	2.92 (2.5−3.3)	2.99 (2.5−3.5)
Alkaline phosphatase (IU/L)	74.93 (46−83)	92.44 (59−102)	126.15 (64−138)	91.43 (52−96)
Alanine transaminase (IU/L)	89.16 (18−45)	128.28 (19−78)	164.87 (16−61)	191.52 (17−84)
Bilirubin (mg/dL)	1.01 (0.4−1.1)	0.87 (0.4−0.9)	2.44 (0.4−1.6)	1.85 (0.5−1.5)
Blood urea nitrogen (mg/dL)	18.76 (12−21)	29.92 (16−36)	32.59 (14−42)	20.36 (11−24)
Cholesterol (mg/dL)	150.14 (114−174)	163.59 (134−189)	141.04 (111−169)	157.87 (122−184)
Creatinine (mg/dL)	1.04 (0.7−1.1)	1.58 (0.8−1.6)	1.64 (0.7−1.7)	1.12 (0.7−1.1)
Invasive diast. press. (mmHg)	58.85 (51−66)	62.65 (53−74)	54.97 (48−70)	59.65 (52−72)
Fractional inspired O2	0.91 (1.0−1.0)	0.82 (0.5−1.0)	0.72 (0.5−1.0)	0.72 (0.5−1.0)
Serum glucose (mg/dL)	129.28 (103−145)	165.74 (114−191)	155.02 (104−175)	148.85 (114−167)
Serum bicarbonate (mmol/L)	23.41 (22−25)	23.31 (21−26)	22.74 (19−26)	23.44 (21−26)
Hematocrit (%)	29.32 (25.3−32.8)	34.48 (30.7−37.8)	31.82 (27.9−36)	33.01 (29.1−36.8)
Heart rate (bpm)	85.43 (79−91)	84.32 (69−97)	95.61 (80−110)	87.83 (74−100)
Serum potassium (mEq/L)	4.49 (4−4.7)	4.28 (3.8−4.5)	4.19 (3.6−4.5)	4.07 (3.6−4.3)
Lactate (mmol/L)	2.76 (1.5−3.3)	2.76 (1.4−3)	2.58 (1.3−2.8)	2.65 (1.3−3.1)
Serum magnesium (mmol/L)	2.22 (1.8−2.4)	1.90 (1.7−2.1)	1.95 (1.6−2.1)	1.80 (1.5−2)
Invasive mean press. (mmHg)	78.86 (69−86)	86.14 (73−99)	86.58 (68−96)	87.13 (73−98)
Serum sodium (mEq/L)	138.42 (136−140)	137.82 (135−140)	138.96 (136−142)	139.33 (137−142)
Non-invasive diast. press. (mmHg)	52.21 (44−59)	61.15 (49−72)	62.03 (50−72)	62.42 (52−73)
Non-invasive mean press. (mmHg)	71.53 (62−79)	78.93 (67−89)	80.55 (68−91)	82.78 (71−94)
Non-invasive syst. press. (mmHg)	110.88 (96−125)	117.46 (101−134)	121.78 (104−138)	126.72 (108−145)
Partial press. of art. CO2 (mmHg)	41.20 (36−45)	40.61 (35−45)	42.50 (34−48)	41.01 (35−45)
Partial press. of art. O2 (mmHg)	295.46 (218−387)	181.58 (89−248)	147.68 (78−185)	188.24 (101−250)
Arterial pH (0-14)	7.39 (7.35−7.44)	7.84 (7.31−7.43)	7.44 (7.3−7.42)	7.46 (7.32−7.43)
Platelets (cells/nL)	170.36 (117−208)	241.44 (181−283)	230.89 (143−287)	219.19 (150−268)
Respiration rate (bpm)	17.55 (14−20)	19.74 (16−23)	21.10 (17−24)	18.95 (16−21)
O2 saturation in hemoglobin (%)	97.48 (97−98)	96.25 (96−98)	94.84 (94−98)	96.99 (97−98)
Invasive systolic press. (mmHg)	117.16 (105−127)	117.65 (100−139)	107.45 (95−137)	123.33 (108−148)
Temperature (°C)	35.57 (35.5−36.6)	36.38 (36−37.1)	36.77 (36.2−37.4)	36.51 (36.1−37.4)
Troponin-I (μg/L)	6.77 (0.8−10.1)	10.05 (0.8−12.4)	5.59 (0.8−7)	7.02 (0.4−6.7)
Troponin-T (μg/L)	1.51 (0.04−0.59)	2.78 (0.17−2.8)	0.33 (0.04−0.25)	0.22 (0.03−0.14)
Urine output (mL)	497.92 (120−615)	365.62 (100−500)	255.39 (70−325)	389.29 (100−500)
White blood cell (cells/nL)	12.98 (9.2−15.5)	12.31 (8.5−14.3)	13.33 (7.8−17)	12.37 (8.4−15.1)

features on a single time period, and then treated all time periods as one. This approach is not as intuitive as the later, but shows a surprisingly better performance. Here we create several filters that combine the patient observations only locally, this is, it does not combine features across time. Alongside a Max Pooling layer, the model extracts some information about risk regions in the patient's feature space over this local combination, and finally, all this information is flattened into a single vector that is the patients representation, then used to predict his outcome. Although this method does not explicit create a representation based on the patient time series, by creating a flatting representation with all the time steps we are also encapsulating time information.

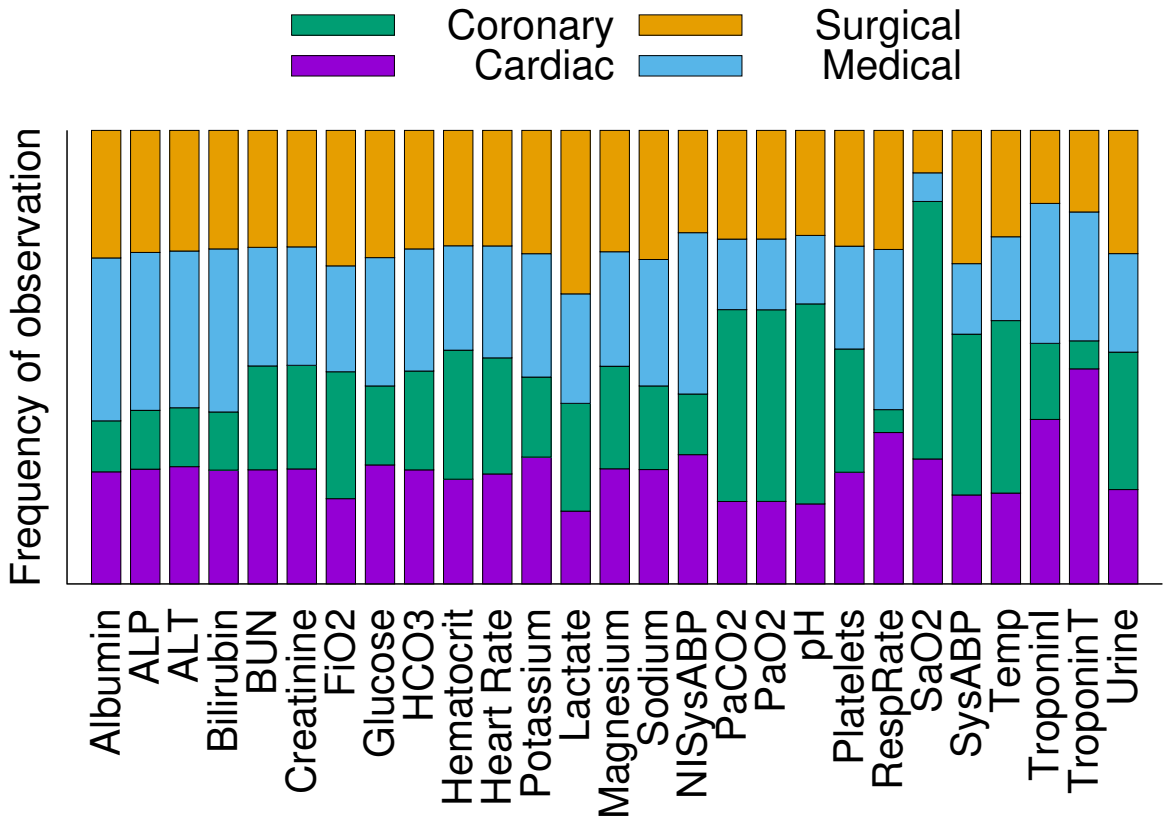

Figure 3.1. (Color online) Relative frequency in which physiological parameters are measured in different ICU domains.

Finally, we have the model that employs a CNN layer followed by a max-pooling layer, thus extracting correlations between physiological parameters measured in the same time period and exploring their simultaneous effects. For instance, it may find that if both temperature and heart rate are high on the same time period, the odds of survival decrease. In a complementary way, the recurrent layer (LSTM) is devoted to learn how changes in observations for a patient affect the corresponding outcome. Intuitively, the recurrent layer captures temporal dependencies, enabling the estimation of patient progress from the current patient state. For instance, if the heart rate was low at the beginning of the stay and then become very high, then the odds of survival decrease. Finally, a dense layer takes the output of the recurrent layer and predicts the patient outcome. This model is shown in Figure **??**.

Naturally, this is a high variance data, since all features come from measuring something as complex as a human being, which leads to the model quickly overfitting the training set. In order to prevent this, we applied several dropout layers **?**, specif-

ically after the input, max pooling and LSTM layers. A dropout layer will choose a random set of neuron each batch, and disable them during training. This will make the other neurons (that were not disabled) generalize more, simulating the effect of training multiple smaller networks and averaging them during test. We drop from 20 to 30 percent of all neurons on each layer. We also apply L2 regularization **?** to the LSTM inner cell neurons and the fully connected layer at the end of the model. This regularization will force each neuron to keep their activation weights low, thus generalizing more. Our loss function was binary cross-entropy, because of its good performance for classification problems with two classes. This loss function is given in the following formula:

$$\ell(\lambda) = -\frac{1}{n}\sum_{i=1}^{n}[y_i log(p_i) + (1-y_i) log(1-p_i)] + \lambda \sum_{j=1}^{k} w_j^2$$

where λ is the set of weights, n is the number of samples in the batch, y_i is the true output of the ith patient, p_i is the predicted output for the ith patient, and k is the number of neurons to regularize.

The final component tested in the neural architecture was the activation function of each layer. For the other layers, a few activation functions were tried:

Linear $f(x) = x$

Sigmoid $f(x) = \frac{1}{1+e^{-x}}$

Tanh $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

Rectifier Linear Unit (ReLU) $f(x) = max(0, x)$

Scaled Exponential Linear Unit (SELU) $f(x) = \lambda x$, if x > 0, $\alpha e^x - \alpha$, otherwise

being x the neuron output. Since this is a binary classification problem, ranged from 0 to 1, we chose to sustain a sigmoid activation on the output layer. This will scale any output to the (0, 1) interval.

In order to choose a set of model parameters that perform well for this task, a hand tuning method was applied. This means that we manually executed tests with different parameter sets and chose the set that performed, making adjustments based on the output of previous executions. Those parameters include the number of neurons on each layer, activation functions (as mentioned above), regularization type and amount, dropout percentage, along with some layer-specific parameters, such as kernel size, for convolutional layers and pool size for max pooling layers.

In summary, our models works by passing each observation through a spatial feature extractor and then the sequence model captures how the extracted spatial

features are associated with patient outcomes over time. Also, dropout operation is performed after each layer of the network.

As not all the descriptors and time-series were available for all records, we had to deal with the problem of missing values. If one variable (either a descriptor or a time-series) was never recorded for a given record, we used the approach called "imputation" and replaced its feature/s with value zero. Because of the normalization step, this approximately corresponds to replacing the missing raw variable with a measure of central tendency, which corresponds to the arithmetic mean for Gaussian-distributed variables and to the geometric mean for log-normal ones. In some cases, the time-series measurement were taken only in the first 24 h or only during the next 24 h. In this case, replacing with zero all the features related to the period with missing measurements could possibly create a non-existing improvement or deterioration trend. Instead, we duplicate the values from the available period, assuming stationarity conditions as default in absence of further measurements.

3.1.3 Switch Layer

3.1.4 Feature Transferability

Our goal is to train multi-domain models to predict patient outcomes over time, which is based on patient observations associated with multiple ICU domains. Although patients associated with a given ICU domain may be better represented by specific features, there still exist some common features that permeate all other ICU domains.

The main intuition that we exploit for feature transferability is that the features must eventually transition from general to specific along our model, and feature transferability drops significantly in higher layers with increasing domain discrepancy [Yosinski et al., 2014]. In other words, the features computed in higher layers must depend strongly on a specific domain, and prediction effectiveness suffers if this domain is discrepant from the target domain. Since we are dealing with many domains simultaneously, we tested multiple transference approaches, which are detailed as follows:

A1: No layer is kept frozen during fine-tuning, i.e., errors are back-propagated through the entire network during fine-tuning.

A2: Only the convolutional layer is kept frozen during fine-tuning.

A3: Convolutional and LSTM layers are kept frozen during fine-tuning, i.e., errors are back-propagated only thought the fully-connected layers during fine-tuning.

A4: Only the convolutional layer is kept frozen during fine-tuning and other layers have their weights randomly initialized for fine-tuning.

A5: Convolutional and LSTM layers are kept frozen during fine-tuning and weights in fully-connected layers are randomly initialized for fine-tuning.

3.2 Application

Here we will present some ways of how a mortality prediction model can improve the work of medical doctors, helping them in their decisions.

3.3 Experiments

In this section, we present the data we used to evaluate our multi-domain model for mortality prediction over time. Then, we discuss our evaluation procedure and report the results of our multi-domain model. In particular, our experiments aim to answer the following research questions:

RQ1: Does domain adaptation improve mortality prediction? Do models that are specific to each ICU domain improve the state-of-the-art models for mortality prediction?

RQ2: Which feature transference approach is more appropriate to each ICU domain?

RQ3: How accurate are dynamic predictions?

RQ4: How meaningful are the mortality risk spaces created from patient trajectories?

3.3.1 Baselines

We considered the following methods in order to provide baseline comparison:

- Traditional classifiers: Support Vector Machines (SVM), Random Forest (RF), Logistic Regression (LR), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and AdaBoost. The main objective of using these baselines is to compare our model with shallow models.
- Training on Target (TT): A CNN−LSTM model is trained using only the target domain data. No source domain data is used. The main objective of using this baseline is to assess the benefits of different feature transference approaches.

- Deep architecture [Che et al., 2015]: A deep network that uses prior-based regularization. The main objective of using this baseline is to compare our model with state-of-the-art results on the PhysioNet data.

3.3.2 Setup

The measure used to evaluate the effectiveness of our models is the standard Area Under the ROC Curve (AUC), as adopted by Che et al. [2015]. Like Johnson et al. [2012], we use five-fold cross validation and relevant hyper-parameters were found using a further internal cross-validation. The results reported are the average of the five runs and are used to assess the overall performance of the models. To ensure the relevance of the results, we assess the statistical significance of our measurements by means of a pairwise t-test [Sakai, 2014] with $p-value \leq 0.05$. Hereinafter, we refer to our model as $CNN-LSTM$.

Chapter 4

Results

The first experiment is devoted to answer RQ1. We present a comparison between shallow and deep models. Table 4.1 shows AUC numbers for predictions performed using information acquired within the first 48 hours after the patient admission. Predictions performed by the baseline models were simply separated according to the ICU domain in which the corresponding patient was admitted, so that we can report AUC numbers for each ICU domain. On the other hand, the CNN−LSTM model employs domain adaptation, and thus is composed of four sub-models that are specific to each of the four domain ICUs. Clearly, domain adaptation improves the accuracy of our models and consistently outperform all baselines considered in this work. Overall, our model shows an AUC number of 0.818, which is considered to provide excellent clinical utility in the field of mortality prediction [Johnson et al., 2014].

The second experiment is concerned with RQ2. We present a comparison between the TT model and models learned following our five feature transference approaches. Table 4.2 shows AUC numbers for predictions performed using information acquired within the first 48 hours after the patient admission. Feature transference is never detrimental when compared with the TT model and they provide substantial gains that are up to 6.7% (Cardiac), 8.3% (Coronary), 8.3% (Medical), and 11.0% (Surgical). These gains seem to be related to the mortality rate associated with each target ICU domain − gains are higher for domains with higher mortality rates.

Finally, we can see from Table 4.2 that the best transference approach varies depending on the target ICU domain. Randomly initializing the weights for fine-tuning does not show to be the best approach, as A4 and A5 were not the best performers for any target ICU domain. It seems that specific temporal patterns play an important role for mortality prediction in the Surgical domain, as A2 was the best transference approach for this domain. For the Medical domain, A3 was the best transference

Table 4.1. AUC numbers for shallow and deep models. Numbers in bold indicate the best models for each ICU domain.

Model	Cardiac	Coronary	Medical	Surgical
AdaBoost	0.572	0.551	0.510	0.531
SVM	0.627	0.572	0.503	0.532
LR	0.629	0.601	0.510	0.517
LDA	0.632	0.602	0.516	0.513
RF	0.610	0.578	0.587	0.623
QDA	0.689	0.668	0.567	0.610
[Che et al., 2015]	0.853	0.802	0.760	0.785
CNN−LSTM	**0.876**	**0.833**	**0.782**	**0.807**

Table 4.2. AUC numbers for different feature transference approaches. Numbers in bold indicate the best transference approach for each target ICU domain.

Target	TT	A1	A2	A3	A4	A5
Cardiac	0.821	**0.876**	0.863	0.864	0.805	0.814
Coronary	0.769	**0.833**	0.812	0.807	0.793	0.784
Medical	0.722	0.749	0.758	**0.782**	0.738	0.741
Surgical	0.727	0.802	**0.807**	0.798	0.756	0.756
Overall	0.760	**0.815**	**0.811**	**0.813**	0.773	0.774

approach, suggesting that spatial and temporal features learned from other domains are already effetive. For the Cardiac and Coronary domains, A1 was the best transference approach, which indicates that specific spatial-temporal features are very important in these domains.

The next set of experiments is devoted to answer RQ3. Figure 4.1 shows AUC numbers obtained with predictions performed using information acquired within the first y hours after the patient admission. As expected, accuracy increases as more information is acquired. From the first 5 to 20 hours, the slopes associated with Cardiac and Coronary domains increase much faster than the slopes associated with Medical and Surgical domains.

Figure 4.2 shows the gains obtained when compared with the work by Che et al. [2015] at different prediction times. The early predictions performed by the CNN−LSTM archtecture are much more accurate than the early predictions performed by Che et al. [2015], particularly in the first hours after the admission. The 10−20 hours period concentrates the more impressive gains, which vary from 4% (Medical) to

almost 8% (Coronary).

The last set of experiments is concerned with RQ4, i.e., to assess how meaningful are the mortality risk spaces. Figure 4.3 shows risk spaces for each ICU domain. These spaces are obtained by gathering patient trajectories, that is, the coordinates (i.e., CNN−LSTM representations) along with the predicted outcome at each time. Risk spaces can also be obtained from raw data and, in this case, the coordinates are simply the entire feature-vector. Risk spaces created from CNN−LSTM representations are much more meaningful than the corresponding spaces obtained from raw data.

Time is also encoded in the risk spaces, and thus we can exploit dynamics, such as the proximity to mortality risky regions or the speed in which the patient condition changes. Figure 4.4 shows such dynamics in mortality risk spaces obtained from CNN−LSTM representations. Dynamics associated with the mortality risk space for the Cardiac and Coronary ICU domains, for instance, are highly discriminative since red and blue curves are separated in the first hours after the patient admission. This may explain the high AUC numbers obtained in these domains. Patients show distinct dynamics, depending on the ICU domain. Patients admitted to the Cardiac and Surgical units, for instance, move much faster than patients admitted to the Coronary and Medical units. Also, the speed increases over time for patients admitted to the Coronary and Medical units.

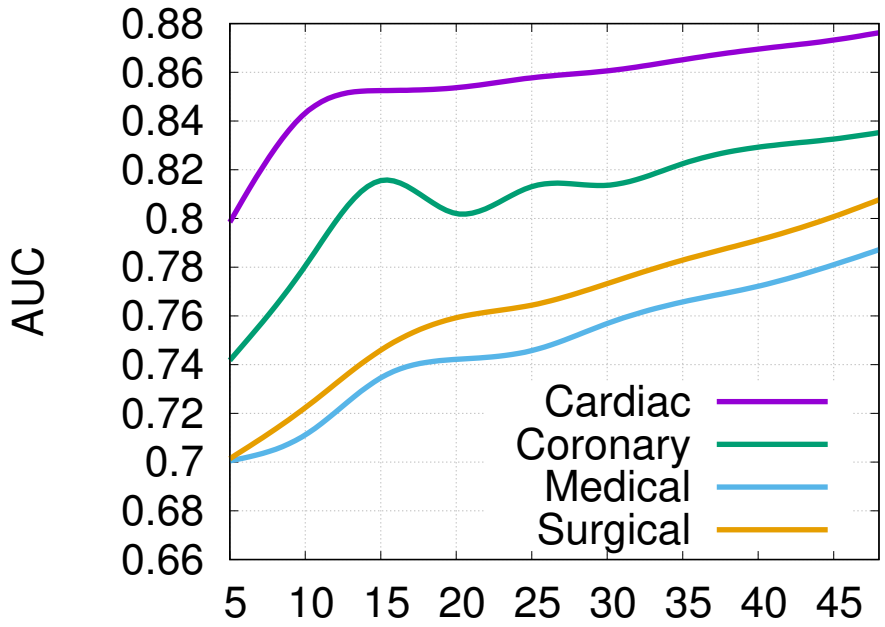

Figure 4.1. (Color online) CNN−LSTM AUC numbers for predictions performed using information within the first y hours after the patient admission ($5 \leq y \leq 48$).

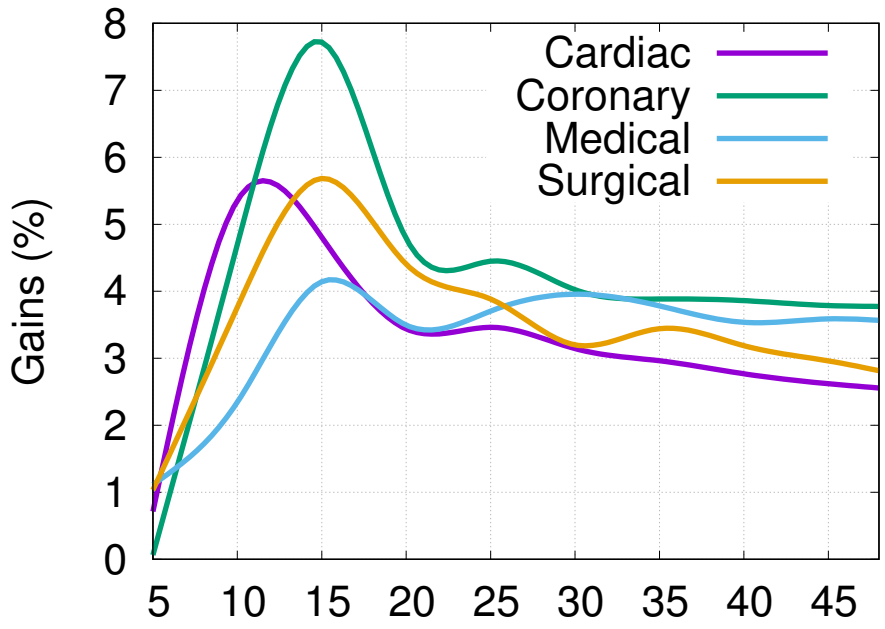

Figure 4.2. (Color online) Gains over [Che et al., 2015] at different prediction times ($5 \leq y \leq 48$).

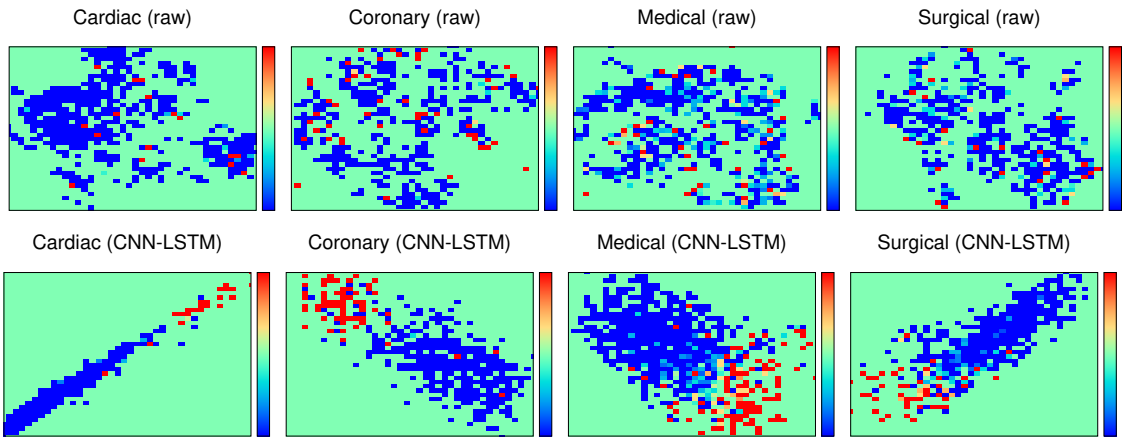

Figure 4.3. (Color online) Mortality risk space for different ICU domains. Regions in red are risky.

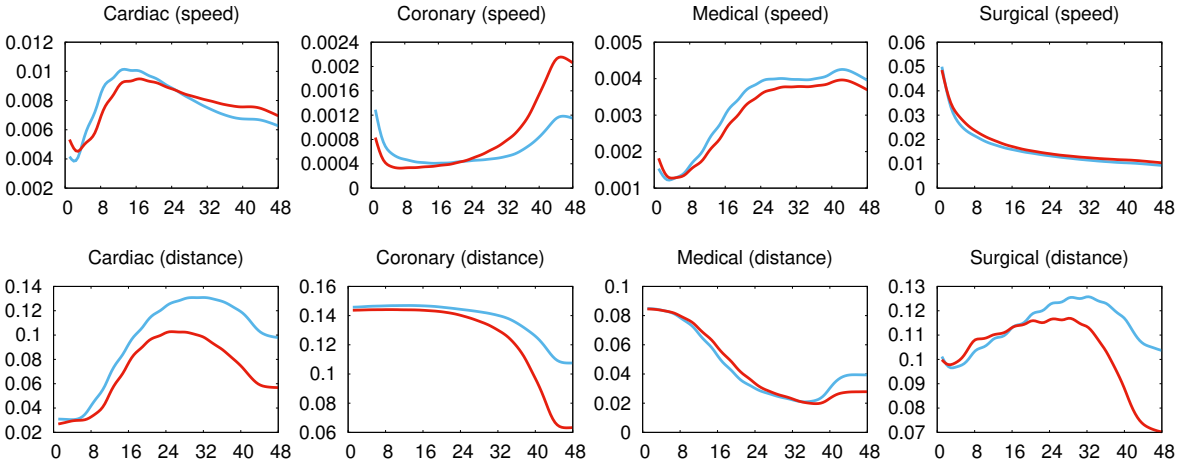

Figure 4.4. (Color online) Dynamics of 48-hour trajectories in different ICU domains. Red curves are computed from trajectories associated with patients that have died. Blue curves are computed from trajectories associated with patients that survived.

Chapter 5

Conclusion

ICU mortality prediction is a domain-specific problem. Thus, a prediction model learned from a sub-population of patients is likely to fail when tested against data from other population. We investigated this problem by considering four sub-populations of patients that were admitted to different ICU domains. We showed that patients within a specific ICU domain are epidemiologicaly and physiologically different from patients within other domains. Nevertheless, patients across ICU domains still share basic characteristics. This motivates us to propose improved mortality prediction models based on domain adaptation. Specifically, our models learn domain invariant representations from time series ICU data while transferring the complex temporal latent dependencies between ICU sub-populations. The proposed models employ spatial and temporal feature extractors, being thus able to perform dynamic predictions during the ICU stay, potentialy leading to earlier diagnosis and a more effective therapy. Finally, our models produce a mortality risk space, and the dynamics associated with patient trajectories are meaningful and can be very discriminative, enabling clinicians to track risky trends and to gain more insight into their treatment decisions or interventions. Our models provide impressive gains (4% to 8%) for early predictions, i.e., predictions within the first $5-20$ hour period after admission. Significant gains (2% to 4%) are also observed for predictions performed based on information acquired during the first 48 hours after admission.

Bibliography

Alemayehu, B. and Warner, K. (2004). The lifetime distribution of health care costs. *Health Services Research*, 3(39):627--642.

Arantes, A. S., Vieira, M. R., Jr., C. T., and Traina, A. J. M. (2003). Operadores de Seleção por Similaridade para Sistemas de Gerenciamento de Bases de Dados Relacionais. In *Proceedings of the Brazilian Symposium on Databases (SBBD)*, pages 341–355.

Barajas, K. C. and Akella, R. (2015). Dynamically modeling patient's health state from electronic medical records: A time series approach. In *Proc. of the 21st ACM SIGKDD*, pages 69--78.

Bera, D. and Nayak, M. (2012). Mortality risk assessment for ICU patients using logistic regression. *Computing in Cardiology*, pages 493--496.

Bhattacharya, S., Rajan, V., and Shrivastava, H. (2017). ICU mortality prediction: A classification algorithm for imbalanced datasets. In *Proc. of the 31st AAAI*, pages 1288--1294.

Bonomi, L. and Jiang, X. (2017). A mortality study for ICU patients using bursty medical events. In *Proc. of the 33rd IEEE ICDE*, pages 1533--1540.

Booth, W. C., Colomb, G. G., and Williams, J. M. (2008). *The Craft of Research.* University Of Chicago Press.

Braganholo, V. P., Davidson, S. B., Heuser, C. A., bla bla, cla cla, and dal dal (2005). Updating Relations through XML Views. In *Anais do Congresso da Sociedade Brasileira de Computação - Concurso de Teses e Dissertações (CTD)*, pages 95–103.

Cai, X., Concha, O., Coiera, E., Martín-Sánchez, F., Day, R., Roffe, D., and Gallego, B. (2016). Real-time prediction of mortality, readmission, and length of stay using electronic health record data. *JAMIA*, 23(3):553--561.

Che, Z., Kale, D., Li, W., Bahadori, M., and Liu, Y. (2015). Deep computational phenotyping. In *Proc. of the 21st ACM SIGKDD*, pages 507--516.

Citi, L. and Barbieri, R. (2012). Physionet 2012 challenge: Predicting mortality of ICU patients using a cascaded SVM-GLM paradigm. *Computing in Cardiology*, pages 257--260.

Cordeiro, R. L. F. (2011). *Data mining in large sets of complex data.* PhD thesis, Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional. Instituto de Ciências Matemáticas e de Computação, USP São Carlos.

da Camara Neto, V. (2011). A classe ppgccufmg. http://vilarneto.com/ppgccufmg.

dos Santos, R. O., de Sá Mesquita, F., da Silva, A. S., and Vilarinho, E. C. C. (2006). Extração de Dados e Metadados em Textos Semi-estruturados usando HMMs. In *Proceedings of the Brazilian Symposium on Databases (SBBD)*, pages 117–131.

Frankland, P. and Weyuker, E. (1993). Provable improvements on branch testing. *IEEE Transactions on Software Engineering (TSE)*, 19(10):962 –975.

Gall, J.-R. L., Lemeshow, S., and Saulnier, F. (1993). A new simplified acute physiology score (SAPS II) based on a european/north american multicenter study. *JAMA, Journal of the American Medical Association*, pages 2957--2963.

Ghassemi, M., Naumann, T., Doshi-Velez, F., Brimmer, N., Joshi, R., Rumshisky, A., and Szolovits, P. (2014). Unfolding physiological state: mortality modelling in intensive care units. In *Proc. of the 20th ACM SIGKDD*, pages 75--84.

Ghassemi, M., Pimentel, M., Naumann, T., Brennan, T., Clifton, D., Szolovits, P., and Feng, M. (2015). A multivariate time-series modeling approach to severity of illness assessment and forecasting in ICU with sparse, heterogeneous clinical data. In *Proc. of the 29th AAAI*, pages 446--453.

Gong, J., Naumann, T., Szolovits, P., and Guttag, J. (2017). Predicting clinical outcomes across changing electronic health record systems. In *Proc. of the 23rd ACM SIGKDD*, pages 1497--1505.

Gong, J., Sundt, T., Rawn, J., and Guttag, J. (2015). Instance weighting for patient-specific risk stratification models. In *Proc. of the 21st ACM SIGKDD*, pages 369--378.

Graff, G. and Birkenstein, C. (2009). *"They Say / I Say": The Moves That Matter in Academic Writing.* W. W. Norton & Company.

Hamilton, S. and Hamilton, J. (2012). Predicting in-hospital death and mortality percentage using logistic regression. *Computing in Cardiology*, pages 489--492.

Horn, B. K. P. (1986). *Robot Vision*. McGraw-Hill Book Company, Cambridge, Massachusetts. MIT Electrical Engineering and Computer Science Series.

Huddar, V., Desiraju, B. K., Rajan, V., Bhattacharya, S., Roy, S., and Reddy, C. (2016). Predicting complications in critical care using heterogeneous clinical data. *IEEE Access*, 4:7988--8001.

Johnson, A., Dunkley, N., Mayaud, L., Tsanas, A., Kramer, A., and Clifford, G. (2012). Patient specific predictions in the intensive care unit using bayesian ensemble. *Computing in Cardiology*, pages 249--252.

Johnson, A., Ghassemi, M., Nemati, S., Niehaus, K., Clifton, D., and Clifford, G. (2016). Machine learning and decision support in critical care. *Proc. of the IEEE*, 104(2):444--466.

Johnson, A., Kramer, A., and Clifford, G. (2014). Data preprocessing and mortality prediction: the physionet/cinc 2012 challenge revisited. *Computing in Cardiology*, pages 157--160.

Krajnak, M., Xue, J., Kaiser, W., and Balloni, W. (2012). Combining machine learning and clinical rules to build an algorithm for predicting ICU mortality risk. *Computing in Cardiology*, pages 401--404.

Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. In *Proc. of the 26th NIPS*, pages 1106--1114.

Lee, D. H. and Horvitz, E. (2017). Predicting mortality of intensive care patients via learning about hazard. In *Proc. of the 31st AAAI*, pages 4953--4954.

Lorensen, W. E. and Cline, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 163--169.

Luo, Y., Xin, Y., Joshi, R., Celi, L., and Szolovits, P. (2016). Predicting ICU mortality risk by grouping temporal trends from a multivariate panel of physiologic measurements. In *Proc. of the 30th AAAI*, pages 42--50.

McMillan, S., Chia, C., Esbroeck, A., Rubinfeld, I., and Syed, Z. (2012). ICU mortality prediction using time series motifs. *Computing in Cardiology*, pages 265--268.

Moro, M. M., Bakalov, P., and Tsotras, V. J. (2007). Early Profile Pruning on XML-aware Publish/Subscribe Systems. In *Proceedings of the International Conference on Very Large Data Bases*, pages 866–877.

Nori, N., Kashima, H., Yamashita, K., Kunisawa, S., and Imanaka, Y. (2017). Learning implicit tasks for patient-specific risk modeling in ICU. In *Proc. of the 31st AAAI*, pages 1481--1487.

Odon de Alencar, R. (2011). Utilizando evidência da wikipedia para relacionar textos a lugares. Master's thesis, Programa de Pós-Graduação em Ciência da Computação. Universidade Federal de Minas Gerais.

Patel, V., Shortliffe, E., Stefanelli, M., Szolovits, P., Berthold, M., Bellazzi, R., and Abu-Hanna, A. (2009). The coming of age of artificial intelligence in medicine. *Artificial Intelligence in Medicine*, 46(1):5--17.

Raghavan, B., Vishwanath, K., Ramabhadran, S., Yocum, K., and Snoeren, A. (2007). Cloud Control with Distributed Rate Limiting. In *Proceedings of the ACM SIGCOMM Conference*.

Sakai, T. (2014). Statistical reform in information retrieval? *SIGIR Forum*, 48(1):3--12.

Seshamani, M. and Gray, A. (2004). A longitudinal study of the effects of age and time to death. *Journal of Health Economics*, 2(23):217--235.

Shashua, A. (1997). On photometric issues in 3D visual recognition from a single 2D image. *International Journal of Computer Vision (IJCV)*, 21(1/2):99--122.

Silva, I., Scott, G. M. D., Celi, L., and Mark1, R. (2012). Predicting in-hospital mortality of ICU patients: The physionet/computing in cardiology challenge. *Computing in Cardiology*, pages 245--248.

Tabak, Y., Sun, X., Nunez, C., and Johannes, R. (2014). Using electronic health record data to develop inpatient mortality predictive model: Acute laboratory risk of mortality score (alarms). *JAMIA*, 21(3):455--463.

Traina, A. J. M. and Traina Jr., C. (2009). Como fazer pesquisa bibliográfica. *SBC Horizontes*, 2(2):30–35.

Turabian, K. L. (2007). *A Manual for Writers of Research Papers, Theses, and Dissertations*. University Of Chicago Press.

Vairavan, A., Eshelman, L., Haider, S., Flower, A., and Seiver, A. (2012). Prediction of mortality in an intensive care unit using logistic regression and hidden markov model. *Computing in Cardiology*, pages 393--396.

Wazlawick, R. S. (2009). *Metodologia de Pesquisa para Ciência da Computação*. Elsevier.

Wu, M., Ghassemi, M., Feng, M., Celi, L., Szolovits, P., and Doshi-Velez, F. (2017). Understanding vasopressor intervention and weaning: risk prediction in a public heterogeneous clinical time series database. *JAMIA*, 24(3):488--495.

Xia, H., Daley, B., Petrie, A., and Zhao, X. (2012). A neural network model for mortality prediction in ICU. *Computing in Cardiology*, pages 261--264.

Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? In *Proc. of the 28th NIPS*, pages 3320--3328.

Zhang, T., Ramakrishnan, R., and Livny, M. (1996). BIRCH: An Efficient Data Clustering Method for Very Large Databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 103–114.

Zobel, J. (2004). *Writing for Computer Science*. Springer.