

UiO : **University of Oslo**

Chi Zhang

Representation and Utilization of Hospital Electronic Health Records Data

Thesis submitted for the degree of Philosophiae Doctor

Oslo Center for Biostatistics and Epidemiology, Department of
Biostatistics

Institute of Basic Medical Sciences, Faculty of Medicine



2022

© **Chi Zhang, 2022**

*Series of dissertations submitted to the
Faculty of Medicine, University of Oslo*

ISBN 978-82-348-0083-2

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Print production: Graphics Center, University of Oslo.

To my family

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of *Philosophiae Doctor* at the University of Oslo. The research presented here is conducted under the supervision of Professor Magne Thoresen and Professor Arnaldo Frigessi at the Oslo Center for Biostatistics and Epidemiology, Faculty of Medicine, University of Oslo.

The thesis is a collection of three papers, presented in chronological order. The common theme to them is the representation and utilization of temporal and relational hospital Electronic Health Records (EHR) data. The papers are preceded by an introductory chapter that relates them together and provides background information and motivation for the work. The first paper is a joint work with Dr Hadi Fanaee Tork. The other two papers are joint work with Dr Signe Søvik, Dr Torsten Eken and Dr Silje Bakken Jørgensen.

Acknowledgements

It has been almost six years since I started my PhD program in August 2016. The journey has been a rather bumpy one: in the first half of my PhD we had no data; in the second half we had no freedom of movement because of the COVID-19 pandemic. Yet here we are, thanks to the tremendous amount of hard work poured into the projects by everyone involved, and our unwavering determination to finish what we started.

First I would like to thank the Faculty of Medicine of the University of Oslo for letting me have the opportunity to work on the interesting projects. Then I would like to thank OCBE, the Oslo Center for Biostatistics and Epidemiology for providing the a great research environment that is inspiring and dynamic. I would also like to thank Big Insight for funding me throughout the PhD projects and providing invaluable networking opportunities.

I would like to express my sincere gratitude to my main supervisor Professor Magne Thoresen for his guidance, patience and support during the entire journey. Magne's calmness gave me confidence to finish the PhD under ever-challenging circumstances. I would like to give a special thanks to him and his family for inviting me to the hytta in Rondane in summer 2020 during the pandemic and to their home in Christmas 2020. I would then like to express my deep gratitude for Professor Arnaldo Frigessi for welcoming me to Norway in 2016. Arnaldo encouraged and supported me to do various activities, including volunteering at Red Cross, giving talks at the annual Big Insight seminars and attending international conferences which I enjoyed very much, and all made me who I am today.

I would like to thank Dr Hadi Fanaee Tork for his scientific advice on the first paper, which built the foundation of the thesis; and I would like to thank Dr Silje Bakken Jørgensen for her insights on hospital acquired infections and clinical interpretations. Dr Signe Søvik and Dr Torsten Eken have invested tremendous amount of time in the second and the third paper, from data acquisition to manuscript revision. Their passion about research had greatly boosted my morale and confidence, which I am mostly thankful for.

I am truly grateful for working with my great colleagues at OCBE and FHI (Folkehelseinstituttet, the Norwegian Institute of Public Health). I thank Dr René Holst for encouraging me to work on advising projects which turned out to be something I enjoy doing. I thank Professor Birgitte Freiesleben de Blasio for introducing me to Arnoldo back in 2015/16 and offering me a job in 2020 that I am passionate about. I thank Dr Richard Aubrey White for inviting me to work on Sykdomspulsen at FHI during the pandemic, and offering me an extremely exciting opportunity to apply my skills in public health.

I have made many friends since I came to Oslo in 2016, and their company has lightened my mood during these admittedly very challenging years. I thank Solveig Engebretsen, Leiv Rønneberg, Niklas Maltzahn, Reza Ghiasvand, Vera Djordjilovic, Ulpu Remes, Owen Thomas, Xiaoran Lai, Simon Lergenmuller and others at OCBE. I thank them for the good food we shared, the good gym and outdoor activities we did, the good cat therapy I received, the good karaoke we had and the good time we enjoyed in the late 2010s. In addition, I thank Jessica Cao for her valuable assistance in the heavy-lifting when I moved to my new apartment.

Due to the pandemic, it has been impossible for me to visit my family in China, and for the most part of 2020/21, nearly impossible to visit my partner in Europe. Nonetheless, I would like to express my deepest gratitude to my parents, who raised me to be independent yet support me with everything they have. Finally, I thank my partner Federico for having my back all these years.

• Chi Zhang

Oslo, March 2022

Summary

Background

Electronic Health Records (EHR) data contain the medical and treatment history of patients, and have become widely adopted in hospitals in the last decade. Hospital EHR data collected during patient visits contain rich information covering the disease history and progression, medication, procedures and diagnoses. The availability of large amount of patient data has brought new opportunities in several research fields, including medicine, epidemiology and method developments using statistical and artificial intelligence tools. In spite of the exciting opportunities, using EHR data for research is challenging. EHR data is inherently multi-modal, longitudinal and noisy. Some EHR data with timestamps are also relational which can reveal additional structural information about the system. Effective extraction and representation of hospital EHR data is a first step to understand the complexity of hospital environment and improve quality of care.

Aims

There are two closely related aims of the thesis. As EHR data exist in various formats, there is the need for specific tools for each type of EHR data. The first aim of the thesis is therefore to develop and experiment with different statistical and computational methods to extract, integrate and represent information from hospital EHR data. In this thesis, I explore data mining algorithms (dynamic time warping), machine learning classification algorithms, network analysis on sequential relational data, regression problems, prediction and variable selection algorithms. The second aim is to demonstrate the broad scope of applications of EHR data in the clinical setting. I used two very different hospital EHR datasets: the MIMIC III database from the US and the AHUS data from Norway. The applications are diverse, for instance, clinical event classification and prediction (paper I and III) could be used for patient risk stratification and early warning. Understanding patient transfer patterns and associated risk factors (paper II) could improve hospital management and logistic efficiency.

Paper I

Temporal EHR data is challenging to work on, as the time series for each clinical feature may be measured at different time frequencies, some more often (e.g. heart rates) and some less often (e.g. test results). In this paper, we propose a combination of dynamic time warping (DTW) and tensor decomposition to represent the unequal time series such that subsequent classification tasks can be performed. The representation feature matrix is then fed into a classifier such as logistic regression or support vector machine to carry out clinical event classification. We demonstrate our method on two cohorts extracted from an openly accessible database, the MIMIC III, aiming to classify in-hospital

mortality. Our method yields promising results compared to other machine learning algorithms commonly used for this application.

Paper II

This paper study the hospital patient transfer patterns at both system (clinical department) level and individual patient level using EHR. Understanding intrahospital transfer pattern is a crucial step for hospital capacity management and infection control. As far as we know, this is the first study that investigates different types of transfers in a variety of departments, both surgical (gastrointestinal surgery and orthopaedics) and medical (gastroenterology and neurology). This is also the first study on patient transfer pattern in a Nordic hospital. With the detailed description of patient movements, the paper illustrates the heterogeneity among different types of patients. We quantify the types and frequencies of transfers, and extract the most common ones at the department level. In addition, we associate number of transfers with clinical risk factors to quantify the risk of excessive transfers. These results are relevant for evidence based method development using simulations to model patient flow, capacity management and infection control, as we show the complexity of hospital systems and the necessity to distinguish different categories of patients. This paper complements the thesis by providing detailed examination of an interesting type of sequential relational EHR data: the location transitions. It provides additional knowledge by quantifying how different patients move within a healthcare facility using graph theory and network analysis.

Paper III

This paper presents a novel framework for predictions in a healthcare setting, with potential applications in hospital operation optimization using EHR. We introduce the term "intervenable prediction" in this paper, which is a two step approach that balance interpretability and prediction accuracy. This serves as the basis for possible intervention strategies. In the first step, a penalized linear regression model that induces sparsity is fitted to the data, with the aim to select variables (risk factors) that can predict the outcome reasonably well. Then in the second step, we use the residuals from the previous step as the new response, and use the lagged responses as the new predictors to fit a second regression model. In this way, the model captures information in both the explanatory variables and the hidden trend in the response itself. We illustrate the method with simulation studies using both purely synthetic data and real EHR data from a Norwegian hospital. We demonstrate the different scenarios in which our method could improve the prediction using the two step approach, and effects of possible intervention strategies. The method can be easily extended to more complex models beyond linear models.

Sammendrag

Bakgrunn

Elektronisk pasientjournal (EPJ) har i stor grad blitt tatt i bruk på sykehus i løpet av det siste tiåret. Sykehus-EPJ-data inneholder rik informasjon angående diagnoser, sykdomshistorie og sykdomsprogresjon, medisinering og prosedyrer. Tilgjengeligheten av store mengder pasientdata har gitt nye muligheter innen flere forskningsfelt, inkludert medisin, epidemiologi og metodeutvikling innenfor statistikk og kunstig intelligens. Til tross for de spennende mulighetene, er det utfordrende å bruke EPJ-data til forskning. EPJ-data er multimodale, longitudinelle og støyfylte. Enkelte typer EPJ-data med tidsinformasjon er også relasjonelle og kan avsløre strukturell informasjon om sykehussystemet. Effektiv uthenting og organisering av sykehus-EPJ-data er et første skritt i retning av å forstå kompleksiteten i et sykehussystem og å forbedre kvaliteten på omsorgen.

Formål

Avhandlingen har to nært beslektede formål. Ettersom EPJ-data finnes i ulike formater, er det behov for spesifikke verktøy for hver type EPJ-data. Det første formålet med oppgaven er derfor å utvikle og eksperimentere med ulike statistiske og beregningsorienterte metoder for å trekke ut, integrere og organisere informasjon fra sykehus-EPJ-data. I denne avhandlingen utforsker jeg data mining-algoritmer, klassifikasjonsalgoritmer innenfor maskinlæring, nettverksanalyse på sekvensielle relasjonelle data, samt regresjonsproblemer og prediksjons- og variabelseleksjonsalgoritmer. Det andre formålet er å demonstrere det brede anvendelsesområdet av EPJ-data i en klinisk setting. Jeg bruker to svært forskjellige sykehus-EPJ-datasett: MIMIC III-databasen fra USA og et datasett fra Akershus Universitetssykehus (AHUS) i Norge. Anvendelsene er forskjellige. For eksempel kan klassifisering og prediksjon av kliniske hendelser (artikkel I og III) brukes til å stratifisere pasienter i forhold til risiko og til tidlig varsling. Å forstå pasientoverføringsmønstre og tilhørende risikofaktorer (artikkel II) kan forbedre sykehusorganisering og logistikk.

Artikkel I

EPJ-data registrert over tid er utfordrende å jobbe med, ettersom tidsseriene for hver klinisk funksjon kan måles ved forskjellige tidsfrekvenser, noen oftere (f.eks. hjertefrekvens) og noen sjeldnere (f.eks. testresultater). I denne artikkelen foreslår vi en kombinasjon av «dynamic time warping» og tensordekomposisjon for å representere ulike tidsserier slik at påfølgende klassifiseringsoppgaver kan gjennomføres. Den resulterende egenskapsmatrisen mates deretter inn i en klassifiseringsalgoritme, som logistisk regresjon eller «support vector machine» for å gjennomføre klassifikasjon med hensyn på for eksempel kliniske hendelser. Vi demonstrerer metoden vår på to kohorter hentet fra en åpent tilgjengelig database, MIMIC III, med sikte på å klassifisere dødelighet på sykehus. Metoden

vår gir lovende resultater sammenlignet med andre maskinlæringsalgoritmer som vanligvis brukes for denne typen anvendelser.

Artikkel II

I denne artikkelen studerer vi mønstre for overføring av pasienter på både systemnivå (type avdeling) og på individuelt pasientnivå ved bruk av EPJ. Å forstå overføringsmønstre innen sykehus er viktig for sykehuskapasitetsstyring og infeksjonskontroll. Så vidt vi vet er dette den første studien som undersøker ulike typer forflytninger ved forskjellige avdelinger, både kirurgiske (gastrointestinal kirurgi og ortopedi) og medisinske (gastroenterologi og nevrologi). Dette er også den første studien som behandler pasientoverføringsmønstre på et nordisk sykehus. Med den detaljerte beskrivelsen av pasientbevegelser vi har tilgang til, illustrerer artikkelen variasjonen mellom ulike typer pasienter. Vi kvantifiserer typer og frekvenser av overføringer, og henter ut de vanligste typene på avdelingsnivå. I tillegg studerer vi sammenhengen mellom antall overføringer og kliniske risikofaktorer for å kvantifisere risikoen for unødvendige / overdrevne overføringer. Disse resultatene er relevante for evidensbasert metodeutvikling ved bruk av simuleringer for å modellere pasientflyt, kapasitetshåndtering og infeksjonskontroll, da vi viser kompleksiteten i et sykehussystem og nødvendigheten av å skille ulike kategorier av pasienter. Denne artikkelen utfyller avhandlingen ved å gjøre en detaljert undersøkelse av en interessant type sekvensielle og relasjonelle EPJ-data: overføring mellom forskjellige lokasjoner. Den gir ytterligere kunnskap ved å kvantifisere hvordan ulike pasienter beveger seg innenfor en helseinstitusjon ved hjelp av grafteori og nettverksanalyse.

Artikkel III

Denne artikkelen presenterer et nytt rammeverk for prediksjoner innen helsevesenet, med potensielle anvendelser innenfor optimalisering av sykehusdrift ved bruk av EPJ. Vi introduserer begrepet "intervenerbar prediksjon" i denne artikkelen. Dette er en to-trinns tilnærming til prediksjon som balanserer tolkbarhet og prediksjonsnøyaktighet. Dette fungerer som grunnlag for mulige intervensjonsstrategier. I det første trinnet blir en regularisert lineær regresjonsmodell tilpasset dataene. Denne modellen søker å velge ut variabler (risikofaktorer) som kan forutsi utfallet rimelig godt. Så, i det andre trinnet, bruker vi restleddene fra forrige trinn som den nye responsen, og de forutgående responsene som nye prediktorer for å tilpasse en ny regresjonsmodell. På denne måten fanger modellen opp informasjon både i forklaringsvariablene og den skjulte trenden i selve responsen. Vi illustrerer metoden med simuleringstudier ved bruk av både rent syntetiske data og ekte EPJ-data fra et norsk sykehus. Vi demonstrerer de forskjellige scenariene der metoden vår kan forbedre prediksjonen ved å bruke to-trinns tilnærming, og effektene av mulige intervensjonsstrategier. Metoden kan enkelt utvides til mer komplekse modeller utover de lineære.

List of Papers

Paper I

Zhang, C., Fanaee-T, H. and Thoresen, M. “Feature extraction from unequal length heterogeneous EHR time series via dynamic time warping and tensor decomposition”. In: *Data Mining and Knowledge Discovery* **35** (2021), 1760-1784. DOI: 10.1007/s10618-020-00724-6.

Paper II

Zhang, C., Eken, T., Jørgensen, S.B., Thoresen, M., Søvik, S. “Effects of patient-level risk factors, departmental allocation and seasonality on intrahospital patient transfer patterns – network analysis applied on a Norwegian single-centre dataset”. Accepted for publication in: *BMJ Open* (2022), **0**:e054545. DOI: 10.1136/bmjopen-2021-054545.

Paper III

Zhang, C., Frigessi, A., Søvik, S., Eken, T., Thoresen, M. “Intervenable predictions of hospital operations using Electronic Health Records”. *Submitted for publication*.

Contents

Preface	iii
List of Papers	ix
Contents	xi
List of Figures	xiii
List of Tables	xv
1 Introduction	1
2 Electronic Health Records	3
2.1 Current state of EHR systems	3
2.2 EHR in daily care	4
2.3 EHR in research	5
2.4 Ethics	7
References	9
3 Data sources	15
3.1 MIMIC III data	15
3.2 AHUS data	15
References	16
4 Aims of the thesis	17
5 Clinical event classification with temporal EHR data	19
5.1 Clinical event prediction and classification	19
5.2 Temporal EHR data	19
5.3 Time series classification	20
5.4 Tensor decomposition	22
References	23
6 Intrahospital patient transfer pattern	25
6.1 Hospital patient flow	25
6.2 Network analysis	26
6.3 Patient flow from EHR location logs: a network approach .	28
References	30
7 Intervenable prediction for hospital operation	33

Contents

7.1	Prediction and intervention in healthcare environment . . .	33
7.2	Regression with sparsity constraints	34
7.3	Intervenable prediction with simulation and EHR data . . .	35
	References	36
8	Summary of the papers	39
9	Discussion	43
9.1	Overview	43
9.2	Paper I	44
9.3	Paper II	47
9.4	Paper III	52
9.5	Data protection	55
9.6	Future perspectives	56
	References	56
10	Conclusion	63
	Papers	66
I	Feature extraction from unequal length heterogeneous EHR time series via dynamic time warping and tensor decomposition	67
II	Effects of patient-level risk factors, departmental allocation and seasonality on intrahospital patient transfer patterns - network analysis applied on a Norwegian single-centre dataset	95
III	Intervenable predictions of hospital operations using Electronic Health Records	127

List of Figures

- 5.1 Illustration of DTW [Tan2019]. a: Alignment for two time series Q (blue) and C (red). b Cost matrix C with warping path W (green). 21

- 6.1 Hospital event from arrival to discharge. Figure adapted from [Ben-Tovim2016] after excluding EECU (Extended Emergency Care Unit). 25
- 6.2 Patient transfer network illustration. For details and abbreviations see Paper II. 28
- 6.3 Examples of network motifs. Numbers are the ordering of edges. 29

List of Tables

5.1 List of notations 22

Chapter 1

Introduction

Electronic Health Records (EHR) data contain the medical and treatment history of patients, and have become widely adopted in hospitals in the last decade. Hospital EHR data collected during patient visits contain rich information covering the disease history and progression, medication, procedures and diagnoses. Modern hospital EHR databases such as the MIMIC III database and eICU initiative can make data on tens of thousands of patients available to researchers. The availability of large amount of patient data has brought new opportunities in several research fields, including medicine, epidemiology and method developments using statistical and artificial intelligence tools.

In spite of the exciting opportunities, using EHR data for research is challenging. Initially recorded for administrative and billing purposes instead of research, EHR data quality is generally insufficient to answer research questions without great amount of pre-processing. EHR data is inherently multi-modal (i.e. recorded in many formats), longitudinal (measured over time) and noisy. Some EHR data with timestamps are also relational which can reveal additional structural information of the system. Effective extraction and representation of hospital EHR data is a first step to understand the complexity of hospital environment and improve quality of care.

There are two objectives of this thesis. The first objective is to explore different statistical and computational methods to extract, integrate and represent information from temporal and relational hospital EHR data. Since EHR exist in various formats, there is the need for using different tools for each data type. In this thesis I explored data mining algorithms (dynamic time warping), machine learning classification algorithms, network analysis on sequential relational data, regression models and prediction problems as well as variable selection algorithms. Some focus on the comparison and development of methods while others focus on the data exploration and interpretation. The field of medical informatics is rapidly expanding and the use of EHR is fundamental. It is necessary to be equipped with skills from various fields (statistic, computer science, medicine) and pick the most appropriate one to solve the problems at hand.

The second objective is to demonstrate the broad scope of potential applications of EHR data in the clinical setting. For instance, clinical event classification and prediction (paper I and III) could be used for patient risk stratification and early warning. Understanding patient transfer patterns and associated risk factors (paper II) could improve hospital management and logistic efficiency. With the ongoing digitalization process, more data sources are becoming available and novel applications will emerge as a consequence.

1. Introduction

This thesis contains three papers. In Paper I, I illustrate the use of data mining algorithms on heterogeneous unequal length temporal EHR data to extract features from ICU patients, and the classification performance using binary classification algorithms. A wide range of algorithms for this type of data are compared. In Paper II, I use methods from network analysis to investigate the patient flow in a large regional hospital in Norway using EHR data. Compared to other studies that analyze patient flow, this paper includes time information (work day, week in the year) and additional patient level risk factors to identify the heterogeneity in different hospital departments. In paper III, I develop a method for intervenable prediction in a hospital environment and demonstrate the use-case using aggregated EHR data at the ward level. This paper is a proof of concept that can be extended by using various models and applied in different contexts. Two datasets have been used to illustrate the methods: the publicly available MIMIC III critical care database (Johnson2016) from Boston, United States; and data from Akershus University Hospital in the Greater Oslo Region, Norway.

The thesis is structured as follow. Chapter 2 provides a brief introduction of the EHR data, which covers the history and current state of adoption, its use in daily care and research and the challenges in both scenarios. Ethics related to EHR use is also introduced in this chapter. In Chapter 3 the two datasets used in the papers are introduced. Chapter 4 states the aim of the thesis. Chapter 5, 6 and 7 provide the background and introduction of paper I, II and III, followed by the Chapter 8 that summarizes all three papers. Chapter 9 contains the discussions: first I discuss the topics that are common in all three papers, then I discuss each paper in detail separately. In the end, I conclude the thesis in Chapter 10.

Chapter 2

Electronic Health Records

Electronic Health Records, or EHR data, are the digital version of patient records. EHR data contain the medical and treatment history of patients that can be used to share information within and across healthcare providers. EHR data are collected when a patient encounters the healthcare system, either in primary care services such as General Practitioner (GP) and community care; or specialized care services such as hospitals.

EHR data can be broadly categorized into the following five categories based on their functionalities [36].

1. Health information and data: symptoms, reasons for appointment, clinical notes, vital signs, treatment outcomes, medical history, diagnoses.
2. Clinical decision support system: drug-drug interactions, drug-allergy alerts, alerts to critical laboratory values, clinical guidelines.
3. Order-entry and result management: medication lists, prescriptions, lab test results
4. Image: radiology test images and reports
5. Administrative: finances and billing, administrative patient data

In this chapter I will first give a brief introduction to the EHR data in daily care: the development and current state, opportunities and challenges associated to the use of EHR. Then I will cover the major research topics that utilize EHR data. The ethical issues related to the use are discussed in chapter ?? at the end of the thesis.

2.1 Current state of EHR systems

EHR systems falls under the scope of Digital Health or Healthcare Information Technology (HIT). The development of HIT and EHR adoption differs from country to country. In the US, hospitals started adopting IT systems as early as in the 1960s [2], however EHR only became widely installed and used after 2009, following the Health Information Technology for Economic and Clinical Health Act (HITECH Act) [18, 33] (United States Public Law, Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009. Section 13410(d)) in the Obama Administration. In 2017, 96% of the US hospitals and 86% of the physician offices had EHR systems. Among Asia-Pacific

2. Electronic Health Records

countries, some have mature systems (Korea, Japan, Common Wealth regions such as Singapore and Hong Kong) while others are experiencing a high rate of development. China, for example, has witnessed an increase from 18.6% to 85.3% (2007-2018) in hospital EHR adoption [34]. European hospitals on average have a high EHR adoption rate (over 95% hospitals in 2015). Some countries had prompted EHR strategies and plans since early 1990s (Northern Europe), and some started the process in the last decade (Italy, 2010) [7].

In Norway the EHR systems are being used in three levels in the healthcare system: Primary healthcare (GP, nursing homes and community care), Specialist healthcare (hospitals) and National Summary Care Record [13]. Among the specialist healthcare facilities, three regions out of four (Southern-Eastern, Northern and Western Region) use the Norwegian solution DIPS, one (Central Region) uses Doculive EPJ.

2.2 EHR in daily care

The biggest motivation for EHR use in healthcare services is to share complete, secure patient information between practitioners. As more and more hospitals and healthcare workers began using EHR systems in the last decade, there has been a mixed view on EHR systems in day-to-day use.

In a survey [36] of over 5000 GPs (general practitioners) in European countries, over 90% use EHR for the following functionalities: prescription and medication, diagnosis, lab test results, medical history and clinical notes and drug-allergy alerts. The benefit of EHR usage falls into three major categories:

- Information management: use of EHR allows practitioners to have fast access to patient information such as medical history and allergy. It also allows faster reimbursement process.
- Clinical communication: use of EHR provides better coordination across different medical teams, for example, drug prescription between physician, pharmacist and nursing staff. It also allows information share across different providers.
- Decision support and error reduction: some EHR systems have embedded algorithms to check drug allergies, drug dosing errors as well as drug-drug interactions. This is particularly useful to avoid harm caused by prescription.

There are also concerns and challenges for making full use of EHR [2]. For physicians, the daily added workload for learning and using the EHR system can have a negative impact on the user experience. Maintenance issues and tool failures can also be prohibiting. EHR systems can also be expensive to install

and maintain. As for the cost reduction brought by EHR for hospitals and for the whole healthcare system, the evidence is mixed.

2.3 EHR in research

The availability of EHR data provides a great data source for research in multiple fields, such as medicine, clinical practice, public health and epidemiology, methodology and model development in statistics as well as algorithms and data management in computer science. In this section, I first give a short introduction to the applications of EHR in two research fields: epidemiology and machine learning; then I discuss some of the challenges of using EHR in research.

2.3.1 Epidemiological research

There are several advantages of using EHR for epidemiological research. Thanks to the large sample sizes, researchers can look at smaller subgroups; in addition, since data already exist in the databases, it is less expensive and time consuming to follow up a cohort [10].

The direct impact of the large volume of EHR data becoming available is that now it is possible to re-evaluate prior findings from smaller studies [12] or to inspire new research directions [4, 41]. For instance, new metrics can be proposed and evaluated with the rich data source [15, 20]. The most common use of EHR in epidemiological research is to discover and evaluate associations between risk factors and outcomes. Such examples include biomarker-disease associations [38], obesity with acute kidney injury [9], obesity with one-year survival [56], sex with clinical outcomes in sepsis patients [54] and many more. Furthermore, some fundamental causes of disease are also available or can be included in EHR: social, behavioral, environmental factors, alcohol and tobacco use, stress and physical activity [10]. This facilitates research that link clinical outcomes with environmental and social factors, such as socioeconomic status with sepsis mortality [43]. Another potential use of EHR is to study stigmatized conditions where participant recruitment is difficult, such as HIV treatment [41], HIV patient and mental health [44] and other mental health outcomes [52].

2.3.2 Machine learning using EHR

During the last decade, EHR data has provided unprecedented opportunities for researchers in the machine learning community to develop and test algorithms in the last decade. The biggest advantage of EHR data is the volume: in both the large size of study cohort, and the number of variables as input and outcome. For instance in the most popular open source MIMIC-III database [29], researchers can build models to perform mortality prediction, length-of-stay predictions, 30 day re-admission predictions and acute kidney injury predictions among many other modeling tasks, using a mixed bag of input data including vital sign time series, demographics, lab results and clinical texts.

2. Electronic Health Records

The most common learning tasks for machine learning using EHR are *disease detection or risk predictions*, where the goal is to develop models to map EHR data to the outcome target. This can be done with statistical models (e.g. generalized linear regressions such as logistic regression; penalized regressions), traditional machine learning algorithms (e.g. Support Vector Machine, ensemble methods, Random Forests) or deep learning depending on the input data, yet it is at its core a supervised learning problem. Mortality prediction (in- or out-of-hospital) has been the target outcome for various benchmark studies using MIMIC-III database [22, 28, 40]. Some clinical adverse events are also of interest, such as hospital acquired infection after surgery [26, 37], seizure [1], myocardial infarction, sepsis onset (see [19] for a complete list of outcome of interest). The input data are typically heterogeneous: they may contain both static (such as age, sex, ethnicity) and temporal (vital signs) data; and they could be both structured (in a table with pre-specified fields) or unstructured texts. The data being multi-modular implies that some data representation (linear or non-linear transformation) is necessary before a discriminatory task (classification) takes place: for example, Natural Language Processing (NLP) is essential when clinical text is included as input.

A few topics other than risk prediction are also getting attention. *Computational phenotyping* embeds EHR data (such as clinical text, medication) to clinical concepts, or phenotypes of interest [3, 46]. This can be done with both classical unsupervised learning algorithms (Principal Component Analysis, K-means clustering), matrix and tensor methods [23, 24] as well as deep learning [53]. Given the importance of data security and privacy protection for using health data, algorithms that enable *privacy preserving learning* to de-identify EHR data or generate synthetic records for research have been proposed in recent years [5, 6, 8, 35].

The evaluation of machine learning algorithms using EHR is typically through their prediction or classification accuracy. Area under receiver operating curve (AUROC), positive predictive value, precision-recall curve (AUPRC) are the most commonly used metrics for binary classifications. There is no single best algorithm that outperforms the others in all situations and datasets, sometimes logistic regression model is among the best performing classifiers [22, 28, 40].

2.3.3 Challenges

Among the many challenges related to using EHR for research, *data quality* challenge is arguably the fundamental one. First of all, EHR are collected during patient encounters with the healthcare system, which suggests that EHR contain more complete data for the ill than the healthy. Lack of data from the healthy population could lead to biased conclusions. EHR systems nowadays have not existed long enough to store many years of data that cover different aspects of a patient. Fortunately this could potentially be solved via linkage to other

registries [10].

Secondly, missing and erroneous data is very common in EHR even for patients whose data have been collected. Measurements could be taken only when physicians think it is necessary, hence we might miss some information [19]; some variables are difficult to observe, such as patient compliance with prescriptions and interventions [55]; diagnostic codes could be missing due to changes in disease definition like updates from ICD-9 to ICD-10 codes [55]. Outcome labels could be unavailable or misleading, for example, diagnostic codes could imply that a patient was screened instead of diagnosed [17]. A few solutions are 1) develop software to facilitate accurate data entry; 2) use more than one database to construct a patient history and verify outcome definitions; 3) understand the cause of missing: is the data missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR), and choose methods that can deal with each type of missingness accordingly [16].

Apart from data quality challenges, there are also a few *analytical* challenges. Bias is a direct consequence of incomplete data. If data used to develop predictive models do not represent the population where the model will be deployed, this model is hardly a correct one. Researchers generally would validate the results on the test data, but rarely on external data, like another hospital [19]. This is partly because there are very few open, multi-center EHR data for research purposes. In order for models to generalize, there is a need for collaborated research using more data sources. Causality analysis and modeling using EHR is also not an easy task. Using existing data collected in the past suggests that the study is observational instead of experimental, where the latter is considered as the golden-standard to draw causal conclusions. On top of these comes the interpretation of machine learning models, which is hard to achieve when the model involves complex architectures like deep learning. Building interpretable machine learning models that can carry out causal inference is a particularly active research field [11, 49].

2.4 Ethics

I will introduce the ethical issues related with EHR use in this section. The first topic is the privacy, consent and data security; and the second topic is the responsible use of EHR in machine learning research.

2.4.1 Privacy, consent and data security

Patient data use (including EHR) for direct care such as legitimate clinical treatment and audit purposes is seen as the primary use of health data and do not require consent [47]. On the contrary, research is considered as the secondary use of health data and require stricter regulations. Consent should be obtained from individual patients, and the research itself should be in line with the public

2. Electronic Health Records

interest [47]. According to [30, 31, 42], patients are generally supportive about their data being used in research. Nevertheless, there are also concerns. One of the most common concerns is that patients fear that they may not have control of who have access to their data and for what purpose[33]. For example, in a study patients expressed their wish to not share data with pharmaceutical companies [21]. They might fear that the data could be linked to other personal data: disclosure of health information could lead to undesired predictive marketing, discrimination in employment, denial of certain insurance coverage [14]. Patients might also have concerns about the competence of the healthcare providers for data protection from attacks [27].

There are a few ways to facilitate the secure use of health data. Legal requirements could reduce the risk of unauthorized disclosure, such examples are the HIPAA Privacy Rule [48], the HITECH Act (2009), the Federal Policy for the Protection of Human Subjects in the United States and the General Data Protection Regulation (GDPR) in EU and EEA (reference 4,5,6 in [33]). There should be consequences for data breaches in the forms of civil and criminal penalties [33]. Encryption and access limits to personnels are also ways avoid unauthorized disclosure. In order to reduce the risk that an individual patient might be identified, data holders could remove pseudonymisation code so that re-identification becomes impossible, and use secure computing systems that do not allow the data to be downloaded outside the system[47]. Education on data security should be provided to personnels handling collection, data entry, data management and storage by IT professionals as well as end-users to promote secure use of patient data[33].

2.4.2 Responsible machine learning using EHR

A few considerations need to be taken in order to use EHR in a research in a responsible way, especially when developing machine learning algorithms. The clinical relevance of the objective should be considered carefully: in simple words, ask good questions and solve actionable questions[51]. If developing an in-hospital mortality prediction tool does not do better than what the clinicians already know, then this research question could be more specific or revised. Similarly, if a prediction model for death does not allow for any intervention, or the strategy is constrained by resources, then this is not actionable intelligence. Developing methods and algorithms that are clinically relevant and actionable requires stakeholder engagement, and physicians must guide, oversee and monitor the adoption of ML/AI solutions[49, 51]; meanwhile solutions that take resource constrains into account could be a promising direction[45].

From the technical point of view, the explainability of the algorithms, data validity and bias induced by either algorithm or data all have ethical implications. These aspects remain challenges in developing useful and fair ML/AI tools for healthcare and beyond, as discussed in the previous section and in [25, 32, 39, 50]. In addition, the reproducibility of results [22, 40] and high standard reporting should also be take into account in scientific research. There are challenges and

great opportunities in EHR and ML/AI research, and ethical considerations must always be prioritized: do no harm.

References

- [1] Ali, H. et al. *Seizure Prediction Using Bidirectional LSTM*. Vol. 1137 CCIS. Springer Singapore, 2019, pp. 349–356.
- [2] Atasoy, H., Greenwood, B. N., and McCullough, J. S. “The Digitization of Patient Care: A Review of the Effects of Electronic Health Records on Health Care Quality and Utilization”. In: *Annual Review of Public Health* vol. 40 (2019), pp. 487–500.
- [3] Banda, J. M. et al. “Advances in Electronic Phenotyping : From Rule-Based Definitions to Machine Learning Models”. In: (2018).
- [4] Banerji, A. et al. “Epidemiology of ACE Inhibitor Angioedema Utilizing a Large Electronic Health Record”. In: *The Journal of Allergy and Clinical Immunology in Practice* vol. 5, no. 3 (), pp. 744–749.
- [5] Baowaly, M. K. et al. “Synthesizing electronic health records using improved generative adversarial networks”. In: *Journal of the American Medical Informatics Association* vol. 26, no. 3 (2019), pp. 228–241.
- [6] Beaulieu-jones, B. K. et al. “Privacy-Preserving Distributed Deep Learning for Clinical Data”. In: (2018). arXiv: [arXiv:1812.01484v1](https://arxiv.org/abs/1812.01484).
- [7] Bonomi, S. “The electronic health record: A comparison of some European countries”. In: *Lecture Notes in Information Systems and Organisation* vol. 15 (2016), pp. 33–50.
- [8] Brisimi, T. S. et al. “Federated learning of predictive models from federated Electronic Health Records”. In: *International Journal of Medical Informatics* vol. 112, no. November 2017 (2018), pp. 59–67.
- [9] Burgess, A. et al. “Obesity, acute kidney injury, and mortality in critical illness”. In: vol. 15, no. 5 (2016), pp. 477–491.
- [10] Casey, J. A. et al. “Using Electronic Health Records for Population Health Research: A Review of Methods and Applications”. In: *Annual Review of Public Health* vol. 37 (2016), pp. 61–81.
- [11] Chen, P. et al. “Interpretable clinical prediction via attention-based neural network”. In: *BMC Medical Informatics and Decision Making* vol. 20, no. Suppl 3 (2020), pp. 1–9.

2. Electronic Health Records

- [12] Dhalwani, N. N. et al. “Women With Celiac Disease Present With Fertility Problems No More Often Than Women in the General Population”. In: *Gastroenterology* vol. 147, no. 6 (2014), 1267–1274.e1.
- [13] e-helse, D. for. “A brief overview of health IT collaboration and interoperability in five countries in 2018”. In: (2019).
- [14] Entzeridou, E., Markopoulou, E., and Mollaki, V. “Public and physician’s expectations and ethical concerns about electronic health record: Benefits outweigh risks except for information security”. In: *International Journal of Medical Informatics* vol. 110, no. December 2017 (2018), pp. 98–107.
- [15] Fan, T. et al. “Nomogram to predict the risk of acute kidney injury in patients with diabetic ketoacidosis: an analysis of the MIMIC-III database”. In: *BMC Endocrine Disorders* vol. 21, no. 1 (2021), pp. 1–11.
- [16] Gharghabi, S. et al. *An ultra-fast time series distance measure to allow data mining in more complex real-world deployments*. Springer US, 2020.
- [17] Ghassemi, M. et al. “A Review of Challenges and Opportunities in Machine Learning for Health.” In: *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science* vol. 2020 (2020), pp. 191–200. arXiv: [1806.00388](https://arxiv.org/abs/1806.00388).
- [18] Gold, M. and McLaughlin, C. “Assessing HITECH 5 Years Later”. In: *Milbank Q.* vol. 94, no. 3 (2016), pp. 654–687.
- [19] Goldstein, B. A. et al. “Opportunities and challenges in developing risk prediction models with electronic health records data: A systematic review”. In: *Journal of the American Medical Informatics Association* vol. 24, no. 1 (2017), pp. 198–208. arXiv: [arXiv:1301.2609v5](https://arxiv.org/abs/1301.2609v5).
- [20] Haines, R. W. et al. “Elevated urea-to-creatinine ratio provides a biochemical signature of muscle catabolism and persistent critical illness after major trauma”. In: *Intensive Care Medicine* vol. 45, no. 12 (2019), pp. 1718–1731.
- [21] Harle, C. A. et al. “Patient preferences toward an interactive e-consent application for research using electronic health records”. In: vol. 25, no. December 2017 (2018), pp. 360–368.
- [22] Harutyunyan, H. et al. “Multitask Learning and Benchmarking with Clinical Time Series Data”. In: (2018). arXiv: [1703.07771](https://arxiv.org/abs/1703.07771).
- [23] Henderson, J., Malin, B. A., and Ho, J. C. “PIVETed-Granite : Computational Phenotypes through Constrained Tensor Factorization”. In: (2018). arXiv: [arXiv:1808.02602v1](https://arxiv.org/abs/1808.02602v1).
- [24] Henderson, J. et al. “Granite: Diversified, Sparse Tensor Factorization for Electronic Health Record-Based Phenotyping”. In: *IEEE International Conference on Healthcare Informatics (ICHI)*. 2017.
- [25] Holzinger, A. et al. “Causability and explainability of artificial intelligence in medicine”. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* vol. 9, no. 4 (2019), pp. 1–13.

-
- [26] Hu, Z. et al. “Automated Detection of Postoperative Surgical Site Infections Using Supervised Methods with Electronic Health Record Data Zhen”. In: (2017), pp. 706–710.
- [27] Hübner, U. H., Egbert, N., and Schulte, G. “Clinical Information Systems - Seen through the Ethics Lens”. In: *Yearbook of medical informatics* vol. 29, no. 1 (2020), pp. 104–114.
- [28] Johnson, A. E. W., Pollard, T. J., and Mark, R. G. “Reproducibility in critical care : a mortality prediction case study”. In: *2nd Machine Learning for Healthcare Conference, PMLR*. Vol. 68. 2017.
- [29] Johnson, A. E. et al. “MIMIC-III, a freely accessible critical care database”. In: *Scientific Data* vol. 3 (2016), p. 160035.
- [30] Kass, N. E. et al. “The use of medical records in research: What do patients want?” In: *Journal of Law, Medicine and Ethics* vol. 31, no. 3 (2003), pp. 429–433.
- [31] Kim, K. K., Joseph, J. G., and Ohno-Machado, L. “Comparison of consumers’ views on electronic data sharing for healthcare and research”. In: *Journal of the American Medical Informatics Association* vol. 22, no. 4 (2015), pp. 821–830.
- [32] Kundu, S. “AI in medicine must be explainable”. In: *Nature Medicine* vol. 27, no. 8 (2021), p. 1328.
- [33] Lee, L. M. “Ethics and subsequent use of electronic health record data”. In: *Journal of Biomedical Informatics* vol. 71 (2017), pp. 143–146.
- [34] Liang, J. et al. “Adoption of electronic health records (EHRs) in China during the past 10 years: Consecutive survey data analysis and comparison of Sino-American challenges and experiences”. In: *Journal of Medical Internet Research* vol. 23, no. 2 (2021).
- [35] Liu, D. et al. “FADL : Federated-Autonomous Deep Learning for Distributed Electronic Health Record”. In: (2018). arXiv: [arXiv:1811.11400v2](https://arxiv.org/abs/1811.11400v2).
- [36] Lupiáñez-Villanueva, F. et al. *Benchmarking Deployment of eHealth among General Practitioners*. 2018, p. 122.
- [37] Menendez, M. E., Janssen, S. J., and Ring, D. “Electronic health record-based triggers to detect adverse events after outpatient orthopaedic surgery”. In: (2016), pp. 25–30.
- [38] Mosley, J. D. “A study paradigm integrating prospective epidemiologic cohorts and electronic health records to identify disease biomarkers”. In: *Nature Communications* (2018).
- [39] Panch, T., Mattie, H., and Atun, R. “Artificial intelligence and algorithmic bias: Implications for health systems”. In: *Journal of Global Health* vol. 9, no. 2 (2019), pp. 1–5.

2. Electronic Health Records

- [40] Purushotham, S. et al. “Benchmarking deep learning models on large healthcare datasets”. In: *Journal of Biomedical Informatics* vol. 83 (2018), pp. 112–134.
- [41] Puttkammer, N. et al. “Development of an Electronic Medical Record Based Alert for Risk of HIV Treatment Failure in a Low-Resource Setting”. In: *PLoS ONE* vol. 9, no. 11 (2014).
- [42] Riordan, F. et al. “Patient and public attitudes towards informed consent models and levels of awareness of Electronic Health Records in the UK”. In: *International Journal of Medical Informatics* vol. 84, no. 4 (2015), pp. 237–247.
- [43] Rush, B. et al. “Association of household income level and in-hospital mortality in patients with sepsis: a nationwide retrospective cohort analysis”. In: vol. 15, no. 5 (2016), pp. 477–491.
- [44] Scorza, P. et al. “HIV and Child Mental Health : A Case-Control Study in Rwanda”. In: *Pediatrics* vol. 134 (2014).
- [45] Shah, N. H., Milstein, A., and Bagley, S. *Making Machine Learning Models Clinically Useful*. 2019.
- [46] Shickel, B. et al. “Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis”. In: *IEEE Journal of Biomedical and Health Informatics* vol. 22, no. 5 (2018), pp. 1589–1604. arXiv: [arXiv:1706.03446v2](https://arxiv.org/abs/1706.03446v2).
- [47] Stockdale, J., Cassell, J., and Ford, E. ““ Giving something back ”: A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland”. In: *Wellcome Open Research* (2019).
- [48] United States. *The Health Insurance Portability and Accountability Act (HIPAA)*. Washington, D.C., 2004.
- [49] Verghese, A., Shah, N. H., and Harrington, R. A. “What This Computer Needs Is a Physician Humanism and Artificial Intelligence”. In: *JAMA - Journal of the American Medical Association* vol. 319 (2018).
- [50] Vokinger, K. N., Feuerriegel, S., and Kesselheim, A. S. “Mitigating bias in machine learning for medicine”. In: *Communications Medicine* vol. 1, no. 1 (2021), pp. 3–5.
- [51] Wiens, J. et al. “Do no harm: a roadmap for responsible machine learning for health care”. In: *Nature Medicine* vol. 25, no. 1 (2019), pp. 15–18.
- [52] Wu, R. and Wang, Q. “Shrinkage estimation for linear regression with ARMA errors”. In: *Journal of Statistical Planning and Inference* vol. 142, no. 7 (2012), pp. 2136–2148.
- [53] Xiao, C., Choi, E., and Sun, J. “Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review”. In: *Journal of the American Medical Informatics Association* vol. 25, no. 10 (2018), pp. 1419–1428.

- [54] Xu, W. et al. “A hybrid modelling method for time series forecasting based on a linear regression model and deep learning”. In: *Applied Intelligence* vol. 49, no. 8 (2019), pp. 3002–3015.
- [55] Yadav, P. et al. “Mining Electronic Health Records: A Survey”. In: vol. 1, no. 1 (2017), pp. 1–41. arXiv: 1702.03222.
- [56] Zhang, Y. et al. “Overweight is associated with better one-year survival in elderly patients after cardiac surgery: A retrospective analysis of the MIMIC-III database”. In: *Journal of Thoracic Disease* vol. 13, no. 2 (2021), pp. 562–574.

Chapter 3

Data sources

3.1 MIMIC III data

The Medical Information Mart for Intensive Care III, or MIMIC III [2] is a large, freely-available single center database that contains deidentified data on patients admitted to critical care units at Beth Israel Deaconess Medical Center, Boston, USA between 2001 to 2012. The data can be accessed and downloaded at <https://mimic.mit.edu> upon the completion of a privacy protection course.

The data types include, but are not limited to structured data such as temporal physiological signs, laboratory test results, static demographic information such as age and gender as well as unstructured data such as free text clinical notes. In this thesis I will focus on the structured temporal and sequential data. Recent works on reproducible studies using MIMIC-III data make it possible to extract consistent patient cohorts and features. In Paper I, I selected two cohorts from MIMIC III database for the experiments, and the selection criteria is in line with [1].

3.2 AHUS data

The AHUS data is a large EHR dataset provided by Akershus University Hospital (Akershus Universitet Sykehus), Norway. Akershus University Hospital is a tertiary hospital serving a population of 560,000 in the Greater Oslo Region. The use of data was approved by the Regional Committee for Medical and Health Research Ethics (Ref.no. 33192) and was considered exempt from patient consent requirements by the institutional Data Protection Officer (Ref.no. 2019/56). Pseudonymized data was extracted on December 6th 2019 by the AHUS Department for Data Extraction and Analysis. The study period is a 365-day period from June 2018 to May 2019. The study index time (T₀) is on a Monday in June 2018, however for personal privacy protection the exact week number is kept by the data extraction personnel and unknown to anyone else, including the researchers using the data. For data that contains temporal records, the granularity of time is hourly.

The extracted deidentified personal level data is stored in the TSD (Tjenester/Service for Sensitive Data), a platform for collecting, storing, analyzing and sharing sensitive data in compliance with the Norwegian privacy regulation. The platform is owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (USIT). (tsd-drift@usit.uio.no)

References

- [1] Johnson, A. E. W., Pollard, T. J., and Mark, R. G. “Reproducibility in critical care : a mortality prediction case study”. In: *2nd Machine Learning for Healthcare Conference, PMLR*. Vol. 68. 2017.
- [2] Johnson, A. E. et al. “MIMIC-III, a freely accessible critical care database”. In: *Scientific Data* vol. 3 (2016), p. 160035.

Chapter 4

Aims of the thesis

There are two closely related aims of the thesis. As EHR data exist in various formats, there is the need for specific tools for each type of EHR data. The first aim of the thesis is therefore to develop and experiment with different statistical and computational methods to extract, integrate and represent information from hospital EHR data. In this thesis I explore data mining algorithms (dynamic time warping), machine learning classification algorithms, network analysis on sequential relational data, regression problems, prediction and variable selection algorithms.

The second aim is to demonstrate the broad scope of applications of EHR data in the clinical setting. I used two very different hospital EHR datasets: the MIMIC III database from the US and the AHUS data from Norway. The applications are diverse, for instance, clinical event classification and prediction (paper I and III) could be used for patient risk stratification and early warning. Understanding patient transfer patterns and associated risk factors (paper II) could improve hospital management and logistic efficiency.

In **Paper I** I aim to develop a method to represent complex temporal EHR data, with the application of clinical event classification. I illustrate the use of data mining algorithms on heterogeneous unequal length temporal EHR data to extract features from ICU patients, and the classification performance using binary classification algorithms. A wide range of algorithms for this type of data are compared in the paper.

In **Paper II** I aim to explore the representation of sequential hospital EHR location logs via networks. I use network analysis, both statically and enhanced with time information, to investigate the patient flow in a large regional hospital in Norway. Compared to other studies that analyze patient flow, this paper includes time information (work day, week in the year) and additional patient level risk factors to identify the heterogeneity in different hospital departments.

In **Paper III** I aim to develop a method that balances predictability and interpretability, with intervenable prediction in a hospital setting as the target application. I demonstrate the use-case using aggregated EHR data at the ward level. This paper is a proof of concept that can be extended using various models and different applications.

The background information for each paper are provided in the chapters 5, 6, 7, and the summary of the papers is presented in chapter 8.

Chapter 5

Clinical event classification with temporal EHR data

5.1 Clinical event prediction and classification

Risk prediction and classification has been an important task for clinicians for decades. Prediction tasks vary greatly. The tasks of interest include but are not limited to patient length of hospital stay forecast, physiologic decline detection, phenotype classification, admission or readmission prediction, discharge prediction and risk of mortality prediction [6, 16, 17, 21, 27]. The purpose of developing predictive models using EHR data is to better understand patient conditions and devise a clinical decision support system that can assist clinicians on decision making.

Various criteria and metrics have been proposed to evaluate and assess the condition of patients admitted to intensive care units (ICU), such as SAPS (Simplified Acute Physiology Score) [15], SOFA score (Sequential or Sepsis-related Organ Failure Assessment) [25] and APACHE (Acute Physiologic And Chronic Health Evaluation) [14]. These criteria aim to measure the severity of disease by combining routine physiological measurements. Generally, a higher score implies greater chance of mortality. These activities have been replaced by modern statistical and machine learning algorithms nowadays.

EHR data are challenging to work with in research. Access to personal level EHR data for research is highly restricted due to privacy protection regulations; the data has questionable quality due to missing measurements and labels; the various formats where data are recorded makes it pre-processing a difficult but necessary task. However, thanks to the efforts in making a large EHR data available to the public [12], it is possible to start make some progress on this front. For the remainder of this chapter, I will focus on the temporal EHR data.

5.2 Temporal EHR data

EHR data of various formats are being collected on patients from their arrival to discharge. Based on sampling frequency, EHR data can be categorized into the following groups [4]:

1. High frequency (seconds): signals such as electrocardiogram ECG that are generated at bedside. This type of data are vulnerable against artifact corruption like sensor falling off and power grid interference. The data must

5. Clinical event classification with temporal EHR data

be aggregated and filtered before any learning tasks or feature extraction can take place.

2. Mid frequency (hours): vital signs, medications, procedures and lab results. Errors could occur due to human registration. Sometimes the time of measurement taken can be more informative than the actual value of the test results (e.g. a test being run at 2 am because the patient is very ill) [4].
3. Low frequency or static (once or by order): structured data (demographics, life-style information, underlying disease) and unstructured data (clinical notes in free text). For the free-text notes, natural language processing techniques need to be applied. However, misspelling, acronym-heavy, and domain specific terms across different diseases can be challenging without expert insight.

In clinical event classification, the most common type of temporal EHR data are mid and low frequency physiologic measurements. These are also the basis of classic disease severity scores [14, 15, 25].

5.3 Time series classification

The difference between time series classification and other classification problems is the additional time dimension, or the ordered sequences. Typically, the input of a classifier is the feature matrix where each row is one observation (or sample, subject), and each column is a feature. For time series or ordered sequences, the additional time dimension needs to be accounted for when constructing a feature matrix.

Time series classification algorithms can be broadly categorized into three groups [26]: feature based, model based and distance based methods. For the *feature based* methods, feature extraction based on e.g. a Fourier transform is done on the time series data, then a conventional classifier such as logistic regression or SVM is used on the extracted features. For *model based* methods, a data generation model is assumed (such as an auto-regressive AR model). The model that is the best fit for the test data is the class output. Finally, for the *distance based* methods, a similarity metric is defined, then a distance is computed for pairs of input series. The classification is done with a distance-based classifier like k-nearest neighbor algorithm.

5.3.1 Dynamic time warping

Dynamic time warping (DTW) is a technique to find the optimal alignment between two time dependent sequences, specifically with time deformation and different speed [13, 18]. Given two time series $\mathbf{x} = (x_1, x_2, \dots, x_N)$ and $\mathbf{y} = (y_1, y_2, \dots, y_M)$, construct a cost matrix $\mathbf{C} \in \mathbb{R}^{N \times M}$ with elements

$c_{n,m} = d(x_n, y_m)$. Here d is a distance measure. With squared Euclidean distance, $d(x_n, y_m) = (x_n - y_m)^2$.

A warping path $W = (w_1, \dots, w_Q)$ is a set of matrix indices that defines a mapping between \mathbf{x} and \mathbf{y} where Q is the length of the warping path. Let $w_1 = (1, 1), w_Q = (N, M)$, indicating that the warping path starts and ends in the opposite corner cells of the matrix (boundary conditions). W also needs to satisfy additional continuity and monotonicity conditions [13]. Let the total cost of a warping path W between \mathbf{x}, \mathbf{y} be

$$TC_W(\mathbf{x}, \mathbf{y}) = \sum_{q=1}^Q c_{w_q}, \quad (5.1)$$

The optimal warping path W^* is the one that minimizes the total cost among all possible paths, and the DTW distance is the total cost associated with W^* ,

$$\begin{aligned} DTW(\mathbf{x}, \mathbf{y}) &= TC_{W^*}(\mathbf{x}, \mathbf{y}) \\ &= \min\{TC_W(\mathbf{x}, \mathbf{y})\}. \end{aligned}$$

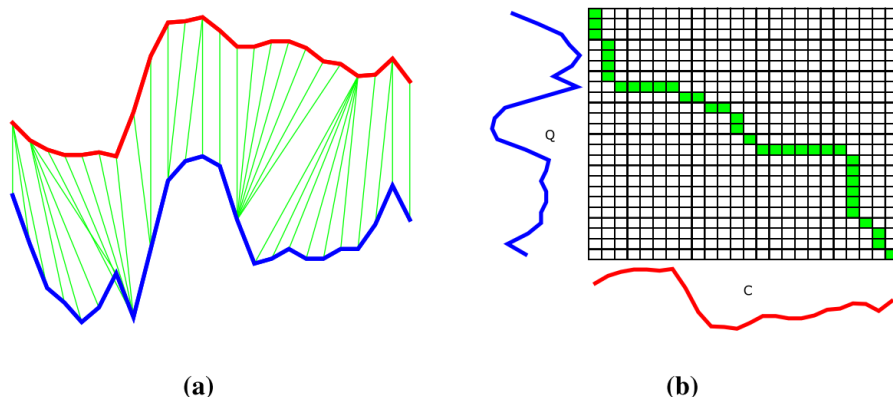


Figure 5.1: Illustration of DTW [Tan2019]. a: Alignment for two time series Q (blue) and C (red). b Cost matrix C with warping path W (green).

It is time consuming to find the optimal warping path. By restricting the difference between possible alignment indices between time series pairs, the search window is narrowed around the diagonal of the warping cost matrix. Two well known global constraints are the Sakoe-Chiba band [22] and Itakura parallelogram[11]. A comparison between these two constraints has been made by [3]. More recent works have investigated learning constraints from the data for faster computation and better accuracy [2, 19, 20, 23]. It is worth mentioning that constraints work well when the time series lengths do not differ much, otherwise the warping path might not exist[5].

5.4 Tensor decomposition

A tensor is a multidimensional array, the number of dimensions is called order, modes or ways. In this work we focus on third order tensors. A slice is a two dimensional section of a tensor with two fixed modes. For example $\mathbf{X}_{1::}$ is a horizontal slice, which is the first layer or top matrix of a tensor.

Table 5.1: List of notations

Symbol	Definition
$\mathbf{X}, \mathbf{D}, \mathbf{M}$	Matrix
\mathbf{X}^T	Matrix transpose
\mathbf{x}_r	r-th column of \mathbf{X}
\mathcal{X}, \mathcal{D}	Tensor
x_{ij}, x_{ijk}	Elements of a matrix and a tensor
$\mathbf{X}_{i::}, \mathbf{X}_{:j:}$	Horizontal, lateral slice of tensor
$\mathbf{X}_{::k}$ or simply \mathbf{X}_k	Frontal slice of tensor
\mathbf{x}, \mathbf{y}	Vector
\circ	Outer product

In this paper we focus on CANDECOMP/PARAFAC or CP decomposition for short. For a third order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, a CP decomposition for a chosen number of components $r = 1, \dots, R$ can be formalized in the following way:

$$\min_{\hat{\mathcal{X}}} \|\mathcal{X} - \hat{\mathcal{X}}\| \quad \text{where} \quad \hat{\mathcal{X}} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \quad (5.2)$$

Here $\mathbf{a}_r, \mathbf{b}_r, \mathbf{c}_r$ are column vectors of size I, J, K . The vectors can be reorganised into factor matrices $[\mathbf{A}, \mathbf{B}, \mathbf{C}]$ where $\mathbf{A} \in \mathbb{R}^{I \times R}, \mathbf{B} \in \mathbb{R}^{J \times R}, \mathbf{C} \in \mathbb{R}^{K \times R}, \mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \dots \quad \mathbf{a}_R]$. If the columns of $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are normalized to unit length, then the weights are absorbed into $\boldsymbol{\lambda} \in \mathbb{R}^R$,

$$\hat{\mathcal{X}} = \sum_{r=1}^R \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r. \quad (5.3)$$

Tensors can capture information with more than two dimensions, and tensor or matrix decomposition has wide applications in signal processing and data mining [1, 24], and has been applied successfully in healthcare informatics, see [7, 8, 9, 10].

References

- [1] Acar, E. et al. “Tensor-based fusion of EEG and FMRI to understand neurological changes in schizophrenia”. In: *Proceedings - IEEE International Symposium on Circuits and Systems* (2017), pp. 1–4.
- [2] Dau, H. A. et al. “Judicious setting of Dynamic Time Warping’s window width allows more accurate classification of time series”. In: *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017* (2017).
- [3] Geler, Z. et al. “Dynamic Time Warping: Itakura vs Sakoe-Chiba”. In: *IEEE International Symposium on INnovations in Intelligent SysTems and Applications, INISTA 2019 - Proceedings* (2019).
- [4] Ghassemi, M. et al. “Opportunities in Machine Learning for Healthcare”. 2018. arXiv: [1806.00388](https://arxiv.org/abs/1806.00388).
- [5] Giorgino, T. “Computing and Visualizing Dynamic Time Warping Alignments in R : The dtw Package”. In: *Journal of Statistical Software* vol. 31, no. 7 (2009).
- [6] Harutyunyan, H. et al. “Multitask Learning and Benchmarking with Clinical Time Series Data”. In: (2018). arXiv: [1703.07771](https://arxiv.org/abs/1703.07771).
- [7] Henderson, J., Malin, B. A., and Ho, J. C. “PIVETed-Granite : Computational Phenotypes through Constrained Tensor Factorization”. In: (2018). arXiv: [arXiv:1808.02602v1](https://arxiv.org/abs/1808.02602v1).
- [8] Henderson, J. et al. “Granite: Diversified, Sparse Tensor Factorization for Electronic Health Record-Based Phenotyping”. In: *IEEE International Conference on Healthcare Informatics (ICHI)*. 2017.
- [9] Ho, J., Ghosh, J., and Sun, J. “Marble : High-throughput Phenotyping from Electronic Health Records via Sparse Nonnegative Tensor Factorization”. In: *Kdd* (2014), pp. 115–124.
- [10] Ho, J. et al. “Limestone: High-throughput candidate phenotype generation via tensor factorization”. In: *Journal of Biomedical Informatics* vol. 52 (2014), pp. 199–211.
- [11] Itakura, F. “Minimum Prediction Residual Principle Applied to Speech Recognition”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* vol. 23, no. 1 (1975), pp. 67–72.
- [12] Johnson, A. E. et al. “MIMIC-III, a freely accessible critical care database”. In: *Scientific Data* vol. 3 (2016), p. 160035.
- [13] Keogh, E. J. and Pazzani, M. J. “Scaling up dynamic time warping to massive datasets”. In: *Principles of Data Mining and Knowledge Discovery* vol. 1704, no. Derriere (1999), pp. 1–11.
- [14] Knaus, W. et al. *APACHE II: A severity of disease classification system*. 1985.

5. Clinical event classification with temporal EHR data

- [15] Le Gall, J.-R., Lemeshow, S., and Saulnier, F. “Simplified Acute Physiology Score (SAPS II) Based on a European / North American Multicenter Study”. In: *JAMA* vol. 270, no. 24 (1993), pp. 2957–2963.
- [16] Lin, Y.-W. et al. “Analysis and Prediction of Unplanned Intensive Care Unit Readmission using Recurrent Neural Networks with Long Short-Term Memory”. In: *PLoS ONE* vol. 14, no. 7 (2019).
- [17] McWilliams, C. J. et al. “Towards a decision support tool for intensive care discharge: Machine learning algorithm development using electronic healthcare data from MIMIC-III and Bristol, UK”. In: *BMJ Open* vol. 9, no. 3 (2019), pp. 1–8.
- [18] Muller, M. “Dynamic Time Warping”. In: *Information Retrieval for Music and Motion*. Springer, Berlin, Heidelberg, 2007. Chap. 4, pp. 69–84.
- [19] Niennattrakul, V. and Ratanamahatana, C. A. “Learning DTW Global Constraint for Time Series Classification”. In: (2009). arXiv: [0903.0041](https://arxiv.org/abs/0903.0041).
- [20] Ratanamahatana, C. A. and Keogh, E. “Making time-series classification more accurate using learned constraints”. In: *SIAM Proceedings Series* (2004), pp. 11–22.
- [21] Ribas Ripoll, V. J. et al. “Sepsis mortality prediction with the Quotient Basis Kernel”. In: *Artificial Intelligence in Medicine* vol. 61, no. 1 (2014), pp. 45–52.
- [22] Sakoe, H. and Chiba, S. *Dynamic Programming Algorithm Optimization for Spoken Word Recognition*. 2. Morgan Kaufmann Publishers, Inc., 1990, pp. 159–165.
- [23] Salvador, S. and Chan, P. “FastDTW: Toward accurate dynamic time warping in linear time and space”. In: *Intelligent Data Analysis* vol. 11, no. 5 (2007), pp. 561–580.
- [24] Sidiropoulos, N. D. et al. “Tensor Decomposition for Signal Processing and Machine Learning”. In: *IEEE Transactions on Signal Processing* vol. 65, no. 13 (2017), pp. 3551–3582. arXiv: [1607.01668](https://arxiv.org/abs/1607.01668).
- [25] Vincent, J. L. et al. “The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure”. In: *Intensive Care Medicine* vol. 22, no. 7 (1996), pp. 707–710.
- [26] Xing, Z., Pei, J., and Keogh, E. “A brief survey on sequence classification”. In: *ACM SIGKDD Explorations Newsletter* vol. 12, no. 1 (2010), pp. 40–48.
- [27] Yoon, J. et al. “Forecast ICU: A prognostic decision support system for timely prediction of intensive care unit admission”. In: *33rd International Conference on Machine Learning, ICML 2016* vol. 4 (2016), pp. 2524–2538.

Chapter 6

Intrahospital patient transfer pattern

6.1 Hospital patient flow

Hospitals are complex systems that must run smoothly to optimize treatment quality and patient safety. Hospital patient flow, as the name suggests, is the physical movement or transfers of patients from one hospital unit to another. It is also referred to as hospital events [1] or patient pathways [7]. A typical chain of hospital events from patient admission to discharge takes the one of the following paths in figure 6.1, although it can differ across different hospitals.

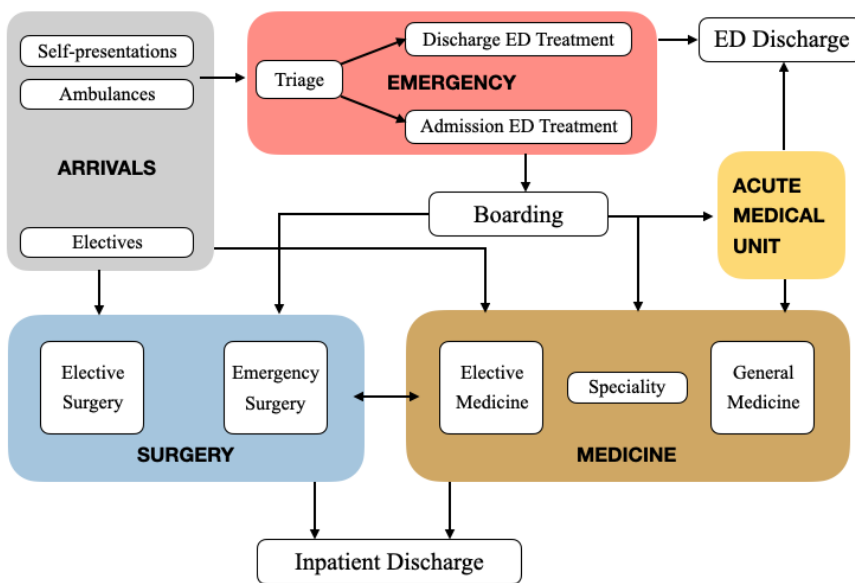


Figure 6.1: Hospital event from arrival to discharge. Figure adapted from [Ben-Tovim2016] after excluding EECU (Extended Emergency Care Unit).

Patient flow from admission to discharge is affected by a few factors, including time of arrival (weekday or weekend, during the day or night), triage types, admission rates, treatment times and resource capacities. There are five common types patients who have distinct needs[7]:

6. Intrahospital patient transfer pattern

1. Emergency: patient who are admitted from Emergency Department (ED), either on their own or by ambulance.
2. Surgical: patients who undergo surgical procedures.
3. Medical: patients who need medical treatment and do not undergo surgery.
4. Day-surgical or day-medical: patients who need minor surgical or medical procedures and do not require a hospital bed.
5. Out-patient: patients who visit the hospital for a few hours for medical or diagnostic visits.

The proper management of patient flow is critical to not only patient safety, but also healthcare worker wellbeing. Delays, queues and inappropriate placement of patients can directly affect patient safety and satisfaction, while inefficient capacity utilization and variability of workload can cause stress for healthcare personnel, which would in turn affect patient safety. These are patient flow problems either due to bad allocation or shortage of capacity, lack of coordination between different units, or presence of bottlenecks that could delay patient throughput [7]. Beyond capacity management, patient flow also plays a part in hospital infection control: patients can spread pathogens to other patients while being moved around; this in turn plays a part in capacity management as separating colonized patients require additional beds in single rooms or double rooms with other colonized patients [6]. An in-depth understanding of patient flow throughout the hospital is therefore key to hospital management and quality of care.

6.2 Network analysis

A short introduction to network terminology and analysis methods is covered in this section, before I put them into the context of hospital patient flow.

6.2.1 Terminology

A network graph $G = (V, E)$ is a mathematical structure that consists of a set of *vertices* V and *edges* E . Vertices are also known as nodes, and edges are also known as links. Elements of E are pairs $\{u, v\}$ of distinct vertices $u, v \in V$. The number of vertices $N_v = |V|$ and number of edges $N_e = |E|$ are called the order and size of the graph G .

The descriptive analysis of a network graph is focused on its structure and characteristics. A graph is *directed* if the vertices defining an edge are ordered, i.e. $\{u, v\}$ is distinct from $\{v, u\}$ for $u, v \in V$. Such edges are *directed edges* or *arcs*, with the direction from head u to tail v . In a directed graph there can be two arcs connecting a pair of vertices of the opposite directions. A graph is *weighted* if the edges have the weight attribute, i.e. their frequencies are of interest. A graph can be both directed and weighted as these are two independent edge

attributes. The *degree* of a vertex v , d_v counts the number of edges in E that links to v . If the graph is directed, the terms in and out degree d_v^{in} , d_v^{out} are used to distinguish the direction. Finally, the *density* of a graph is the frequency of realized edges relative to potential edges,

$$density(G) = \frac{|E|}{|V|(|V| - 1)/2} \quad (6.1)$$

The methods and metrics in graph theory and network analysis are extensive. For instance, finding the shortest path connecting any two vertices (*shortest path distance*); measuring how important a vertex is to the system (with different *centrality* metrics); discovering groups that are closely connected (*community detection* or *graph partitioning*). Nevertheless, these metrics and methods are not particularly relevant to our application in understanding hospital patient flows. This is further explained in the discussions (Chapter 9).

6.2.2 Network visualization

The visualization of a network is fundamental to understand the network structure in an intuitive manner. The simplest way is merely to draw the vertices randomly and connect the them with lines, although when the network is big, the graph can be unreadable or not informative. Drawing conventions, aesthetics and constraints have been proposed [2], and many layout algorithms have since been implemented for convenient network visualization. In figure 6.2 I illustrate the patient transfer network using AHUS data. The details can be found in Paper II.

6.2.3 Temporal network

Network has a natural temporal flavor to it: edges are formed consequently rather than all at once. Nonetheless, the majority of network analyses are focused on static networks, either due to the complexity brought by the time, or due to the availability of time-stamped relational data. A temporal or dynamic network is a time-indexed graph $G(t) = (V(t), E(t))$ where time t varies discretely or continuously, and $V(t), E(t)$ are vertices and edges present at time t .

The classic way to represent a network evolving over time is through the construction of an *aggregated static graph*. To put in simple terms, t is divided in consecutive non-overlapping intervals and one graph is constructed for each interval. This is equivalent to taking snap-shots of the whole, aggregated network over time. Other analyses can be carried out on each graph. It is common to compute some network metric, e.g. clustering coefficient for each interval then plotted over time to understand how network evolves during 0-12 hours, 12-24 hours and so on. Alternatively, visualizations can be created at each time interval for a qualitative comparison.

Temporal motif is another way of understanding temporal networks. A *motif* or *graphlet* is a small subgraph occurring in a network [4]. In a temporal network,

6. Intrahospital patient transfer pattern

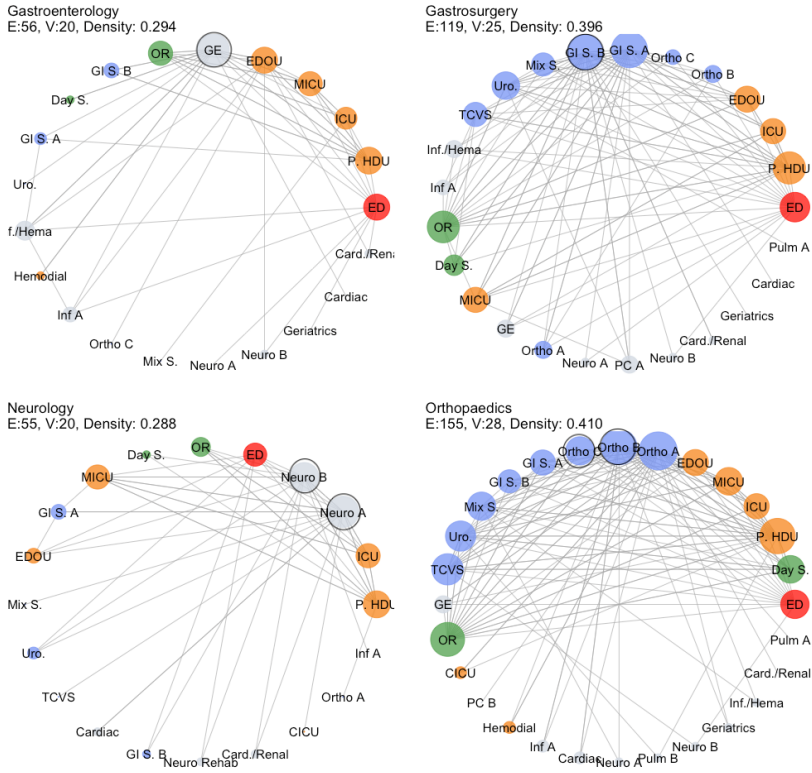


Figure 6.2: Patient transfer network illustration. For details and abbreviations see Paper II.

motifs are used to study the temporal and topological structure of the network. As per [5], a k -node, l -edge motif is a sequence of l edges:

$$M = (u_1, v_1, t_1), (u_2, v_2, t_2), \dots, (u_l, v_l, t_l) \quad (6.2)$$

where $t_1 < t_2 < \dots < t_l$. Intuitively, motifs are time-stamped and therefore ordered edges involving a small number of vertices. A real life example of such motifs is reciprocation of telephone calls in a social network: person A calls B, B calls C, etc. Figure 6.3 shows some simple examples of motifs. The study of temporal motifs is relatively recent, with a large focus on counting motifs efficiently in a large networks [3, 5].

6.3 Patient flow from EHR location logs: a network approach

In this section I describe the two perspectives to understand patient flow using EHR data, first by examining the system itself (hospital operation), then by examining the individual actors in the system first (patients or healthcare workers).

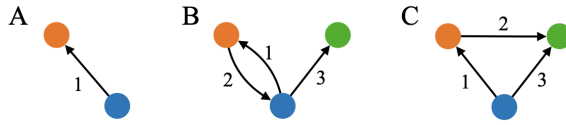


Figure 6.3: Examples of network motifs. Numbers are the ordering of edges.

The EHR data contain rich information to study the patients and hospital operation in detail, as it is possible to link time stamped events (operation procedure, measurements, medication) with patient whereabouts. The terms *flow*, *transfer*, *movement* are used interchangeably. The following information are required to produce patient flow network:

- Location for each patient as vertices;
- Time when each transfer happens;
- Specify time granularity (e.g. weekly, daily) and level of aggregation (e.g. patient level, departmental level, hospital level).

Other patient attributes can be attached to each transfer. For brevity, I only focus on the network construction here.

6.3.1 System-wide analysis

The system-wide analysis is essentially the structural analysis of the patient transfer network. The transfer network construction depends on the question of interest: are we interested in all the possible connections and system bottlenecks? A static network would be sufficient to answer this question. If we want to identify how the connection changes in the monthly or weekly scope, it becomes necessary to keep the time information. Direction of each transfer is also an important aspect.

In Paper II, the *system* is each one of the four departments in the hospital. They are very different in terms of service types (both surgical and medical). For the static networks we ignored the time of each transfer and constructed one network for each department for all transfers that happened in the one-year study period. To keep the time information, we have constructed the weekly networks; we also kept weekday to explore the difference between the work days (Monday to Friday) and weekends (Saturday to Sunday). Metrics applied to both static and temporal transfer networks are: size in terms of edge and vertex counts, connectedness in terms of density, vertex functionality and importance in terms of in- and out-degree. For temporal network metrics, mean and standard deviation are reported.

We also investigate certain types of transfer by categorizing all transfers into groups: Bed-to-Bed, ED-to-Any, and others. By averaging the number of

transfers per patient in each department, it can reveal the differences in the types of service provided for patients and the complexity of hospital system.

6.3.2 Patient-based analysis

The patient-based analysis using network is a form of *temporal motif* analysis. The previous system-wide analysis breaks the complete trajectory of individual patients and only the transfer itself is important. Here, the complete patient trajectory of interest.

We investigate the two extremes of transfer trajectories: the most common and the least common trajectories. In motif analysis, counts or frequency of certain motifs are the fundamental quantities of interest. These are also what we try to quantify in our study. The results can inspire further analysis on the potential causes for exceedingly long trajectories and if some transfers could be avoidable.

References

- [1] Ben-tovim, D. et al. “Simulation Modelling Practice and Theory Hospital Event Simulation Model : Arrivals to Discharge – Design , development and application”. In: *Simulation Modelling Practice and Theory* vol. 68 (2016), pp. 80–94.
- [2] Di Battista, G. *Graph Drawing: the Aesthetics-Complexity Trade-Off*. Berlin Heidelberg: Springer Berlin Heidelberg, 1999, pp. 92–94.
- [3] Kovanen, L. et al. “Temporal Motifs”. In: (2013), pp. 135–159. arXiv: 1108.1780.
- [4] Milo, R. et al. “Networks Network Motifs : Simple Building Blocks of Complex Networks”. In: vol. 824, no. 2002 (2008).
- [5] Paranjape, A., Benson, A. R., and Leskovec, J. “Motifs in temporal networks”. In: *WSDM 2017 - Proceedings of the 10th ACM International Conference on Web Search and Data Mining* (2017), pp. 601–610. arXiv: 1612.09259.
- [6] Shenoy, E. S. et al. “A Discrete Event Simulation Model of Patient Flow in a General Hospital Incorporating Infection Control Policy for Methicillin-Resistant Staphylococcus Aureus (MRSA) and Vancomycin-Resistant Enterococcus (VRE)”. In: vol. 38, no. 2 (2018), pp. 246–261.

- [7] Villa, S., Prenestini, A., and Giusepi, I. “A framework to analyze hospital-wide patient flow logistics : Evidence from an Italian comparative study”. In: *Health policy* vol. 115, no. 2-3 (2014), pp. 196–205.

Chapter 7

Intervenable prediction for hospital operation

7.1 Prediction and intervention in healthcare environment

Prediction is an important task in healthcare. An accurate prediction of the target quantity can facilitate good planning to utilize available resources. For instance, in [3] the platelet usage was predicted to reduce waste due to the short storage time of blood product, and in [6] the weekly average medicine expenditure of patients was predicted to help with healthcare expenditure planning. Aside from the prediction itself, it is often interesting to find out what factors are contributing to the predictions. In a patient-centered prediction model, possible risk factors are age, sex, lifestyle indicators such as smoking drinking. Sometimes time factor represented by weekday or weekend can also provide additional information[3, 8].

There exist a wide range of predictive models including both statistical and machine learning approaches. Despite the popularity of complex machine learning methods, regressions remain one of the most commonly used models. Regression models have a simple formulation that encourages interpretation, and their theoretical properties have been studied extensively. Moreover, the development of penalized regression models made it convenient to include thousands of variables at the same time. Some forms of penalties also produce sparse regression coefficients, which act as a variable selection procedure.

In healthcare, it is natural to ask questions about intervention. What would be the effect of changing one variable? However this is not a trivial problem to solve. Randomized controlled trials are the golden standard to measure intervention effects, yet not all intervention strategies are suitable for experiments. The adoption of electronic health records in healthcare makes available a huge quantities of information, which could enable the design of intervention strategies and measure effects. One possibility is to build a predictive model that is accurate as well as interpretable, and use it to guide further intervention studies. This is what we are trying to do in Paper III.

7.2 Regression with sparsity constraints

In this section I provide a brief introduction to the regression problem with sparsity constraints. Without loss of generality, only linear regression is discussed.

We observe N observations of some response variable y_i and p predictor variables $x_i = (x_{1i}, \dots, x_{pi})$, $i = 1, \dots, N$. A linear regression model assumes the following relationship:

$$y_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j + \epsilon_i \quad (7.1)$$

where β_0 and $\beta = (\beta_1, \dots, \beta_p)$ are unknown parameters (regression coefficients) and ϵ_i is an error term. In standard regression models, the errors are assumed independent. The intercept β_0 can be conveniently folded into the β by adding a **1** column into the design matrix X . The least squares estimates for the parameters optimize the objective function

$$\min_{\beta} \sum_{i=1}^N (y_i - \sum_{j=0}^p x_{ij}\beta_j)^2. \quad (7.2)$$

7.2.1 Variable selection

The motivation of variable selection is to produce a more parsimonious model that contain fewer predictors to enhance the interpretability, reduce the variance of the model and improve prediction. Broadly speaking, there are two categories of variable selection algorithms: the filter methods and the wrapper methods [4, 12]. The filter methods rank the importance for each variable first, then combine the selected variables together. They are fast, however they risk missing out interactions and selecting highly correlated variables. Wrapper methods search for combinations of variables that maximize model performance, which makes them computationally intensive.

In a low dimensional regression setting where $n \gg p$, variable selection can be done with backward selection, forward selection or a mix of both. Sometimes domain knowledge plays a part in the variable selection as well. Discussions on variable selection in low dimensional regression can be found in [1, 5]. When $n < p$, the more appropriate solution is to use regularization.

7.2.2 Sparsity via the lasso

The lasso (Least Absolute Shrinkage and Selection Operator [9]) is a widely used method for regularization. The optimization problem of least squares becomes

$$\min_{\beta} \sum_{i=1}^N (y_i - \sum_{j=0}^p x_{ij}\beta_j)^2 \quad \text{subject to } \|\beta\|_1 \leq t \quad (7.3)$$

where $\|\beta\|_1 = \sum_{j=0}^p |\beta_j|$ and t is a positive bound. Writing $\mathbf{y} = (y_1, \dots, y_N)$ as the vector of the response and X as the $N \times p$ design matrix ($p+1$ if including the intercept term), this optimization problem is commonly expressed in the Lagrangian form:

$$\min_{\beta} \left\{ \|\mathbf{y} - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\} \quad (7.4)$$

for some $\lambda \geq 0$.

The tuning parameter λ controls the amount of shrinkage on the regression coefficients. The larger λ is, the heavier the penalty becomes and as a result, some coefficients become 0. When $\lambda = 0$ the coefficients are the same as least squares estimates. The choice of λ is typically made by cross-validation.

7.2.3 Other regularizations

A wide range of penalties exist in the literature for regression problems. A few well known examples include the adaptive lasso [13] that provides the "Oracle Property", i.e. consistent variable selection; and the elastic-net [14] that encourages strongly correlated variables to be in or out of the model together. The R package `glmnet` [2] provides a fast and easy-to-use implementation of the elastic-net.

Some slightly more complex penalties allow investigators to include domain knowledge, beliefs or preference on the predictors. For example, one might hope that some variables have non-negative sign [7]. Sometimes it could make sense to group variables together based on the functionalities and real-world implications. The group lasso [10] imposes a group structure on the predictors, where a subgroup of predictors are either all non-zero, or all shrunk to zero. Hierarchy between variables can also be imposed via overlapping groups [11].

7.3 Intervenable prediction with simulation and EHR data

In this project we propose a framework for intervenable predictions in the hospital setting. The key concept is the balance of predictability and interpretability, achieved via a two step regression approach. In the first step we use a linear penalized regression model: the purpose is to select variables to explain the outcome. If all explanatory variables that generate the response variable are measured, and are included in the regression model in the correct way, the prediction would not be far off; however in reality it is hard to know the true data generating mechanism, or to measure all the relevant variables. This is the motivation for the second step after the initial regression and variable selection. We use the residuals as the new response, and use the original lagged response as the explanatory variables in the second step. This is equivalent to an autoregressive model. By combining step one which explains the outcome using

explanatory variables, and step two that captures the unexplained components, our method balances interpretability and prediction accuracy.

I illustrate our method using two sets of simulation studies. In the first set of simulations I use a combination of simulated periodic and non-periodic time series as explanatory variables to generate the outcome time series. In the second simulation study, due to the lack of outcome variable in the AHUS dataset, I simulate the pseudo-infection metrics using the real hospital EHR series as explanatory variables. In both examples I illustrate the variable selection in step one, and prediction improvement after step two.

References

- [1] Desboulets, L. D. D. “A review on variable selection in regression analysis”. In: *Econometrics* vol. 6, no. 4 (2018), pp. 1–27.
- [2] Friedman, J., Hastie, T., and Tibshirani, R. “Regularization Paths for Generalized Linear Models via Coordinate Descent”. In: *Journal of Statistical Software* vol. 33, no. 1 (2010), pp. 1–22.
- [3] Guan, L. et al. “Big data modeling to predict platelet usage and minimize wastage in a tertiary care system”. In: *Proceedings of the National Academy of Sciences of the United States of America* vol. 114, no. 43 (2017), pp. 11368–11373.
- [4] Guyon, I. and Elisseeff, A. “An introduction to variable and feature selection”. In: *Journal of Machine Learning Research* (2003).
- [5] Heinze, G., Wallisch, C., and Dunkler, D. “Variable selection – A review and recommendations for the practicing statistician”. In: *Biometrical Journal* vol. 60, no. 3 (2018), pp. 431–449.
- [6] Kaushik, S. et al. “AI in Healthcare: Time-Series Forecasting Using Statistical, Neural, and Ensemble Architectures”. In: *Frontiers in Big Data* vol. 3, no. March (2020).
- [7] Owrang, A., Bresler, Y., and Jansson, M. “Model selection with covariance matching based non-negative lasso”. In: *Signal Processing* vol. 170 (2020).
- [8] Swierkowski, P. and Barnett, A. “Identification of hospital cost drivers using sparse group lasso”. In: *PLoS ONE* vol. 13, no. 10 (2018), pp. 1–19.
- [9] Tibshirani, R. “Regression shrinkage and selection via the Lasso”. In: *Journal of the Royal Statistical Society B* vol. 58, no. 1 (1996), pp. 267–288.

-
- [10] Yuan, M. and Lin, Y. “Model selection and estimation in regression with grouped variables”. In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* vol. 68, no. 1 (2006), pp. 49–67.
 - [11] Zhao, P., Rocha, G., and Yu, B. “The composite absolute penalties family for grouped and hierarchical variable selection”. In: *Annals of Statistics* vol. 37, no. 6 A (2009), pp. 3468–3497. arXiv: [0909.0411](https://arxiv.org/abs/0909.0411).
 - [12] Zhu, Z., Ong, Y. S., and Dash, M. “Wrapper-filter feature selection algorithm using a memetic framework”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* vol. 37, no. 1 (2007), pp. 70–76.
 - [13] Zou, H. “The adaptive lasso and its oracle properties”. In: *Journal of the American Statistical Association* vol. 101, no. 476 (2006), pp. 1418–1429.
 - [14] Zou, H. and Hastie, T. “Regularization and variable selection via the elastic net”. In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* vol. 67, no. 5 (2005), p. 768.

Chapter 8

Summary of the papers

Paper I: Feature extraction from unequal length heterogeneous EHR time series via dynamic time warping and tensor decomposition.

Published in *Data Mining and Knowledge Discovery* (2021) 35:1760-1784.
DOI: 10.1007/s10618-020-00724-6

Temporal EHR data is challenging to work on, as the time series for each clinical feature may be measured at different time frequencies, some more often (e.g. heart rates) and some less often (e.g. test results). In this paper we propose a combination of dynamic time warping (DTW) algorithm and tensor decomposition to represent the unequal time series such that subsequent classification tasks can be performed. The representation feature matrix is then fed into a classifier such as logistic regression or support vector machine to carry out clinical event classification. We demonstrate our method on two cohorts extracted from an openly accessible database, the MIMIC III, aiming to classify in-hospital mortality. Our method yields promising results compared to other machine learning algorithms commonly used for this application.

Our method is an alternative approach to tackle the classification of multivariate time series beyond deep learning and specifically, recurrent neural networks, which have been the most widely used methods for temporal EHR data classification in recent years. We demonstrate the success of the classic time series mining algorithm (DTW) in the field of healthcare EHR data, which might inspire further method developments using distance based methods.

Relating back to the thesis aim, this paper uses an existing data mining algorithm, originally for speech recognition, for a new type of data (temporal EHR) and yields interesting results. This is an addition to the existing literature on clinical event classification which focus primarily on neural networks, and can be extended to other applications.

Paper II: Effects of patient-level risk factors, departmental allocation and seasonality on intrahospital patient transfer patterns - network analysis applied on a Norwegian single-centre dataset. Accepted for publication in *BMJ Open*(2022) 0:e054545. DOI: 10.1136/bmjopen-2021-054545.

This paper studied the hospital patient transfer patterns at both system (clinical department) level and individual patient level using Electronic

8. Summary of the papers

Health Records (EHR). Understanding intrahospital transfer pattern is a crucial step for hospital capacity management and infection control. As far as we know, this is the first study that investigates different types of transfers in a variety of departments, both surgical (gastrointestinal surgery and orthopaedics) and medical (gastroenterology and neurology). This is also the first study on patient transfer pattern in a Nordic hospital. With the detailed description of patient movements, the paper illustrates the heterogeneity among different types of patients. We quantified the types and frequencies of transfers, and extracted the most common ones at the department level. In addition, we associated number of transfers with clinical risk factors to quantify the risk of excessive transfers. These results are relevant for evidence based method development using simulations to model patient flow, capacity management and infection control, as we show the complexity of hospital systems and the necessity to distinguish different categories of patients.

From a methodology perspective, we applied graph theory and methods in network science. In recent years network analysis has become popular in patient flow analysis, although the focus has been on the system as a whole over an entire period of time; in other words, static networks. Instead, in our work we also analyzed the evolution and difference of the networks in each week, and we compared the network differences between weekdays to weekends. We also utilized the notion of network motifs to describe the most common transfer types for each patient.

This paper complements the thesis by providing detailed examination of an interesting type of sequential relational EHR data: the location transitions. It provides additional knowledge by quantifying how different patients move within a healthcare facility using graph theory and network analysis, which can be used for future method developments and simulations.

Paper III: Intervenable predictions of hospital operation using Electronic Health Records [*Submitted to BMC Medical Informatics and Decision Making*]

This paper presented a novel framework for predictions in a healthcare setting, with potential applications in hospital operation optimization using Electronic Health Records. We introduce the term "intervenable prediction" in this paper, which is a two step approach that balance interpretability and prediction accuracy. This serves as the basis for possible intervention strategies.

In the first step, a penalized linear regression model that induces sparsity is fitted to the data, which attempts to select variables (risk factors) that can predict the outcome reasonably well. Then in the second step, we use the residuals from the previous step as the new response, and use the lagged response variable as the new predictors to fit a second regression model. In

this way, the model captures information in both the explanatory variables and the hidden trend in the response itself. We illustrate the method with simulation studies using both purely synthetic data and real EHR data from a Norwegian hospital. We demonstrate the different scenarios in which our method could improve the prediction using the two step approach, and effects of possible intervention strategies. The method can be easily extended to more complex models beyond linear models.

Chapter 9

Discussion

9.1 Overview

The thesis constitutes of three distinct yet related papers. The common theme and type of data is hospital electronic health records. Due to the development of electronic devices that collect patient data and digitalisation of health registries, EHR contain a much broader range of data in different formats compared to traditional healthcare data. This creates more opportunities for researchers to study new problems, but also brings new methodological challenges. This is an interdisciplinary field that requires knowledge in databases, computer science, statistics and medicine, which makes working with EHR data an inspiring yet challenging learning experience.

This thesis is my attempt to explore some of the many possibilities that hospital EHR data brings. The three papers have different focuses. Some are more focused on the data representation: specifically, how to make complex data into useful formats that can shed light on the patients (paper I, II); some are more focused on the potential application of streaming EHR data in real time to improve hospital logistic operations (paper III). Some are more experimental and based on statistical and machine learning algorithms (paper I, III) while others are more exploratory and require medical expert insights (paper II).

From the methodological perspective, this thesis contains not only statistical methods but also graph theory and data mining algorithms. This is not surprising given the richness and complexity of EHR data. I have used two single center hospital EHR (one American, one Norwegian) that are similar in terms of data collection (patient measurements with time stamps). However, one important difference between these two datasets is the way data were anonymised. Therefore it becomes infeasible to carry out certain analyses on the American data. The choices of methodology are naturally based on the questions of interest as well as types of variables at hand, such as the unique sequential relational data with consistent time stamps.

In the next three sections I will discuss each paper in detail. Each section covers contributions and findings, methods, challenges and limitations. In the end I will discuss the future perspectives that are common for all papers.

9.2 Paper I

9.2.1 Contribution and findings

The paper *Feature extraction from unequal length heterogeneous EHR time series via dynamic time warping and tensor decomposition* presents a novel method for temporal EHR data classification. Temporal EHR data is complex to work on as the feature time series may be measured at different time points for different periods of time. Such unequal length time series needs special algorithms to represent before they can be fed into any classification algorithms. We proposed a combination of dynamic time warping and tensor decomposition to tackle this issue. The DTW algorithms computes a distance-based metric for each feature that is further fed into some form of classifier (logistic regression, support vector machine). The distance matrices are used in a novel way for classification: the proposal of pivot patient makes sure that all features are considered together in contrast to the typical use of classifying each feature alone and then aggregate.

Our method produces classification results that are better than or equal to other state-of-the-art machine learning algorithms for this type of classification problems. We found that the representation extracted by a global DTW (computed on the entire sequence) do not differ significantly from constrained DTW. Among the classifiers we experimented with for the DTW representations, logistic regression has been the best performing one compared to more complex classifiers.

Implications: our method is an alternative approach to tackle the classification of multivariate time series beyond deep learning and specifically, recurrent neural networks, which have been the most widely used methods for temporal EHR data classification in recent years. We demonstrated the success of the classic time series mining algorithm (DTW) in the field of healthcare using temporal EHR data, which might inspire further method developments using distance based methods. The use of a popular publicly available dataset, the MIMIC III data and a standardized data processing pipeline suggests that the result is reproducible.

9.2.2 Methods, challenges and limitations

EHR time series as features for a classification problem

There exist various strategies to work on temporal EHR data as inputs for classification problems. In our paper we used a combination of a distance based method and tensor decomposition to lower the dimension of the original input data. The time dimension of each feature is compressed into a time-less distance matrix and stacked into a tensor, then the three dimensional tensor is decomposed, where one of the feature matrices is used for further dimension reduction. The end product is a matrix where each row is one observation (sample), which is used as inputs for further classification. Reference [49] is

closely related to our work in the sense that they also apply DTW algorithm to unequal length physiological series to compute the distance matrix. The difference is that they stopped at the stack of distance matrices and used a KNN classifier instead. Both solutions seem to work decently in terms of final classification.

Apart from utilizing a distance metric for signals, there are two directions for making use of EHR time series as features: one direction is to simplify the input by summarizing time series into summary statistics such as mean, standard deviation, min and max and skewness[23, 25, 31]. Another direction is to use a non-linear transformation such as recurrent or convolutional neural networks to transform the three dimensional input (patient by feature by time) into a two dimensional feature matrix [25, 49, 58]. There is no conclusion as for which representation is the best in general as there are limited data of this kind for large scale experiments.

DTW for multivariate features

Time series classification using dynamic time warping algorithms is most common in univariate problems where there is only one feature for each sample. When there are more than one feature per sample, a few solutions are possible: computing one distance matrix for every feature and stack them into a three dimensional tensor; summing every distance matrix together (the "independent DTW" [63]); or compute the DTW warping path on a cumulative distance considering all features together (the "dependent DTW" [63]). The first solution (distance tensor) is what our method and reference [49] adopted, and it is essentially an independent DTW. According to [63], whether independent or dependent DTW is superior depends on the data in application.

DTW constraints

Dynamic time warping (DTW) is a type of elastic measures. As discussed by [66], elastic measures are typically time consuming to compute, especially when collected in an ensemble for Nearest Neighbor classification. They have proposed a fast solution for the Elastic Ensemble within the Hierarchical Vote Collective of Transformation based Ensembles (HIVE-COTE), which is the state-of-the-art time series classification algorithm [38, 39]. Most variations of DTW try to tackle the issue of singularity: one point can be assigned to align with many consecutive points on the other series, leading to over compression[2]. A detailed analysis on the optimal warping path of DTW was provided in [18] where they proposed a sliding window DTW and time alignment measurement to tackle the issue of singularity. Some other alternatives are the continuous DTW [50], weighted DTW [2, 30] and DTW based on Mahalanoibis distance for multivariate time series [46].

Choice of classifier

After the representation of original temporal EHR data into a feature matrix,

the choice of classifier is rather trivial. In our experiments we have found that logistic regression (LR) is the best performing classifier. It is interesting that LR is frequently among the best performing classifiers for binary classifications despite its simplicity [25, 31]. LR is also a typical choice of classifier after neural network representations. The representation stage, either through simple summary statistics (max, mean et cetera) or complex deep representations or distance-based representations (our approach), seem to be the vital stage.

Use of pivot patient and distance matrix

In this paper we proposed the use of a pivot patient after we computed the distance tensor. By using the pivot patient, we make sure that all the rest of the observations in the feature matrix used for classification have a DTW distance based on the same patient. This is one way to encourage the interpretability of the feature matrix. The choice of pivot is made by optimizing the cross-validation error.

This is a novel use of DTW distance matrices. In the classical DTW or other distance-based time series classification frameworks, the distance matrix, say for feature A, is directly fed into a distance-based classifier such as KNN [71]. In the case of multivariate features, one can either sum up the independent DTW distances on each feature into one total distance matrix then do KNN (as suggested in [63]), or do the other way around: do KNN feature by feature, produce a classification result for each feature, then combine them together in some way [49]. Our method by using a pivot patient makes sure that all features are considered together, hence is an interesting alternative of using distance matrices in the multivariate scenario.

Interpretation

There are two ways to look at the interpretation of our method. The first is the patient-wise distance and the second is the factor loading post decomposition. A distance matrix is generally symmetrical although this might not be the case for DTW distances. Nonetheless, it consists of the distance between any two patients for a specific feature. When a pivot patient is fixed, this ensures that the input matrix for the classifier are distances or transformed distances against the same patient.

The factor loading of the tensor is another angle of looking at interpretations. In the literature of tensor based phenotyping [28, 29, 56] the decomposed three factors are often a way of interpreting the components that makes up the corresponding disease phenotype. For our method the interpretation is slightly different, since the tensor itself contains distances instead of the original values for each patient. The interpretation of a specific component (one column of the factor) would therefore be the relative dis-similarity of all features in this component.

Computational cost

Classification problems generally have the time complexity that depends on the learning phase and data size. If a non-distance based method is used, then the computation is mostly during the training stage, not the test stage where only prediction is made. If a distance-based method is used, the distance of the test set with a training sample also needs to be computed. This is true for a KNN classifier, which requires computing the distance of the new test samples with the training samples.

For a classification of time series or sequential data using distance-based methods, the situation is more complicated. The cost of the distance measure plays a part: using a lock-step measure such as Euclidean distance (i -to- i correspondence of both input series) is cheaper than an elastic measure (one-to-many) such as DTW, where the optimal warping path needs to be found [1]. DTW has quadratic complexity, and depends on the size of the longest series. Using constrained DTW can possibly reduce the computational cost by restricting the possible path.

Prediction or forecast

The distinction between prediction and/or classification and forecast is that the latter is yet to happen in the future. In principle our method can be extended to make forecasts, however the requirement of computing the distance matrix between test samples and pivot could make it time consuming and unpractical for forecasting tasks. Our method is more suitable for a first step towards deeper understanding of how patients differ. DTW-KNN type methods also have this issue. On another note, the Long-Short Memory Neural Networks (LSTM) or other deep learning methods that only take first 24 or 48 hours for training are designed to produce early warnings, hence are more suitable for forecasting.

Generalization to other datasets

In our experiments we made use of the MIMIC III [32] database. Specifically, we used two sub-cohorts: the sepsis cohort and the acute kidney injury cohort. The cohort definition and extraction algorithm is defined in [31] where the authors have released the code for reproducible cohort creation. Nonetheless, there could be some overlaps between these two cohorts, and due to the lack of other publicly available EHR data of similar type, we do not yet know how well the method will generalize.

9.3 Paper II

9.3.1 Contributions and implications

The paper *Effects of patient-level risk factors, departmental allocation and seasonality on intrahospital patient transfer patterns: network analysis applied on a Norwegian single-centre dataset* provides a detailed analysis of the intrahospital patient flow pattern in a large Norwegian hospital during a one-year period in 2018-2019. This paper studied the hospital patient transfer patterns at both

system (department) level and individual patient level using Electronic Health Records (EHR). As far as we know, this is the first study that investigates different types of departments: surgical (gastrointestinal surgery and orthopaedics) and medical (gastroenterology and neurology). This is also the first study on patient transfer patterns in a Nordic hospital. The Paper II is based on data pre-COVID19 pandemic and is therefore informative on how the hospital operates during normal periods.

Understanding the transfer patterns at healthcare facilities is crucial for patient safety. The two most important aspects at play are infection control and hospital capacity management. Various studies have attempted to address these two matters separately. For instance, authors of [12, 14, 57, 69] discussed the transfer pattern associated with hospital acquired infections, and some also address hospital ward planning [44]. When it comes to capacity management, the more optimized and efficient the hospital operates (smooth patient flow), the less loss of time incurs. This suggests not only better patient safety but also reduced costs as well as better staff workload planning[8]. One particular aspect to measure capacity, the emergency department (ED) process has been studied extensively, mostly using simulations [5, 10, 45, 51, 64]. Nonetheless, the simulation studies typically focus on only selected departments and wards, and only assign parameters to represent and model the overall processes. This abstraction can overlook important differences across different types of wards. In this paper we identified the heterogeneity of different hospital wards during different time periods, which is an important step towards evidence based planning and realistic simulations, for both infection control and capacity management. This is an important contribution to system level quality improvement [43], which can also inspire healthcare management education[3].

From a methodological point of view, this paper applies network analysis on time-stamped EHR location logs. In addition to visualizing all the transfers together, it is one of the few studies that analyze temporal networks by aggregated static graphs over time, and it is the first that analyses the complete patient trajectory using temporal motifs. Compared to other studies that use discrete event simulation [74], or multi-state models for patient processes which typically focus on the abstraction of activities or wards, network analysis has a different focus: on description, not prediction or modeling.

9.3.2 Findings

Network analysis on patient transfers

We found that transfer network patterns vary greatly across patient types. Specifically, patients in surgical departments (gastrointestinal surgery, orthopedics) have greater network size: more wards (vertices) are involved, and more transfer types (edges) occur compared to medical departments. This is expected as operation rooms are involved. In addition, many seemingly unrelated wards with respective departments are connected as well. This illustrates the

complexity of hospital systems where all locations are inter-connected. This phenomenon is well established as in [4, 34, 65].

We enriched network with time information, by constructing aggregated weekly networks. Within each week, we also differentiate by work days (Monday to Friday) and weekends (Saturday and Sunday). We found that the network sizes differ greatly in the one year period. Apart from [4], other studies have not used the time information. Seasonality plays a big role in large volume of surgical transfers: in summer season there are less scheduled surgeries.

Transfer patterns at the individual level

We looked at the transfer patterns at the individual level from different angles. Temporal motifs (in the form of transfer chains) have been extracted and analyzed. We have identified and quantified the most common types of transfers. These findings can be used as additional evidence for future simulation models such as in [26, 40, 62]. The seasonal and weekly patterns could be used for detailed ED process simulations [70] and hospital occupancy and discharge simulations [59].

We have found that the vast majority of transfers are short (two or three transfers). This is the basis for our investigations of the risk factors associated to many number of transfers, as we were interested to know if the long transfer chains can be characterized and eventually prevented. Compared to [9] that investigated the relationship between intrahospital transfers and hospital acquired infections, we did not have infection outcome in our study. Nonetheless, we argue that the causal direction between number of transfers and infection risk is not clear-cut. Better questions and careful variable constructions are necessary to investigate this further.

9.3.3 Methods, challenges and limitations

Network analysis as the choice of model

The patient transfer study is partly inspired by the availability of time-stamped patient location data. Network is a natural choice to understand relational data, such as the patient flow within the system (hospital or department, in our study). In studies that apply network analysis on intrahospital transfers[4, 34, 65], it is common to use all wards in the whole hospital as vertices to illustrate the connectivity and bottlenecks in the system. For our study, we did not have access to data at the whole hospital. However, what differentiate our work from the rest is that we stratify by departments, as each department has different features.

It is not uncommon to make use of network analysis for healthcare related studies in recent years. Networks for single hospital settings typically focus on patient mobility associated with infections (*Clostridium difficile* is often the target) [12, 44]. Some studies extend the networks to many healthcare facilities at a national level [17]. Some other types of networks such as healthcare

collaboration [13, 16] are less common. Network analysis as the choice of model allow the investigators to look at the hospital (hospitals) at a system level, which can be quite convenient for some applications. For example, in [17], by identifying only two percent of all facilities that are at critical positions in the US healthcare facilities network, it is possible to monitor 80 percent of all *Clostridium difficile* infections nationwide. This is a good strategy for optimized surveillance efficiency.

Network metric choices

In this paper I have used the fundamental network characteristics to describe the intrahospital transfers, namely the degree (in and out), edge and vertex counts as well as density. Directions are used when appropriate. These metrics are simple yet captures important features of the system. Similar results can be seen in [4, 34].

Some other common metrics to describe networks such as *clustering coefficients, degree distribution, centrality, shortest path* have not been used in this paper, which have appeared in the studies mentioned above. There are a few reasons behind this. The intrahospital transfers have meaningful clinical context where patients have specific medical needs, and the networks are inherently temporal and directed. This suggests that shortest path metric (i.e. the minimum number of edges that exist between point A to B, which is a good metric to measure how fast information spread) is not as clinically relevant in this setting: if a patient in ward A needs to be treated in ward B before he reaches ward C, he *will have* to visit ward B even though A and C are connected for other patients. Clustering coefficients and centrality are not applicable for another reason: our network is incomplete for the whole hospital system, and our networks are relatively small compared to other studies. We are not trying to identify the small world property as suggested in [34], but evolution and temporal differences over one year period and between weekdays and weekends. Our choices of metrics are suitable for our purposes.

Time stamps for network construction

The EHR data used in this study has consistent hourly time stamps *across* different patients. For the purpose of privacy protection, the index time $t = 0h$ is on an unknown day in June 2018 where only personnel who extracted the data from the data warehouse know which day it is. This feature is non-trivial: it is possible to identify events that are happening at the the same time. In contrast, EHR data that use a privacy protection strategy such as randomly assign distinct index time to each patient (such as MIMIC-III database) will not allow the construction of a temporal network. This is possibly one reason for the sparsity of temporal network analysis on patient transfers.

Transfer chain (motif) analysis

In this paper we analyzed the typical transfer patterns at the patient level (chains). In computer science and network science communities these reappearing

graphlets are called motifs [36, 47, 52]. Recent efforts in network sciences have been developing new fast algorithms to extract motifs in big temporal networks, especially social networks. Although the repeating patterns is intuitive in the patient transfer context, there is few mentioning of motifs in healthcare and medicine. Patient transfer chains are known and understood by physicians who work in the front line with patients, however their quantification and frequency are not well documented. Also, it is feasible in our medium sized dataset, yet how it scales in bigger datasets is unknown. Time lapse between stages are also interesting to investigate. The most interesting point in my view is to apply our findings with current simulation methods (discrete event, multi-state models) to investigate the spread of infections or hospital capacity.

Risk factors for many transfers

Our secondary objective was to identify risk factors for number of transfers. In order to avoid artificially boosting the number of transfers by in-and-out of operation room transfers, we coded operations and perioperative wards into one single location. We used a regression approach where the response is the number of transfers using departments, patient demographics, antibiotics and NEWS scores (first 48 hours average) as covariates. The biggest challenge for this task is the covariate construction: it is very difficult to meaningfully represent the changing patient clinical conditions over the course of the entire visit into one static feature. For example, the NEWS score is designed to capture the physiology derangement for *early* warning. However, NEWS is also recorded over the entire patient stay which should also provide information on the patient condition at each time of measurement. It requires a clear formulation of the research question and interpretation before the construction of the explanatory variables.

Another challenge is an ever lasting topic in statistics: causality. The direction of causal relationship is unclear in this study, or in many observational studies in healthcare. Does having antibiotics (or infection) result in longer length of hospital stays and more numbers of transfers? Or does more transfers lead to being infected? An intuitive but overly simplistic solution is to look at the time of event, whichever happened first 'caused' the other. Nevertheless, most patients seemed to have antibiotics administered until they were discharged so there was no easy way to tell which happened first. Even with such detailed temporal data, it is still very difficult to answer causal questions. These issues need to be addressed with more precise research questions, and more appropriate methodology.

Lack of laboratory data

In the AHUS dataset we used for this paper, we do not have laboratory data for infection status from each patient. This is a limitation of our study as we could not carry out analysis on the important topic of hospital acquired infections. The use of antibiotics (which we had) could serve as a proxy for being infected yet it was not optimal even after we exclude prophylactic antibiotics for

surgical patients. The inclusion of lab data should be implemented in the future studies.

9.4 Paper III

9.4.1 Contributions and implications

The paper *Intervenable predictions of hospital operations using Electronic Health Records* provides a framework for making predictions that balance interpretability and accuracy. This is the basis for making interventions to avoid undesirable outcomes in a healthcare environment and improve quality of care. The appropriate application can be allocating more resources (such as nurses) to wards that need them to prevent adverse events. This paper focuses on the idea of making predictions where stakeholders (hospital management) could identify the risk factor involved and potentially make a change, rather than the exact application. When the availability of data ceases to be an obstacle, variants of our methods could be deployed as appropriate.

From a methodological point of view, this paper has a few novelties. We use a two step approach: the first one is a penalized linear regression analysis that induces sparsity, which attempts to select variables (risk factors) that can predict the outcome. Then in the next step we use the residuals from the previous step as the new response, and the lags of the original response as the explanatory variables. In this way, the model utilizes not only the information in the explanatory variables, but also the unmodel part left in the residual. In our paper we only present simple models (linear regression as the first step), however the idea can be extended to other linear and non-linear models.

9.4.2 Methods, challenges and limitations

Regression as the choice of model

Regression models, including linear and generalized linear regression models (GLM), generalized additive model (GAM), have been applied in various of application fields for decades. They are simple and flexible: the response variables is expressed as a weighted sum of original or transformed explanatory variables, and the response can take the form of real values, binary or counts.

In the Interpretable Machine Learning (IML) field, regression models are considered as inherently interpretable [48]. This is thanks to the weighted-sum formulation of the regression models and the rich literature on the theoretical properties of these models. It is also worth mentioning that when it comes to prediction and forecasting (either numerical outcome or classes), simple, interpretable models such as logistic regression and ARIMA models are still among the best performing models among much more complex methods such as neural networks [22, 25, 31, 41].

The challenge for non-penalized regression models arises when there are many explanatory variables, especially when $n < p$. The interpretability will be difficult and there might not be a unique solution. A standard approach is to use regularization such as ridge or lasso regression.

Sparsity and variable selection

In this paper we achieve sparsity via the lasso and a form of hierarchical structure in the regularization. One of the benefits of limiting the number of variables is to make the model simpler for the human brain to understand, especially when there are many variables. It could help with understanding the data generating mechanism and provide possible directions for the future experimental design. The medical scores used to identify patients at risk are common in clinical practices, and they are derived from clinical variables by assign them with weights, similar to regression coefficients[68]. Here the variable selection is purely based on domain knowledge. There exist a wide range of penalties that are possible for regression models, many of which allow domain knowledge or beliefs to be included in the data-driven variable selection process.

There are also several limitations in variable selection via sparsity inducing norms or regularization. The first limitation is the selection procedure. One way to produce sparse models, the stepwise selection is a very common strategy for low dimensional problem. However adding or dropping one variable at a time might be problematic when interaction exist, and stepwise selection requires computing a large number of competing models. The lasso itself is known to not handle interaction or correlated variables well. This might be mitigated via some structures on the penalty, such as grouped or hierarchical selection.

Another limitation is the inference on parameters after selection. The lasso estimators, in contrast to the least-squared estimators, are biased towards zero. While it has the benefit of shrinking some variables to exactly zero hence producing sparse results, it is difficult to make inference even with resampling techniques such as bootstrap. There are techniques to address this issue, however they are beyond the scope of this discussion [37, 67].

Penalized regression with time series

Although the theoretical works on lasso and other penalties are mostly focused on scenarios where observations are independent, penalized regressions are also frequently applied to time series data. The theoretical properties of the model and sparsity consistency (whether the correct variables with non zero coefficients could be identified) are summarized in [33] and reference therein. In the numerical experiments by [33] using various penalties (lasso, elastic-net and their adapted versions) on autoregressive (AR), autoregressive distributed lag (ADL) and vector-autoregression (VAR) models, the selection is mostly affected by the correlation between explanatory variables. This is consistent with our own experiments, where the variable selection stage is able to identify the correct variables in uncorrelated scenario despite the data are sequentially correlated.

Beyond the penalties and procedures that are widely used for independent observations, there also exist procedures for sequentially correlated data [24, 27, 53, 72, 73].

Forecasting

Forecast is a challenging task. When a simple univariate model such as ARIMA is applied, the forecast is made using the available data of the series itself. When a regression model or autoregressive model with exogenous variables is applied, it is also necessary to know the forecast values of the exogenous or explanatory variables, which increase the uncertainty even further. Only when the exogenous variables are known in advance or we have high confidence in (such as schedules) will the final forecast of the response be more certain. Usually seasonality or trend are very informative, and time series forecast models typically decompose the data into components to enhance the forecast.

Explainable AI

Explainable artificial intelligence (XAI) has recently become an important topic especially in healthcare applications. There have been various methods trying to "open the black box", i.e. understand how the model works so that medical decisions can be made with trust. XAI can be internally explainable methods, or post-hoc or externally explainable methods [20, 42]. In Paper III we use the linear regression model which is the simplest internally explainable method, where the outcome is a weighted sum of risk factors. Externally explainable methods such as the Local Interpretable Model-agnostic Explanations (LIME) [60], Shapely values [19] could also be used to understand which input are meaningful as intervention variables.

Intervention and causality

Healthcare machine learning is mostly done on observational data, and this poses challenge to understanding causality. One example given in [21] is that asthmatic patients admitted for pneumonia are treated more aggressively hence lowering the sub-population mortality, hence a simple model will learn that asthma is protective, which is untrue. Special methodology is required to infer causal relationship using observational data [55].

For time series data, there exist a few approaches to infer intervention effects that do not fall under the causal inference frameworks (Structural Equation Modeling[54] and potential outcome [61]). In marketing applications it is common to infer the effects of interventions, and one approach in practice is the Bayesian Structural Time series [11]. This approach to construct the counterfactuals where an intervention is *not* in place is similar to what we are doing in Paper III, i.e. prediction using inputs without intervention. Alternatively, interrupted time series analysis is a popular method to assess the intervention effect in time series in public health applications [6, 7, 15, 35].

Real life application

Unfortunately, for this paper I did not have a suitable response variable from the AHUS dataset. This is a limitation of the study. We could only demonstrate hypothetical scenarios based on the model we assumed that generated the outcome variable. In the future I hope we could have access to more variables in the EHR datasets and construct realistic examples.

9.5 Data protection

Here I would like to briefly describe my experience on use of EHR and ethics. Privacy protection on health data is important, however the way it is implemented can have various implications to research: from access to data, to the research questions that can be addressed.

MIMIC data

The access to MIMIC data had been a smooth experience as the data is publicly available. A privacy protection course was required in the application for data access, which took a few hours to complete. Once the access is granted, the data can be downloaded to one's personal computer.

The MIMIC III data constitutes of 26 separate data files and are linked by unique patient ID. The time of admission for each patient is randomly shuffled and projected into the future, so there is no possibility to identify which patients were in the hospital at the same time. This protects privacy of the patients while also render certain research topics that require time information impossible (e.g. season, real time of the day, weekend). Yet the independence of patients is suitable for most machine learning problems such as disease progression prediction on individual patients.

AHUS (Norwegian) data

The access to the Norwegian hospital data had been a long and laborious experience. Initially EHR data from Oslo University Hospital (Oslo Universitetssykehus, OUS) was the designated data source for this PhD project back in 2016. However after the data usage application was approved by the Regional Ethical Committee (REK), we still did not receive the promised data due to bureaucratic reasons even until late 2018. This had been a big obstacle to the PhD project. Eventually we applied for access to EHR data at another Norwegian hospital, Akershus University Hospital (AHUS) and the application was approved in late 2019.

The AHUS data is deidentified by the data extraction personnel at AHUS hospital and then stored in the TSD (Tjenester/Service for Sensitive Data, tsd-drift@usit.uio.no). This is a secure server that only granted researchers have access to, and all analyses need to be done here. Only the project owner is allowed to export aggregated results. The time information is processed in a different way compared to MIMIC data. The data extraction period is roughly one year, however we as investigators only know the start of the study period is

one Friday in June 2018, but do not know the exact date. The time recording for each patient is not shuffled as in MIMIC data hence it is possible to compute the number of patients at a certain ward during a specific time period.

9.6 Future perspectives

In addition to the points raised above, there are a few things that can be done in the future. As of now, access to EHR data for research is still challenging even in countries with good EHR systems in place. Efforts need to be made in order to make data open and still secure. More data similar to the MIMIC III database and eICU are desirable so that researchers can evaluate their models in more than one dataset. The generalisability is crucial for any AI system to be useful.

Complex raw data require robust data processing pipelines. For individual researchers this is a non-trivial task. Some efforts to build consistent cleaning pipelines have been made on the MIMIC III database although they are very data-specific. It would be ideal if a universal data pipeline exists and saves researchers time on this task. Nonetheless, cleaning the data from its raw form would let the researcher know the data better; therefore a hybrid solution or routine could be interesting.

There are many directions for future research using EHR beyond making predictions on patient health status, which is still the most popular theme. One direction forward would be understanding a model's decision process, so that both clinicians and patients can understand the decisions made by machines. This is crucial for building trust in AI and patient safety. Beyond healthcare applications, interpretable machine learning systems are also high demand. There is going to be an exponential growth in this field.

References

- [1] Abanda, A., Mori, U., and Lozano, J. A. "A review on distance based time series classification". In: *Data Mining and Knowledge Discovery* vol. 33, no. 2 (2019), pp. 378–412. arXiv: 1806.04509.
- [2] Anantasech, P. and Ratanamahatana, C. A. *Enhanced weighted dynamic time warping for time series classification*. Vol. 797. Springer Singapore, 2019, pp. 655–664.

-
- [3] Bean, D. M., Taylor, P., and Dobson, R. J. “A patient flow simulator for healthcare management education”. In: *BMJ Simulation and Technology Enhanced Learning* vol. 5, no. 1 (2019), pp. 49–51.
- [4] Bean, D. M. et al. “Network analysis of patient flow in two UK acute care hospitals identifies key sub-networks for A&E performance”. In: *PLoS ONE* vol. 12, no. 10 (2017), pp. 1–16.
- [5] Ben-tovim, D. et al. “Simulation Modelling Practice and Theory Hospital Event Simulation Model : Arrivals to Discharge – Design , development and application”. In: *Simulation Modelling Practice and Theory* vol. 68 (2016), pp. 80–94.
- [6] Bernal, J. A. L. et al. “Erratum: Correction: Association between the 2012 Health and Social Care Act and specialist visits and hospitalisations in England: A controlled interrupted time series analysis (PLoS medicine (2017) 14 11 (e1002427))”. In: *PLoS medicine* vol. 15, no. 2 (2018), e1002527.
- [7] Bernal, J. L., Cummins, S., and Gasparrini, A. “Difference in difference, controlled interrupted time series and synthetic controls”. In: *International Journal of Epidemiology* vol. 48, no. 6 (2019), pp. 2062–2063.
- [8] Blay, N. et al. “Intrahospital transfers and the impact on nursing workload”. In: *Journal of Clinical Nursing* vol. 26, no. 23-24 (2017), pp. 4822–4829.
- [9] Boncea, E. E. et al. “Association between intrahospital transfer and hospital-acquired infection in the elderly: A retrospective case-control study in a UK hospital network”. In: *BMJ Quality and Safety* (2021), pp. 1–10.
- [10] Bristol, A. A. et al. “A Systematic Review of Clinical Outcomes Associated with Intrahospital Transitions”. In: *Journal for Healthcare Quality* vol. 42, no. 4 (2020), pp. 175–187.
- [11] Brodersen, K. H. et al. “Inferring causal impact using bayesian structural time-series models”. In: *Annals of Applied Statistics* vol. 9, no. 1 (2015), pp. 247–274.
- [12] Bush, K. et al. “Predicting hospital-onset Clostridium difficile using patient mobility data: A network approach”. In: *Infection Control and Hospital Epidemiology* vol. 40, no. 12 (2019), pp. 1380–1386.
- [13] Curtis, D. E. et al. “Healthcare worker contact networks and the prevention of hospital-acquired infections”. In: *PLoS ONE* vol. 8, no. 12 (2013).
- [14] Cusumano-Towner, M. et al. “A social network of hospital acquired infection built from electronic medical record data”. In: *Journal of the American Medical Informatics Association* vol. 20, no. 3 (2013), pp. 427–434.
- [15] Derde, L. P. et al. “Interventions to reduce colonisation and transmission of antimicrobial-resistant bacteria in intensive care units: An interrupted time series study and cluster randomised trial”. In: *The Lancet Infectious Diseases* vol. 14, no. 1 (2014), pp. 31–39.

- [16] Ebadi, A. et al. “A quest for the structure of intra- and postoperative surgical team networks: does the small-world property evolve over time?” In: *Social Network Analysis and Mining* vol. 9, no. 1 (2019), p. 0.
- [17] Fernández-Gracia, J. et al. “Influence of a patient transfer network of US inpatient facilities on the incidence of nosocomial infections”. In: *Scientific Reports* vol. 7, no. 1 (2017), pp. 1–9.
- [18] Folgado, D. et al. “Time Alignment Measurement for Time Series”. In: *Pattern Recognition* vol. 81 (2018), pp. 268–279.
- [19] Fryer, D., Strumke, I., and Nguyen, H. “Shapley values for feature selection: The good, the bad, and the axioms”. In: *IEEE Access* vol. 3 (2021), pp. 3–10. arXiv: [2102.10936](https://arxiv.org/abs/2102.10936).
- [20] Ghassemi, M., Oakden-Rayner, L., and Beam, A. L. “The false hope of current approaches to explainable artificial intelligence in health care”. In: *The Lancet Digital Health* vol. 3, no. 11 (2021), e745–e750.
- [21] Ghassemi, M. et al. “Opportunities in Machine Learning for Healthcare”. 2018. arXiv: [1806.00388](https://arxiv.org/abs/1806.00388).
- [22] Green, K. C. and Armstrong, J. S. “Simple versus complex forecasting: The evidence”. In: *Journal of Business Research* (2015).
- [23] Guo, C., Lu, M., and Chen, J. “An evaluation of time series summary statistics as features for clinical prediction tasks”. In: *BMC Medical Informatics and Decision Making* vol. 20, no. 1 (2020), pp. 1–20.
- [24] Hansheng, W., Guodong, L., and Tsai, C. L. “Regression coefficient and autoregressive order shrinkage and selection via the lasso”. In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* (2007).
- [25] Harutyunyan, H. et al. “Multitask Learning and Benchmarking with Clinical Time Series Data”. In: (2018). arXiv: [1703.07771](https://arxiv.org/abs/1703.07771).
- [26] Hasan, I. and Bahalkeh, E. “Evaluating intensive care unit admission and discharge policies using a discrete event simulation model”. In: vol. 96, no. 6 (2020), pp. 501–518.
- [27] Haselimahhadi, H. and Vinciotti, V. “Penalised inference for lagged dependent regression in the presence of autocorrelated residuals”. In: *Metron* (2018).
- [28] Ho, J., Ghosh, J., and Sun, J. “Marble : High-throughput Phenotyping from Electronic Health Records via Sparse Nonnegative Tensor Factorization”. In: *Kdd* (2014), pp. 115–124.
- [29] Ho, J. et al. “Limestone: High-throughput candidate phenotype generation via tensor factorization”. In: *Journal of Biomedical Informatics* vol. 52 (2014), pp. 199–211.
- [30] Jeong, Y.-s., Jeong, M. K., and Omitaomu, O. A. “Weighted dynamic time warping for time series classification”. In: *Pattern Recognition* vol. 44, no. 9 (2011), pp. 2231–2240.

-
- [31] Johnson, A. E. W., Pollard, T. J., and Mark, R. G. “Reproducibility in critical care : a mortality prediction case study”. In: *2nd Machine Learning for Healthcare Conference, PMLR*. Vol. 68. 2017.
- [32] Johnson, A. E. et al. “MIMIC-III, a freely accessible critical care database”. In: *Scientific Data* vol. 3 (2016), p. 160035.
- [33] Kock, A. B., Medeiros, M., and Vasconcelos, G. *Penalized Time Series Regression*. Vol. 52. 2020, pp. 193–228.
- [34] Kohler, K. and Ercole, A. “Can network science reveal structure in a complex healthcare system? A network analysis using data from emergency surgical services”. In: *BMJ Open* vol. 10, no. 2 (2020), pp. 1–8.
- [35] Kontopantelis, E. et al. “Regression based quasi-experimental approach when randomisation is not an option: Interrupted time series analysis”. In: *BMJ (Online)* (2015).
- [36] Kovanen, L. et al. “Temporal Motifs”. In: (2013), pp. 135–159. arXiv: [1108.1780](https://arxiv.org/abs/1108.1780).
- [37] Lee, J. D. et al. “Exact post-selection inference, with application to the lasso”. In: *Annals of Statistics* vol. 44, no. 3 (2016), pp. 907–927. arXiv: [1311.6238](https://arxiv.org/abs/1311.6238).
- [38] Lines, J. and Bagnall, A. “Time series classification with ensembles of elastic distance measures”. In: *Data Mining and Knowledge Discovery* vol. 29, no. 3 (2015), pp. 565–592.
- [39] Lines, J., Taylor, S., and Bagnall, A. “HIVE-COTE : The Hierarchical Vote Collective of Transformation-based Ensembles for Time Series Classification”. In: *2016 IEEE 16th International Conference on Data Mining (ICDM)* (2016), pp. 1041–1046.
- [40] Liu, A. et al. “Multistate model of the patient flow process in the pediatric emergency department”. In: *PLoS ONE* vol. 14, no. 7 (2019), pp. 1–16.
- [41] Makridakis, S., Spiliotis, E., and Assimakopoulos, V. “Statistical and Machine Learning forecasting methods: Concerns and ways forward”. In: *PLoS ONE* vol. 13, no. 3 (2018), pp. 1–26.
- [42] Markus, A. F., Kors, J. A., and Rijnbeek, P. R. “The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies”. In: *Journal of Biomedical Informatics* vol. 113, no. July 2020 (2021), p. 103655. arXiv: [2007.15911](https://arxiv.org/abs/2007.15911).
- [43] Martinez, D. A. et al. “An Electronic Dashboard to Monitor Patient Flow at the Johns Hopkins Hospital: Communication of Key Performance Indicators Using the Donabedian Model”. In: *Journal of Medical Systems* vol. 42, no. 8 (2018).
- [44] McHaney-Lindstrom, M. et al. “Network analysis of intra-hospital transfers and hospital onset clostridium difficile infection”. In: *Health Information and Libraries Journal* vol. 37, no. 1 (2020), pp. 26–34.

- [45] Mckinley, K. W. et al. “Discrete event simulation modelling to evaluate the impact of a quality improvement initiative on patient flow in a paediatric emergency department”. In: *Emerg Med J* (2020), pp. 193–199.
- [46] Mei, J. et al. “Learning a Mahalanobis Distance-Based Dynamic Time Warping Measure for Multivariate Time Series Classification”. In: *IEEE Transactions on Cybernetics* vol. 46, no. 6 (2016), pp. 1363–1374.
- [47] Milo, R. et al. “Networks Network Motifs : Simple Building Blocks of Complex Networks”. In: vol. 824, no. 2002 (2008).
- [48] Molnar, C., Casalicchio, G., and Bischl, B. “Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges”. In: *Communications in Computer and Information Science* vol. 1323, no. 01 (2020), pp. 417–431. arXiv: 2010.09337.
- [49] Moor, M. et al. “Early Recognition of Sepsis with Gaussian Process Temporal Convolutional Networks and Dynamic Time Warping”. In: (2019). arXiv: 1902.01659.
- [50] Munich, M. E. and Perona, P. “Continuous Dynamic Time Warping for translation- invariant curve alignment with applications to signature verification”. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision* (1999).
- [51] Ong, M. S. and Coiera, E. “Safety through redundancy: A case study of in-hospital patient transfers”. In: *Quality and Safety in Health Care* vol. 19, no. 5 (2010), pp. 1–7.
- [52] Paranjape, A., Benson, A. R., and Leskovec, J. “Motifs in temporal networks”. In: *WSDM 2017 - Proceedings of the 10th ACM International Conference on Web Search and Data Mining* (2017), pp. 601–610. arXiv: 1612.09259.
- [53] Park, H. and Sakaori, F. “Lag weighted lasso for time series model”. In: *Computational Statistics* vol. 28, no. 2 (2013), pp. 493–504.
- [54] Pearl, J. “Causal inference in statistics: An overview”. In: *Statistics Surveys* vol. 3 (2009), pp. 96–146.
- [55] Pearl, J. “Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution”. In: (2018), pp. 1–8. arXiv: 1801.04016.
- [56] Perros, I. et al. “SPARTan : Scalable PARAFAC2 for Large & Sparse Data”. In: *KDD*. 2017.
- [57] Piotrowska, M. J. et al. “Impact of inter-hospital transfers on the prevalence of resistant pathogens in a hospital–community system”. In: *Epidemics* vol. 33, no. January (2020), p. 100408.
- [58] Purushotham, S. et al. “Benchmarking deep learning models on large healthcare datasets”. In: *Journal of Biomedical Informatics* vol. 83 (2018), pp. 112–134.
- [59] Qin, S. et al. “Hospital occupancy and discharge strategies : a simulation-based study”. In: *Internal Medicine Journal* vol. 47 (2017), pp. 894–899.

-
- [60] Ribeiro, M. T., Singh, S., and Guestrin, C. “‘Why should I Trust You?’ Explaining the Predictions of Any Classifier”. In: (2016). arXiv: **1602.04938**.
- [61] Rubin, D. B. “Causal inference using potential outcomes: Design, modeling, decisions”. In: *Journal of the American Statistical Association* vol. 100, no. 469 (2005), pp. 322–331.
- [62] Shenoy, E. S. et al. “A Discrete Event Simulation Model of Patient Flow in a General Hospital Incorporating Infection Control Policy for Methicillin-Resistant Staphylococcus Aureus (MRSA) and Vancomycin-Resistant Enterococcus (VRE)”. In: vol. 38, no. 2 (2018), pp. 246–261.
- [63] Shokoohi-Yekta, M. et al. “Generalizing DTW to the multi-dimensional case requires an adaptive approach”. In: *Data Mining and Knowledge Discovery* vol. 31, no. 1 (2017), pp. 1–31.
- [64] Somlo, D. R. M., Repenning, N. P., and Mangi, A. A. “In-Hospital Delays Result in Worse Patient Outcomes and Higher Cost After Cardiac Surgery”. In: *The Annals of Thoracic Surgery* vol. 106, no. 4 (2018), pp. 1143–1149.
- [65] Tahir, H. et al. “Relevance of intra-hospital patient movements for the spread of healthcare-associated infections within hospitals - a mathematical modeling study”. In: *PLOS Computational Biology* vol. 17, no. 2 (2021), e1008600.
- [66] Tan, C. W., Petitjean, F., and Webb, G. I. “FastEE : Fast Ensembles of Elastic Distances for time series classification”. In: *Data Mining and Knowledge Discovery* (2019).
- [67] Taylor, J. and Tibshirani, R. “Post-selection inference for L1-penalized likelihood models”. In: *Canadian Journal of Statistics* ().
- [68] Ustun, B. and Rudin, C. “Supersparse linear integer models for optimized medical scoring systems”. In: *Machine Learning* vol. 102, no. 3 (2016), pp. 349–391. arXiv: **1502.04269**.
- [69] Vilches, T. N. et al. “The role of intra and inter-hospital patient transfer in the dissemination of healthcare-associated multidrug-resistant pathogens”. In: *Epidemics* vol. 26, no. June 2018 (2019), pp. 104–115.
- [70] Whitt, W. and Zhang, X. “A data-driven model of an emergency department”. In: *Operations Research for Health Care* vol. 12 (2017), pp. 1–15.
- [71] Xing, Z., Pei, J., and Keogh, E. “A brief survey on sequence classification”. In: *ACM SIGKDD Explorations Newsletter* vol. 12, no. 1 (2010), pp. 40–48.
- [72] Xue, Y. and Taniguchi, M. “Modified LASSO estimators for time series regression models with dependent disturbances”. In: *Statistical Methods & Applications* vol. 0123456789 (2020).

9. Discussion

- [73] Yoon, Y. J., Lee, S., and Lee, T. “Adaptive LASSO for linear regression models with ARMA-GARCH errors”. In: *Communications in Statistics: Simulation and Computation* (2017). arXiv: [03610918.2015.1096372](https://arxiv.org/abs/1510.09637) [[10.1080](https://doi.org/10.1080)].
- [74] Yousefi, M., Yousefi, M., and Fogliatto, F. S. “Simulation-based optimization methods applied in hospital emergency departments : A systematic review”. In: *Simulation* vol. 96, no. 10 (2020), pp. 791–806.

Chapter 10

Conclusion

In this thesis I introduced novel methods to extract, integrate and represent information from heterogeneous temporal hospital electronic health record EHR data. Three distinctive papers covering different clinical applications (clinical event classification, patient flow management, logistic optimization) are included in the thesis, making use of a wide range of statistical, machine learning and graph theory tools. The thesis illustrates the opportunities and challenges in healthcare in the Big Data Era, and at same time demonstrates the methodological solutions to tackle these challenges. I believe the field of medical informatics using EHR data is only getting more exciting where practitioners and teams with interdisciplinary expertise can and will excel. However, it is the patients and everyone using the healthcare services who will truly benefit from the advancement in this field.

Papers

Paper I

Feature extraction from unequal length heterogeneous EHR time series via dynamic time warping and tensor decomposition

Chi Zhang, Hadi Fanaee-T, Magne Thoresen

Published in *Data Mining and Knowledge Discovery* (2021) **35**: 1760-1784. DOI: 10.1007/s10618-020-00724-6.



Feature extraction from unequal length heterogeneous EHR time series via dynamic time warping and tensor decomposition

Chi Zhang¹ · Hadi Fanaee-T² · Magne Thoresen¹

Received: 6 January 2020 / Accepted: 9 November 2020 / Published online: 4 January 2021
© The Author(s), under exclusive licence to Springer Science+Business Media LLC, part of Springer Nature 2021

Abstract

Electronic Health Records (EHR) data is routinely generated patient data that can provide useful information for analytical tasks such as disease detection and clinical event prediction. However, temporal EHR data such as physiological vital signs and lab test results are particularly challenging. Temporal EHR features typically have different sampling frequencies; such examples include heart rate (measured almost continuously) and blood test results (a few times during a patient’s entire stay). Different patients also have different length of stays. Existing approaches for temporal EHR sequence extraction either ignore the temporal pattern within features, or use a predefined window to select a section of the sequences without taking into account all the information. We propose a novel approach to tackle the issue of irregularly sampled, unequal length EHR time series using dynamic time warping and tensor decomposition. We use DTW to learn the pairwise distances for each temporal feature among the patient cohort and stack the distance matrices into a tensor. We then decompose the tensor to learn the latent structure, which is consequently used for patient representation. Finally, we use the patient representation for in-hospital mortality prediction. We illustrate our method on two cohorts from the MIMIC-III database: the sepsis and the acute kidney failure cohorts. We show that our method produces outstanding classification performance in terms of AUROC, AUPRC and accuracy compared with the baseline methods: LSTM and DTW-KNN. In the end we provide a detailed analysis on the feature importance for the interpretability of our method.

Keywords Electronic health records · Dynamic time warping · Tensor decomposition · Patient similarity

Responsible editor: Panagiotis Papapetrou.

Extended author information available on the last page of the article

1 Introduction

Electronic health record, or EHR data is patient data routinely generated from health institutions, including demographics, diagnoses, vital measurements, clinical notes, laboratory test results and medical images. EHR data can provide valuable information for analytical tasks including but not limited to disease detection and classification, medical concept embedding and data augmentation (Xiao et al. 2018). However, EHR data can be challenging due to multi-modality of data types, lack of outcome labels and missingness, temporality and irregularity (Ghassemi et al. 2018; Kruse et al. 2016). Temporal EHR sequences are difficult to model due to two sources of variability in length of sequence: feature-wise and subject-wise. Different features (also known as variables or parameters) can vary greatly in terms of measurement frequency, from measured nearly continuously (blood pressure) to daily or whenever necessary (laboratory test). Different patients (or subjects) can have varying periods of stays in the hospital or intensive care units. It is crucial to represent the data in meaningful ways to proceed with further analytical tasks such as clinical event predictions, hence the heterogeneity in sequence length poses challenges.

There have been several approaches to represent temporal EHR data. One simple approach is to compute sample statistics (minimum, maximum, mean, standard deviation, number of measurements, first measurement) for features at predefined intervals, for instance the first 48 h of the hospital stay (Harutyunyan et al. 2018; Johnson et al. 2017; Guo et al. 2020a). Classification tasks are then completed using classifiers such as logistic regression or gradient boosting machines. This approach produces human readable and interpretable features and can adapt to both feature and subject sequence length variability, however loses the temporal dependency which is valuable for modeling pathophysiologic evolution and disease progression (Luo et al. 2016; Alaa and van der Schaar 2018). Recent developments in deep learning, especially Recurrent Neural Networks (RNN) are able to capture the temporal pattern in multiple features. Long Short Term Memory (LSTM) networks are a type of state-of-the-art RNN, and have shown many successful applications in temporal healthcare data representation and classification. Such examples include sepsis prediction (Scherpf et al. 2019), unplanned intensive care unit readmission (Lin et al. 2019), mortality risk monitoring (Kaji et al. 2019; Purushotham et al. 2018) and other clinical event detection and diagnosis (Lipton et al. 2016). Bidirectional LSTM (BiLSTM) is a variation of LSTM which takes both forward and backward sequence dependency into account, and has been successful in disease inference and predictions (Yu et al. 2020; Guo et al. 2020b).

LSTMs are typically trained on a specified window (first 24 or 48 h of patient records) therefore ignore the irregular sequence lengths (Suresh et al. 2018; Purushotham et al. 2018; Song et al. 2018; Lei et al. 2018). When there are missing values or variables, it is necessary to impute: either with mean, zero or with more complex approaches such as Gaussian Processes (Lipton et al. 2016; Moor et al. 2019). Despite of their outstanding performance in classification tasks, most deep learning methods require a large amount of data to train and remain complex with tens of thousands of hyperparameters that are hard to interpret (Lipton 2016). Recent works to improve interpretability in deep learning using 'attention' mechanisms still require complex architecture (Alaa and van der Schaar 2018; Song et al. 2018). We need therefore

a more transparent tool that can account for temporal sequences sampled at various frequencies for irregular length of periods from different patients.

Modeling unequal length temporal EHR features directly in the raw form requires either extracting the summary statistics at a snapshot or segmenting the sequences into a regular window, as outlined above. On the other hand, if we model the relations between sequences such as similarity instead of the raw sequence itself, the segmentation could be avoided. It is natural to study the similarity or distance (we use these two terms interchangeably) as patients with similar conditions might display similar patterns of physiological trends (Luo et al. 2016). This forms clusters of patients that can be used for personalized predictions and treatments (Che et al. 2017; Ruffini et al. 2017) and to help understand the underlying patient characteristics, also known as phenotyping (Ho et al. 2014a, b; Perros et al. 2017). Powerful data mining tools such as dynamic time warping (Keogh and Pazzani 1999) can align and compare two time series of unequal lengths, and has proven effective in EHR temporal sequence learning (Che et al. 2017; Moor et al. 2019). The distances computed for different features then need to be integrated in some way for further classification tasks. Luo et al. (2016) used frequent subgraph mining to group patients with similar temporal trends, then used subgraph groups to predict 30-day mortality. Moor et al. (2019) proposed to use a hybrid of dynamic time warping (or DTW in short) and the K-nearest neighbor ensemble algorithm to classify each feature, then ensemble the predictive score together to predict sepsis onset. Outside healthcare related applications, nearest neighbor type classifiers with some distance metric remains one of the most powerful time series classification methods (Tan et al. 2019; Bagnall et al. 2016).

Instead of classifying each feature individually and then integrate, an alternative to collect all features together is to put the DTW distance matrices into a multidimensional array: a tensor. In this way data from more than two dimensions can be captured conveniently. This tensor contains information about all features that were originally irregular at feature level and subject level, and its decomposition can provide useful insights on the characteristics of the features and the cohort. We therefore propose a novel method to represent irregular length temporal EHR data via dynamic time warping and tensor decomposition. Instead of using a fixed window of data, we use the full patient sequences from various features that typically differ for each patient. We learn the patient-pairwise feature distance for each feature using dynamic time warping. Based on these distance matrices we construct a third order tensor, then decompose the tensor using CANDECOMP/PARAFAC decomposition (Kiers 2000). Our approach is referred to as DTW-CP. The learned latent feature matrix contains information that can further produce patient representation for supervised learning tasks.

We test DTW-CP on two different cohorts from the open MIMIC-III critical care database with an in-hospital mortality prediction experiment, and compare with baseline results produced by LSTM. With sufficient number of latent components, DTW-CP has consistently better classification performance on both cohorts in three metrics. We provide a detailed analysis of the features and learned latent components to provide insight on which features contain more information for the classification performance.

The rest of the paper is organized as follow. Section 2 provides background information for dynamic time warping and CP decomposition, and describes our methodology of patient representation in detail. Section 3 outlines the experimental evaluation and implementation details. Section 4 provides results for the experiments and analysis of feature importance. Section 5 discusses the strength, limitation and future works and conclude the paper.

2 Methodology

We give a brief review of dynamic time warping and tensor decomposition in Sect. 2.1, then describe our method for patient time series representation in Sect. 2.2.

2.1 Background

We first introduce the notations used in the paper (consistent with Kolda and Bader 2009). A tensor is a multidimensional array, the number of dimensions is called order, modes or ways. In this work we focus on third order tensors. A slice is a two dimensional section of a tensor with two fixed modes. For example $X_{1::}$ is a horizontal slice, which is the first layer or top matrix of a tensor (Table 1).

2.1.1 Tensor decomposition

Tensor decomposition has wide applications in signal processing and data mining (Sidiropoulos et al. 2017; Acar et al. 2017), and has been applied successfully in helathcare informatics (Ho et al. 2014a; Henderson et al. 2017, 2018). In this paper we focus on CANDECOMP/PARAFAC or CP decomposition for short. For a third order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, a CP decomposition for a chosen number of components

Table 1 List of notations

Symbol	Definition
X, D, M	Matrix
X^T	Matrix transpose
x_r	r-th column of X
\mathcal{X}, \mathcal{D}	Tensor
x_{ij}, x_{ijk}	Elements of a matrix and a tensor
$X_{i::}, X_{:j}$	Horizontal, lateral slice of tensor
$X_{::k}$ or simply X_k	Frontal slice of tensor
x, y	Vector
\circ	Outer product

$r = 1, \dots, R$ can be formalized in the following way:

$$\min_{\hat{\mathcal{X}}} \|\mathcal{X} - \hat{\mathcal{X}}\| \quad \text{where} \quad \hat{\mathcal{X}} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \tag{1}$$

Here $\mathbf{a}_r, \mathbf{b}_r, \mathbf{c}_r$ are column vectors of size I, J, K . The vectors can be reorganised into factor matrices $[[\mathbf{A}, \mathbf{B}, \mathbf{C}]]$ where $\mathbf{A} \in \mathbb{R}^{I \times R}, \mathbf{B} \in \mathbb{R}^{J \times R}, \mathbf{C} \in \mathbb{R}^{K \times R}, \mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_R]$. If the columns of $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are normalized to unit length, then the weights are absorbed into $\boldsymbol{\lambda} \in \mathbb{R}^R$,

$$\hat{\mathcal{X}} = \sum_{r=1}^R \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r. \tag{2}$$

More details on tensors and the CP decomposition can be seen in (Rabanser et al. 2017; Kolda and Bader 2009) and references therein.

2.1.2 Dynamic time warping

Dynamic time warping (DTW) is a technique to find the optimal alignment between two time dependent sequences, specifically with time deformation and different speed (Keogh and Pazzani 1999; Muller 2007). Given two time series $\mathbf{x} = (x_1, x_2, \dots, x_N)$ and $\mathbf{y} = (y_1, y_2, \dots, y_M)$, construct a cost matrix $\mathbf{C} \in \mathbb{R}^{N \times M}$ with elements $c_{n,m} = d(x_n, y_m)$. Here d is a distance measure. With squared Euclidean distance, $d(x_n, y_m) = (x_n - y_m)^2$.

A warping path $W = (w_1, \dots, w_Q)$ is a set of matrix indices that defines a mapping between \mathbf{x} and \mathbf{y} where Q is the length of the warping path. Let $w_1 = (1, 1), w_Q = (N, M)$, indicating that the warping path starts and ends in the opposite corner cells of the matrix (boundary conditions). W also need to satisfy additional continuity and monotonicity conditions (Keogh and Pazzani 1999). Let the total cost of a warping path W between \mathbf{x}, \mathbf{y} be

$$TC_W(\mathbf{x}, \mathbf{y}) = \sum_{q=1}^Q c_{w_q}, \tag{3}$$

The optimal warping path W^* is the one that minimizes the total cost among all possible paths, and the DTW distance is the total cost associated with W^* ,

$$\begin{aligned} DTW(\mathbf{x}, \mathbf{y}) &= TC_{W^*}(\mathbf{x}, \mathbf{y}) \\ &= \min\{TC_W(\mathbf{x}, \mathbf{y})\}. \end{aligned}$$

It is time consuming to find the optimal warping path. By restricting the difference between possible alignment indices between time series pairs, the search window is narrowed around the diagonal of the warping cost matrix. Two well known global

constraints are the Sakoe–Chiba band (Sakoe and Chiba 1978) and Itakura parallelogram (Itakura 1975). A comparison between these two constraints has been made by Geler et al. (2019). More recent works have investigated learning constraints from the data for faster computation and better accuracy (Ratanamahatana and Keogh 2004; Niennattrakul and Ratanamahatana 2009; Salvador and Chan 2007; Dau et al. 2017). It is worth mentioning that constraints work well when the time series lengths do not differ much, otherwise the warping path might not exist (Giorgino 2009).

2.2 Representation of EHR time series

In this section we describe the workflow of representing patient time series of unequal length and sampling frequency. Each unique variable of such physiological time series such as temperature or white blood cell count is referred to as a *feature*. We use the term *distance* and *similarity* interchangeably. Denote the patient index $i, i = 1, \dots, N$ and feature index $k, k = 1, \dots, K$. The length of stay for different patients varies, leading to patient-specific time index denoted by $t_i = (t_{i1}, \dots, t_{iT})$. The temporal sequence of feature k associated to patient i is recorded as

$$p_{ik} = (p_{ik,t_{i1}}, \dots, p_{ik,t_{iT}}). \tag{4}$$

2.2.1 Learning latent feature structure

Due to the irregularity in lengths of feature sequences across patients and features, we transform the problem from modeling the individual feature itself for all patients to modeling the similarity of feature between pairs of patients. As dynamic time warping (DTW) can align and compute the distances between pairs of univariate sequences with varying lengths, for each feature k , we compute the distance between each pair of patients (i, j) denoted by

$$d_{ijk} = DTW(p_{ik}, p_{jk}). \tag{5}$$

The procedure is illustrated in Fig. 1a. This forms a third order pairwise distance tensor $\mathcal{D} \in \mathbb{R}^{N \times N \times K}$ where the three modes correspond to patient, patient, feature respectively (Fig. 1b). Each frontal slice $\mathbf{D}_{::k}$ represents the pairwise distance matrix for feature k . Elements in the same slice $\mathbf{D}_{::k}$ have 0 as diagonal elements, $d_{iik} = 0$ for $i = 1, \dots, N$.

We then proceed by decomposing the tensor \mathcal{D} via CP decomposition with chosen number of components R (Fig. 1c). The motivation for this step is twofold. On the one hand, using CP allows us to learn the latent variables from a complex set of data of multiple unequal-length time series across different patients; on the other hand we reduce the dimensionality of the data and make it possible to carry out predictive tasks. CP produces three factor matrices $M_1 \in \mathbb{R}^{N \times R}, M_2 \in \mathbb{R}^{N \times R}, M_3 \in \mathbb{R}^{K \times R}$ that are the combinations of the vectors from the rank-one components. They represent patient, patient and feature modes. We refer to M_3 as the *feature factor* matrix where each element $M_{k,r}$ is the loading or weight for feature k on component r .

2.2.2 Patient representation for prediction

We further examine the distance tensor \mathcal{D} from another perspective: its lateral slices $\mathbf{D}_{:i}$. Our approach is similarity based, hence it is necessary to have a common key or *pivot* patient to compare with. A pivot patient is defined as the patient I in the cohort whose features of other patients $i = 1, \dots, N, i \neq I$ are compared to. For instance, the first slice on the left $\mathbf{D}_{:1} \in \mathbb{R}^{N \times K}$ contains DTW distances for all features comparing patient $I = 1$ with all other patients (Fig. 1d). Such a matrix is referred to as a *pivot distance matrix*. Each pivot distance matrix is partial as it only contains distances compared with one key patient. In order to complete a predictive task such as mortality classification, directly using the distance matrix as input creates problems because there is no rule as for which lateral slice (i.e. which pivot patient) to choose. Each component of the feature factor matrix \mathbf{M}_3 , however, contains feature information (loadings) collected from all patient pairs that can be used for prediction. We produce patient representation $P_I \in \mathbb{R}^{N \times R}$ by projecting the pivot distance matrix onto the feature factor matrix (Fanaee-T et al. 2013) as shown in Fig. 1e,

$$P_I = \mathbf{D}_{:I} \mathbf{M}_3. \tag{6}$$

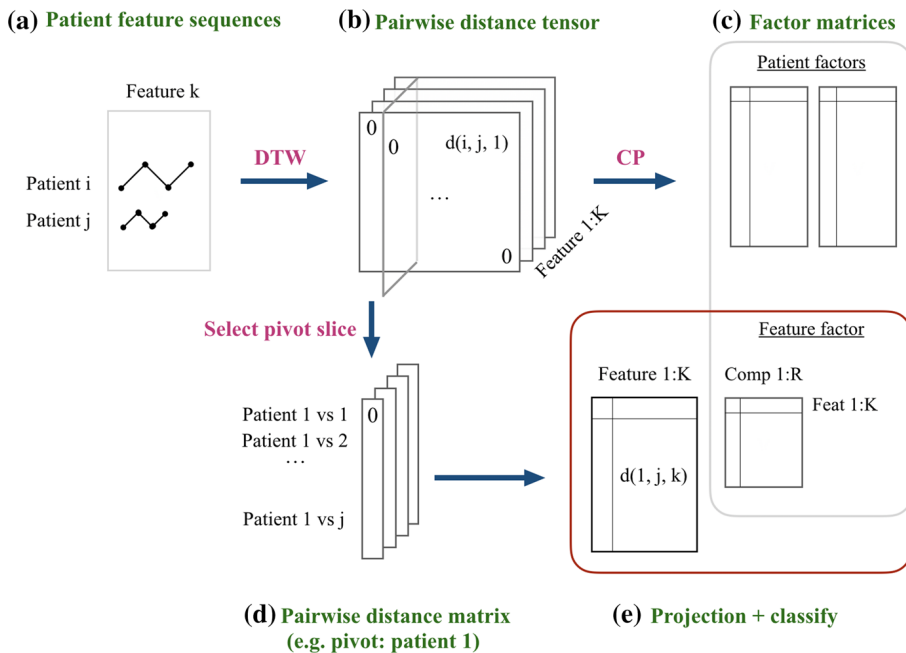


Fig. 1 a, b, c: Procedure of learning latent feature factor, where $d(i, j, 1)$ is the DTW distance between patients (i, j) for feature 1. d, e: Learning patient representation for prediction. Similarly, $d(1, j, k)$ is the DTW distance between patients $(1, j)$ for feature k

2.2.3 Training and testing procedure

Now we describe the workflow of the training and testing. First the data set is randomly split into training and test sets with 70/30 proportion and class stratification. In the training set, DTW distances are computed for all features. The distance tensor is constructed and then decomposed, producing the feature factor matrix for a pre-chosen number of latent components. To create the projection matrix for classification, we choose an arbitrary lateral slice (of pivot patient I) from the training tensor, and project it onto the feature factor matrix. In the illustration of Fig. 1, $I = 1$ but it can be different. This projection is used for training the classifier. For the test set, it is necessary to compute the DTW distance for feature sequences between the test subjects and the pivot patient I , then make the projection onto the feature factor matrix. This is because we need to make the distance representation consistent: both training and test distances for projection need to be compared with the same pivot.

3 Experimental evaluation

We carry out experiments using a publicly available database, the Medical Information Mart for Intensive Care (MIMIC III) database (Johnson et al. 2016). This is a single center database that contains information about patients admitted to critical care units at Beth Israel Deaconess Medical Center, Boston, USA. The data types include, but are not limited to structured data such as temporal physiological signs and laboratory test results, static demographic information such as age and gender, as well as unstructured data such as free text clinical notes. In the current work we will focus on the structured temporal data. Recent works on reproducible studies using MIMIC-III data make it possible to extract consistent patient cohorts and features. We select two cohorts for our experiments, and our selection criteria is in line with (Johnson et al. 2017).

3.1 Cohort and feature selection

3.1.1 Sepsis cohort

The first cohort we examine is a subset of the sepsis cohort originally studied in (Ribas Ripoll et al. 2014) then reproduced by Johnson et al. (2017). We choose patients who have a sepsis diagnosis (ICD-9 code 995.92 or 785.52) and Simplified Acute Physiology Scores (SAPS) (Le Gall et al. 1993). We only keep patients who have been in the ICU for no more than seven days (168 h), making a cohort of 1425 ICU stays in total. Of these patients, 38.9% are associated with a mortality outcome. Our study period is longer than other works using DTW or similarity-based methods that used only 12 or up to 48 h (Luo et al. 2016; Moor et al. 2019). It is of interest to see whether DTW still works well for longer sequences. We design an incremental inclusion criterion: group 1 contains all subjects with below 24 h (1 day) records, group 2 contains subjects with below 48 h (2 days) records and so on, until 7 days. This suggests that patients within groups with shorter stays are also included in those

Table 2 Information for the sepsis and the acute kidney injury (AKI) cohorts

Cohort index	Length of stay (h)	N patient (Case, Control)
Sepsis 1	[5, 24]	225 (136, 89)
Sepsis 2	[5, 48]	546 (240, 306)
Sepsis 3	[5, 72]	833 (338, 495)
Sepsis 4	[5, 96]	1048 (410, 638)
Sepsis 5	[5, 120]	1202 (468, 734)
Sepsis 5	[5, 144]	1329 (521, 808)
Sepsis 5	[5, 168]	1425 (554, 871)
AKI 1	[5, 24]	652 (189, 463)
AKI 2	[5, 48]	1676 (370, 1306)
AKI 3	[5, 72]	2448 (515, 1933)
AKI 4	[5, 96]	2959 (611, 2348)
AKI 5	[5, 120]	3284 (683, 2601)
AKI 6	[5, 144]	3521 (757, 2764)
AKI 7	[5, 168]	3705 (801, 2904)

Note that for the AKI experiment we use 50 fixed size of stratified random samples (500 subjects in total with 150 cases, 350 controls) for each subgroup 1–7

with longer stays. In this way we can observe DTW's performance on data with smaller and larger sequence length variability.

3.1.2 Acute kidney injury cohort

The second cohort is the acute kidney injury (referred to as AKI in the rest of the paper) cohort based on Johnson et al. (2017). We select patients who have ICD-9 diagnosis of acute kidney injury (code 584.9) who have no more than seven days stay, similar to the previous section. We end up with a cohort of 3705 patients (21.6% mortality). Similar to the previous cohort, we segment the cohort into seven incremental groups: below 24, 48, 72, 96, 120, 144, 168 h corresponding to 1 to 7 days. We modify the experiment slightly to assess the stability of our method in a more controlled scenario. We fix two aspects of the cohorts: sample size and class distribution. We perform experiments on 50 *random samples* of fixed size 500 subjects from the five subgroups corresponding to length of stay, shown in Table 2. The class distribution within each sample is set to 30% case (dead) and 70% control (alive). This produced 350 random samples in total.

3.1.3 Feature selection

In the temporal EHR prediction literature there have been some frequently used physiological and laboratory test variables (Johnson et al. 2017; Moor et al. 2019; Luo et al. 2016; Suresh et al. 2018). The majority of these features are the same, such as heart rate, oxygen saturation and others. Nonetheless, there are some study-specific

Table 3 Extracted features and abbreviations for our experiment

Heart rate	HR	Mean blood pressure	MBP
Systolic blood pressure	SBP	Diastolic blood pressure	DBP
Respiratory rate	RR	Temperature	Temp
Oxygen saturation	SpO2	Glasgow coma scale total	GCS
GCS motor, verbal, eyes	GCS_m, v, e	Urine output	UO
Endotracheal flag	EndoFlag		
Anion gap	AG	Albumin	ALB
Immature Band forms	Band	Base excess	BE_bg
Bilirubin	BIL	Blood Urea Nitrogen	BUN
Bicarbonate	HCO3*	Carboxyhemoglobin	CoHB_bg
Calcium	Ca_bg	Chloride	CL*
Creatinine	CR	Glucose	Glu*
Glucose chart	Glu_c	Hematocrit	HCT*
Hemoglobin	HGB*	Lactate	LAC*
Methemoglobin	MetHb_bg	International Normalized Ratio	INR
Partial pressure (Oxygen)	PO2_bg	Partial pressure (CO2)	PCO2_bg
pH	pH	Platelets	PLT
Prothrombin time	PT	Partial thromboplastin time	PTT
Potassium	K*	Sodium	Na*
Total CO2 concentration	totalCO2_bg	White blood cell count	WBC

The top panel consists of vital signs as well as urine output, Glasgow coma scales and endotracheal flag. The bottom panel contains laboratory test variables. *bg*: arterial blood gas measurement

* Indicate that this feature has more than one measurement source, the other being blood gas

features included in each paper, for instance, Luo et al. (2016) uses volumes of gas exchanged per minute which is not included in other studies. For consistency, we extract a reproducible set of features from Johnson et al. (2017), listed in Table 3. Repeated feature names such as glucose is due to multiple sources of data produced in different test procedures, as explained by the authors (finger-stick glucose or arterial blood gas glucose). The final number of features is 52.

3.2 Implementation details

For features other than lab test variables, we use the period starting from patient admission into ICU until their discharge (from '0h' to end of stay illustrated in Fig. 2). For lab test features, we include an extended period of 24 h before admission (from '-24h' to the end of stay). These features are typically measured less frequently, and an additional period may contain useful information (Johnson et al. 2017). Each feature is standardized by subtracting the mean and dividing by the standard deviation of its own cohort.

We evaluate our DTW-CP method on a binary classification task: in-hospital mortality. DTW is carried out on 52 standardized features for the training set, producing

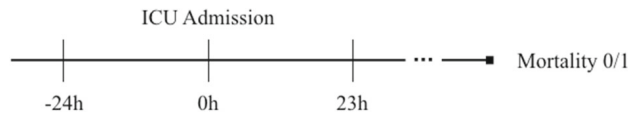


Fig. 2 Time intervals for feature extraction from an individual patient's ICU stay

52 pairwise distance matrices. We use three options for the warping path of DTW: without any constraint, Itakura parallelogram, and Sakoe–Chiba constraint with bandwidth that is half of the maximum length of the two series of interest. In the case when a constrained warping path does not exist, we use the unconstrained warping path to compute the distance. The distance matrices are then stacked into a third order tensor as described in Fig. 1 for CP decomposition. The selected number of components to decompose into is from 2 to 30. The rest of the procedure is as described in Sect. 2.2. We use logistic regression, support vector machine with linear and radial basis function kernel as our classifiers. The tuning parameters for SVM are chosen via 5-fold cross validation. We use three options of pivot patient in our experiments: (1) a random pivot such as the first patient (Sect. 4.1.1, part 1); (2) all patients as pivots (Sect. 4.1.1, part 2); (3) we choose 10 random pivots, split the training set into training and validation data and fit the models with each pivot. The one that has the best validation AUC is picked as the final pivot (Sect. 4.1.2). The metrics to evaluate the classification performance on the test data are Area under Receiver Operating Characteristic curve (AUROC, or AUC in the rest of the paper), Area Under Precision Recall curve (AUPRC) and accuracy (defined by the proportion of correct classifications). The use of AUPRC is to provide a better metric when the class distribution is imbalanced. We report the average of the above three metrics over the random splits from the test sets from each experiment.

We consider two types of comparison methods: K-nearest neighbor combined with dynamic time warping (DTW-KNN), and Long Short Term Memory (LSTM) neural networks. For DTW-KNN, we use the DTW distance computed in the previous task. For all features, we sum up the pairwise DTW distances matching the patient index: the resulting matrix is the multivariate DTW distance matrix with elements $dm_{i,j} = \sum_{k=1}^{52} d_{i,j,k}$ for patients i, j . This is equivalent to the independent multivariate DTW distance (Shokoohi-Yekta et al. 2017). We experiment KNN classifiers with $k = 1, 3, 5$.

There are numerous variations of LSTM architectures (Harutyunyan et al. 2018; Song et al. 2018). A typical LSTM application of temporal EHR data requires each patient record to have at least 24 h of records, then only take the first 24 h for modeling, indicated as the interval between '0h' to '23h' in Fig. 2. While producing good classification results with huge amount of training data, this inclusion criterion ignores patients with shorter records. We adjust this approach to make patient inclusion more flexible. For cohorts with shorter than 24 h records (day1), we make predictions on data periods of both 12 and 18 h for subjects who have at least 12 and 18 h records, respectively. For cohorts with longer records we use 12, 18, 24 h. We use the average performance over these windows as our final metric for that cohort.

We fit LSTM type models of three different architectures for the hidden layers: (1) one LSTM hidden layer; (2) two LSTM hidden layers and (3) one bidirectional LSTM hidden layer (BiLSTM). We use the rectified linear unit (ReLU) activation for hidden layers, and the sigmoid activation for the dense output layer to complete the binary classification. We test two different numbers of units, 64 and 128, for the LSTM layers. We use RMSProp as our optimizer. The batch size is fixed at 32. We train each model with 20 epochs and we use an early stopping if the validation loss stops decreasing for 5 epochs. It is uncommon to have more than two LSTM layers in practice as the number of parameters to estimate explodes. An optimal set of hyperparameters for LSTM does not exist in the literature and the impact of number of units or architecture can be insignificant (Reimers and Gurevych 2017). Our choice of configuration should be representative for this type of methods. We compute the average AUC, AUPRC and accuracy over the random splits from the test set for each window.

Software for implementation: R (version 3.6.1) has been used for data preparation, DTW (with `dtw` package created by Giorgino 2009) and classification. MATLAB Tensorlab (Vervliet et al. 2016) has been used for CP decomposition. Keras (Chollet 2015) with TensorFlow backend has been used for LSTM models.

4 Results

In this section, we report the classification performance tested on both the sepsis and the acute kidney injury (AKI) cohorts, followed by the analysis of features using data from the sepsis cohort. We answer the following questions: (1) how does our method perform in data sets with different combinations of feature sequence heterogeneity compared to the baseline methods? (2) are we able to identify features that are important for the patient representation and the classification performance?

4.1 Classification performance

4.1.1 DTW-CP performance analysis

Figure 3 compares the classification performance (measured by AUC) using DTW distances computed with three warping path options (unconstrained, Itakura parallelogram, and Sakoe–Chiba band) as features, and logistic regression (LR) and linear SVM as classifiers for the sepsis cohort for each group (sepsis 1–7, Table 2) over 10 random splits of the training and test sets. The pivot is fixed at the first patient. We use ‘group’ and ‘day’ interchangeably in the rest of the paper. On the x-axis is the number of components or latent features as predictors. Overall, different constraints do not give very different classification performance. Computation times for the constraints can be found in Sect. 4.3. When the number of components grows, the AUC increases and stabilizes for all groups. This upward-then-stable pattern is common for tensor-based phenotyping and prediction applications (Ho et al. 2014a,b). The exception is day1 (dark green) with the LR classifier: after component 20 to 25 the performance start to deteriorate yet still stays above 0.8. This was not the case for

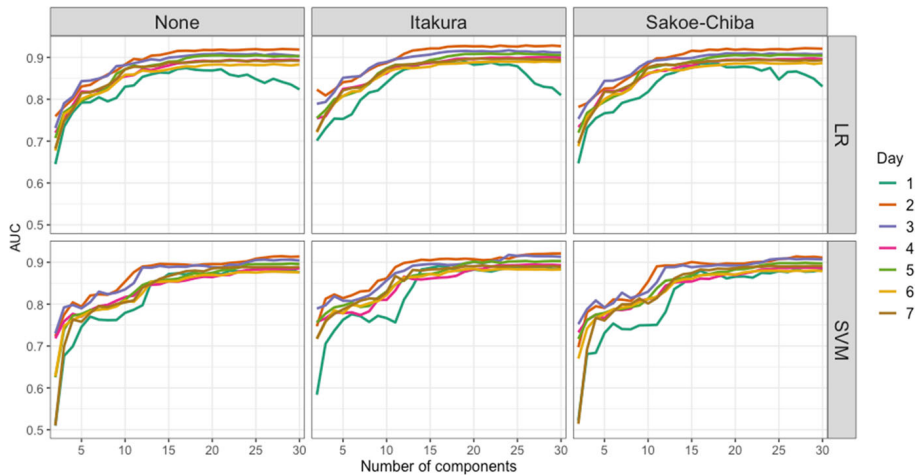


Fig. 3 AUC of DTW-CP with LR and linear SVM classifier on 1–7 groups for the sepsis cohort, fixed pivot at the first patient. DTW distances are computed with no constraint ('None'), Itakura parallelogram, and Sakoe–Chiba band. Lines represent the mean AUC over 10 randomly split test sets for each number of components from 2 to 30

the SVM classifier. This indicates that the optimal component r for day1 with LR is smaller than 30; alternatively a classifier with penalization could be applied to mitigate the problem.

Similarly, we show the mean AUC for the acute kidney injury (AKI) cohort for group 1–7 over 50 *randomly sampled* test sets under different DTW constraints and classifiers (Fig. 4). Similar to the sepsis cohort results, the AUC displays an upward-then-stable trend as the number of components increases. As the day grows (hence the variation among the sequences within the cohort) the performance slightly deteriorates. The patterns in the AKI cohort is more consistent than the sepsis cohort and less variable.

We then test the performance of DTW-CP over different choices of pivots: we carry out classification tasks using all the pivot patients in the training tensor for each day from the sepsis cohort (with only one random split) with component $R = 30$ and no DTW constraint, and report the mean and standard error of the metrics in Table 4. The results can be compared with Figs. 3 and 6. Apart from day1 where the classification performance is slightly worse and with higher LR standard errors, the other metrics fluctuate with an SE around 0.02. It is not straight-forward to identify the potential outliers in the cohort because there are multiple features, and all distances are relative to which pivot to compare with. Instead of iterating over all possible pivots, one way to choose the pivot is to randomly choose a few (for instance, 10) and pick one that produces the best validation AUC.

4.1.2 Comparison with KNN and LSTM

In this section we compare DTW-CP with KNN and LSTM methods. We make use of the procedure described in the previous section: we randomly choose 10 training pivot patients and take the pivot with best validation AUC as the optimal pivot. The

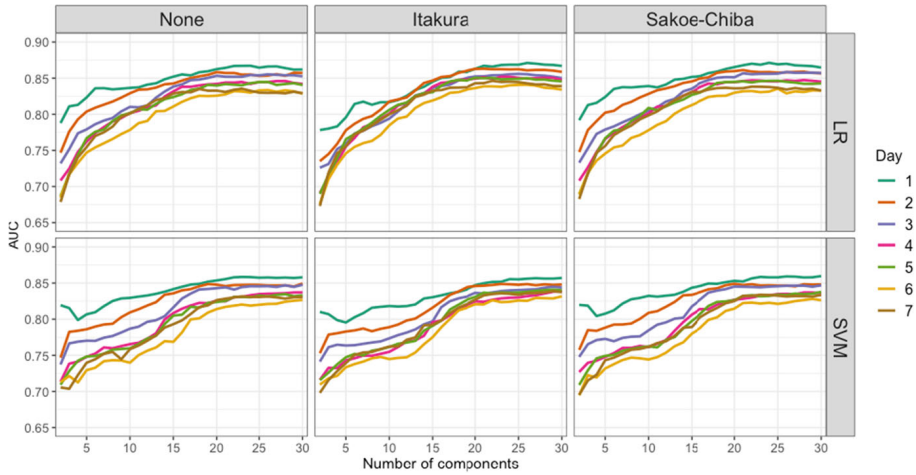


Fig. 4 AUC of DTW-CP with LR and linear SVM classifier on 1–7 groups for the AKI cohort, fixed pivot at the first patient. DTW distances are computed with no constraint ('None'), Itakura parallelogram, and Sakoe–Chiba band. Lines represent the mean AUC over 50 randomly split test sets for each number of components from 2 to 30

Table 4 Performance (mean, SE) on the sepsis cohort with different pivot patients

Metric/Classifier	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
AUC LR							
Mean	0.836	0.876	0.900	0.881	0.852	0.874	0.884
SE	0.044	0.020	0.018	0.021	0.020	0.018	0.022
AUC SVM							
Mean	0.873	0.871	0.889	0.864	0.832	0.863	0.872
SE	0.030	0.023	0.020	0.023	0.024	0.020	0.024
Accuracy LR							
Mean	0.773	0.809	0.850	0.819	0.798	0.830	0.841
SE	0.042	0.021	0.019	0.020	0.019	0.019	0.016
Accuracy SVM							
Mean	0.806	0.805	0.831	0.800	0.788	0.822	0.830
SE	0.040	0.021	0.020	0.020	0.020	0.020	0.019
AUPRC LR							
Mean	0.853	0.883	0.890	0.874	0.801	0.820	0.850
SE	0.050	0.017	0.018	0.022	0.026	0.030	0.024
AUPRC SVM							
Mean	0.897	0.876	0.879	0.848	0.778	0.802	0.827
SE	0.041	0.021	0.022	0.026	0.031	0.029	0.030

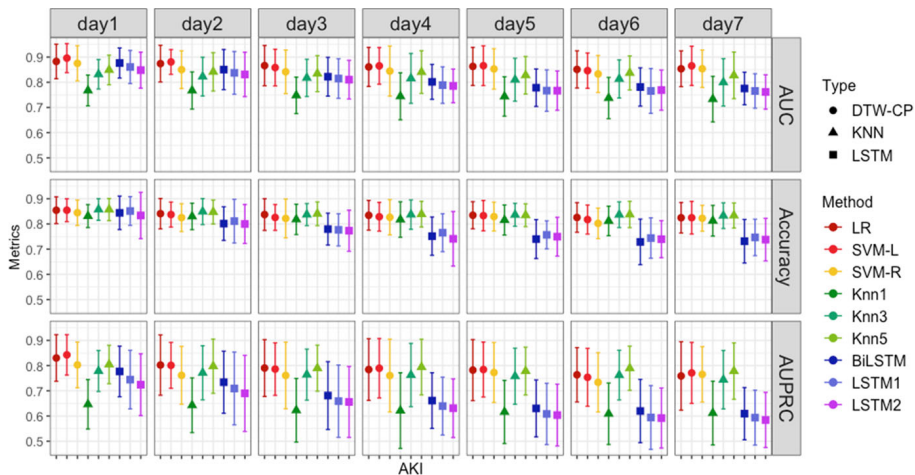


Fig. 5 Three metrics (mean, 95% CI) for the AKI cohort, comparing DTW-CP with LSTM, KNN over 7 groups. DTW-CP (LR = logistic regression, SVM-L = SVM with linear kernel, SVM-R = SVM with radial basis function kernel) performance are extracted at component 30 with unconstrained DTW

results are averaged over 10 random splits from the sepsis cohort, and 50 random splits from the AKI cohort. We focus on the case with 30 components. In Fig. 5 we illustrate the three metrics from the AKI cohort. It can be seen that for DTW-CP, LR and SVM-L classifiers produce similar results, while SVM-R is slightly worse. The impact of adding sequence length variation (from day 1 to day 7) is not as obvious as in Fig. 4, and the AUC is higher after selecting the optimal pivot. Compared with the baseline methods, DTW-CP with LR or SVM-L produce the best AUC. When it comes to the accuracy and AUPRC, DTW-CP is constantly better than LSTM methods, and has better or similar performance as the best KNN method from day 1 to day 5. As the prediction horizon increases, the deterioration of the LSTM methods is more obvious. This is not surprising, as there is not enough information to predict in a long term by using only the first 24 h without huge amounts of training data. BiLSTM seems to have the best performance among the LSTM methods.

In Fig. 6 we show the performance comparison on the sepsis cohort. DTW-CP outperforms LSTMs and KNN ($k = 1$) in all metrics on all groups except day1. It also produces better or equal performance in all metrics as the best KNN in day 3, 4 and 5. In day 2, 6 and 7, DTW-CP has comparable or marginally lower performance than the best KNN ($k = 5$) in one of the three metrics. We also observe that the performance of DTW-CP in day 1 is worse than the other groups in terms of AUC and accuracy, although the metric values are still decent. This is consistent with Fig. 3 and Table 4.

We make the following comments on the performance differences between the sepsis and the AKI cohort. With DTW-CP, when we use any random pivot (such as the first patient, Figs. 3, 4), the overall performance in terms of AUC is better in the sepsis than the AKI cohort, with an exception day1. The worse performance in sepsis-day1 compared to other days in the sepsis data is probably due to much fewer samples (only 225, see Table 2); when the sample size is larger (AKI), day1 has better performance than all other days. In all the other days, sepsis has larger sample size than AKI, which

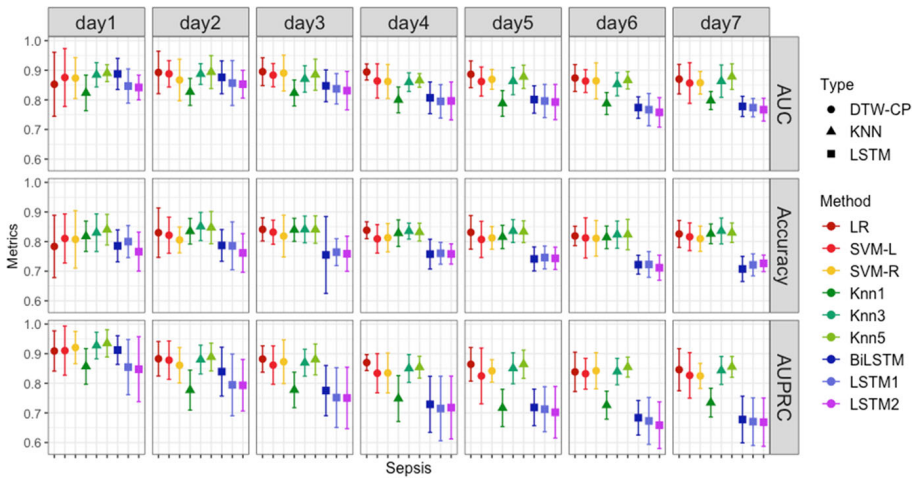


Fig. 6 Three metrics (mean, 95% CI) for the sepsis cohort, comparing DTW-CP with LSTM, KNN over 7 groups. DTW-CP performance are extracted at component 30 with unconstrained DTW

could explain why performance is still competitive when sequence length variation gets bigger. With controlled sample size in all its subgroups, the AKI cohort displays rather constant deterioration as length variation grows (day1, 2 has better AUC than day6, 7). We summarize that DTW-CP could perform better under two conditions: when there is more data, and when the sequences are shorter. If we select the pivot that produces that best validation AUC among a few randomly chosen ones, then the sequence length variation has less impact on the performance.

4.2 Analysis of feature importance

Following the good classification performance, we further investigate the interpretability using data from the sepsis cohort. The aim is to understand which features play an important role in the patient representation. We look at three aspects, namely the measurement frequency of the features, the distance matrix for one pivot patient and the learned latent feature matrix from CP decomposition. The feature names and abbreviations are consistent with Table 3.

4.2.1 Measurement frequency

The features we use vary greatly in terms of measurement frequency, and consequently, in terms of total number of measurements and length of sequence. Figure 7 illustrates the average number of measurements for patients in the sepsis cohort. The time stamp of feature recording is rounded to the nearest hour; if more than one measurement per hour is made, an average is taken. The total number of hours of patient stay in hospital or intensive care unit (length of stay, LOS) is therefore the maximum number of measurements for this patient. The cohort mean (median) length of stay is 56.65 (52) h. Vital features such as heart rate, blood pressures and oxygen saturation are

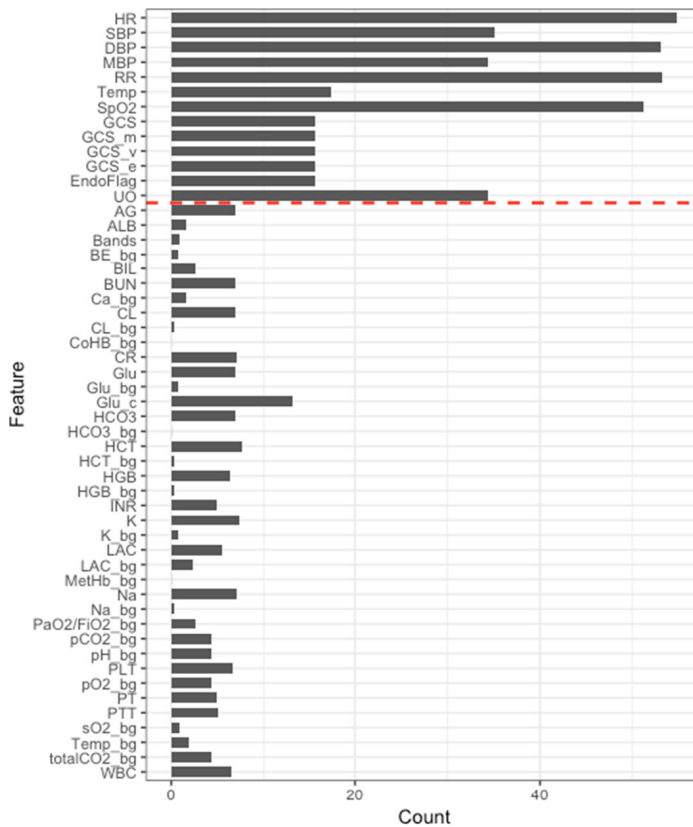


Fig. 7 Average number of total measurements for features in sepsis cohort. Cohort mean (median) length of stay is 56.65 (52) h. The red dashed line distinguishes the non-lab and lab test features

measured almost hourly while laboratory tests are taken only a few times during a patient's entire hospital stay. At the same time, even within the same feature (i.e. heart rate), the number of measurements can vary across patients given different LOS.

4.2.2 Distance matrix

To deal with the heterogeneity of time series outlined in the previous section, we work with the similarity (distance) between patient pairs computed via DTW. Figure 8 presents a heatmap for a pivot distance matrix for an arbitrary patient, as described in Fig. 1. It is important to point out that this matrix varies for different pivot patients. The X-axis represents the subject index of the cohort. Each colored element represents the DTW distance for each individual patient compared to the pivot for the corresponding feature, plotted on the y-axis. The features are ordered in the same way as Fig. 7. The top rows represent very frequently measured features (vitals and procedures) having close to zero distances with low variability, colored in deep red. Most blood gas test results (end with `_bg`) are measured very infrequently and display the same pattern as the vitals. This effect could be interpreted as follows: frequently measured features are

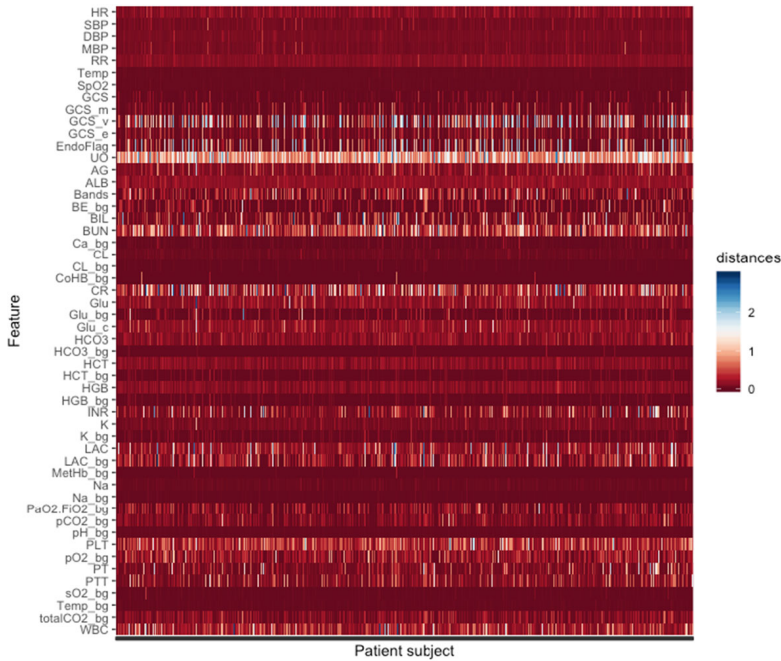


Fig. 8 Distance matrix for one patient from the sepsis cohort

vital signs that are inherently similar, hence little distance; while features with very few measurements simply contain too little information.

Noticeably, for this pivot patient the urine output, GCS measurements, blood urea nitrogen, creatinine, lactate, PaO2/FiO2 ratio, platelet counts and white blood cell count display higher distance variability, colored in white and blue. We assume features with high variance provide more information for classification.

4.2.3 Latent feature matrix

The pivot distance matrix only contains DTW distances of one particular patient compared to others in the cohort, therefore it is patient-specific. Tensor decomposition (CP) provides a useful tool to summarize information from the whole cohort. The latent feature matrix of the sepsis cohort is 52 rows (feature) by $R = 2, \dots, 30$ columns (component). By examining each component, we can identify which feature was important or unimportant by examining the magnitude of its loadings. In contrast to Principal Component Analysis (PCA), the first component from CP does not necessarily correspond to the direction explaining the largest variance: there is no ordering among the components. We illustrate with an example of three arbitrary components out of 30 from the CP decomposition in Fig. 9, as it is infeasible to visualize more than three dimensions. We normalize loadings of each component to unit length. In this particular factorization, it can be observed that most features have low factor weights or loadings, as they are concentrated around 0, and some are more spread out.

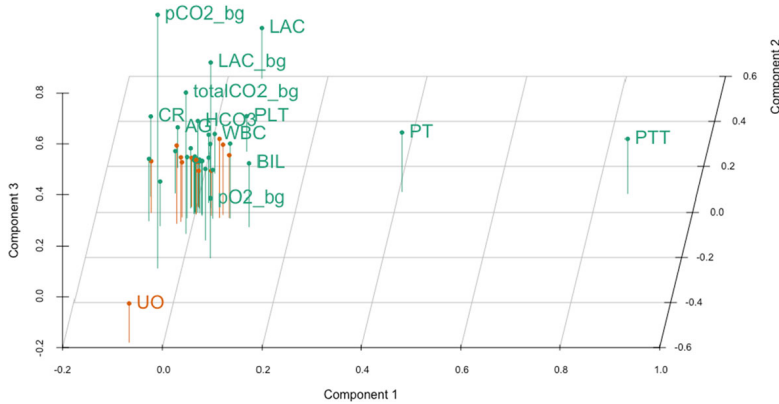


Fig. 9 Normalized factor loadings for three arbitrary components out of 30 from the CP decomposition, corresponding to the sepsis day 1 cohort. Green and Red color indicate feature categories Lab and Vital. Only loadings of magnitude greater than 0.1 in any direction is labeled for readability

To further investigate the importance of each particular feature, we calculate the average factor loading in the following way. For the decomposed feature factor matrix $M \in \mathbb{R}^{52 \times R}$ where each row corresponds to the k_{th} feature's R_{th} component loading, and we define the average loading of feature k as the average magnitude of all its R loadings. We illustrate the average loadings for the sepsis tensor decomposition over 7 days with fixed number of components, $R = 30$ in Fig. 10. Comparing with Fig. 7 it can be observed that the factor loading does not correspond with the measurement frequency: temperature and SpO2 are measured rather frequently but have low loadings across all 7 days constantly; creatinine, lactate, PaO2/FiO2 fraction are measured fewer times but have greater loadings. Regarding trend corresponding to one to seven day data, features display various patterns: increasing (GCS verbal), constant (heart rate) and decaying (lactate). This examination also reveals which features play very little role (close to zero loading for all 7 days) in the patient tensor structure.

From the factor loadings we can try to link to the physical meanings of feature importance. Urine output is measured frequently and is a marker for acute kidney injury that is associated with high hospital mortality (Legrand and Payen 2011; Zhang et al. 2014). Lactate (serum and blood gas) has both shown up as important features, and lactate level elevation is associated with increased risk of death (Sanderson et al. 2018; Filho et al. 2016; Trzeciak et al. 2007). The other features such as PaO2/FiO2 ratio (Allardet-Servent et al. 2009), glucose (Park et al. 2013), creatinine, bilirubin, platelet counts, INR (Murali et al. 2014; Li et al. 2018) are indicators for functionality in different organs, and GCS scores (Ting et al. 2010) provides information for the mobility of a patient. Our method could be one step forward to understanding which features are most indicative for classification for similar datasets, in contrast to including all features available and utilizing models with complex architecture. It is crucial to point out that physiological patterns are extremely complex especially for critically ill patients, and all interpretations are data and context dependent. Therefore any use of machine learning models need to be carefully verified by clinicians.

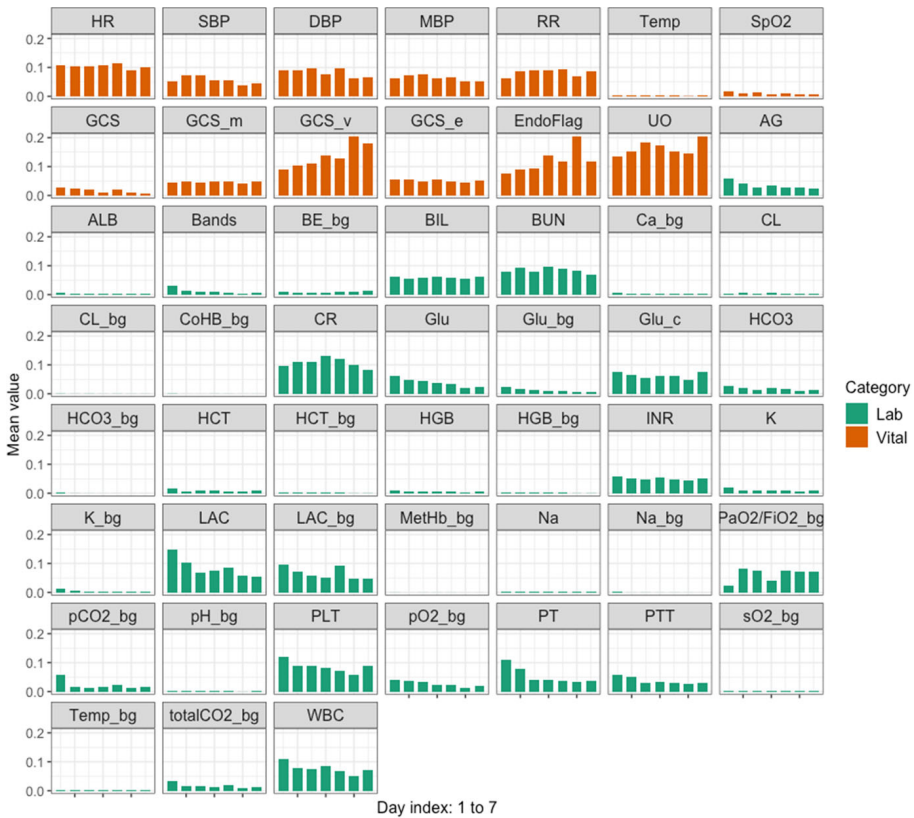


Fig. 10 Average loading for 52 features over all $R = 30$ components for the CP decomposition, sepsis data, day 1 to 7

4.3 Scalability of DTW and CP

We provide the execution time for Dynamic time warping and CP decomposition for the sepsis dataset. Computations are performed on a High Performance Computer running Red Hat Enterprise Linux 7. The hardware includes Intel®Xeon®Platinum 8160 (2.10 GHz) CPU and 1TB of RAM.

The average time for DTW computations in hours (mean, standard deviation) for all features is reported in Table 5. Itakura parallelogram and Sakoe–Chiba constraint (of bandwidth half of the maximum sequence length) improve the DTW speed compared to unconstrained DTW. The higher standard deviation in the unconstrained DTW is due to longer time required for features with longer sequences, such as heart rate (Table 6).

We also provide the time required for CP decomposition with varying size of tensors and number of components to decompose. The computation is carried out using MATLAB tensorlab toolbox. We report the execution time in seconds for the sepsis data set, day 1 to 7 subgroups (averaged over 10 random splits) where the dimension of target tensor grows from $158 \times 158 \times 52$ to $998 \times 998 \times 52$.

Table 5 Average DTW execution time (h) for 52 features

Constraint	Mean	SD
None	1.047	0.351
Itakura	0.980	0.178
Sakoe–Chiba	0.904	0.096

Table 6 CP decomposition execution time (seconds) into 10, 20, 30 components

Data index	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Size of patient mode	158	383	584	734	842	931	998
10	13.01	36.37	56.62	80.66	90.49	98.74	114.09
20	18.81	111.99	134.69	156.64	190.05	229.46	323.64
30	143.30	169.91	211.02	275.45	283.10	302.19	356.80

5 Discussions and conclusion

We have proposed a novel approach, a hybrid of dynamic time warping with tensor decomposition (DTW-CP) to tackle a prevalent but challenging issue with temporal EHR sequences: varying sampling frequency among features for different lengths of patient stays. Our approach utilizes DTW to learn information about feature similarities for patients in the cohort, and consequently uses tensor decomposition to learn the latent feature structures. In addition, we have done a detailed analysis of the temporal features used in many clinical prediction applications using the MIMIC-III database. We illustrated that the importance of a feature (i.e. high factor loading from decomposition) is not directly related to how often it is measured, and linked the ‘important’ features to their clinical interpretations.

Among all the works using DTW or tensor decomposition in healthcare, we are the first to combine these two. Moreover, we have extended the DTW time period to up to seven days, and illustrated how classification performance changes with different variation in sequence length. We carried out careful experiments using (1) distance matrices computed by different DTW constraints (Itakura parallelogram, Sakoe–Chiba band versus unconstrained DTW); (2) different pivot options; (3) different classifiers (logistic regression, linear and radial basis function kernel SVM). By comparing with two baseline methods: LSTM with three architectures, and DTW-KNN methods, we have shown that our method is able to outperform them in three different metrics. We also give interpretations of the classification performance with different data sets and different settings.

DTW-CP is a similarity (distance) based approach, this has two implications. Firstly it is necessary to compute the distance between all pairs of patients in the cohort for each feature. This step can be time consuming when the sequences are long and when the cohort is large, as pointed out in Moor et al. (2019) (who did not use any constraint, but used fastDTW in their implementation). Although DTW computation time can be reduced with constraints, it can only be used when the sequence length do not differ much; also it is unclear which constraint is the best (Geler et al. 2019). Secondly, the

interpretation of features is based on patient similarity instead of the feature value themselves. This means there is always a need for pivot patient to compare the rest of the cohort with, to make the interpretation meaningful. We have chosen to optimize the choice of pivot based on maximizing AUC. This choice should of course be guided by which metric is most important for any given application.

Our choice of decomposition algorithm (CP) does not have non-negative constraint, hence the interpretation of latent feature matrix distinguishes itself from Ho et al. (2014a, b); Afshar et al. (2018) and others where each component is a combination of positive phenotype memberships. There is no standard way to choose the number of components to decompose into, hence we suggest that in practice this should be where the classification performance stabilizes. Lastly, we have only utilized temporal EHR sequences. Most works on patient clustering and clinical event predictions include static demographic data in addition to the dynamic data (Suresh et al. 2018; Purushotham et al. 2018), thus combining static data with temporal sequences is a direction we could investigate further. Possible solutions include coupled matrix and tensor factorization (CMTF).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Acar E, Levin-Schwartz Y, Calhoun VD, Adali T (2017) Tensor-based fusion of EEG and FMRI to understand neurological changes in schizophrenia. In: Proceedings—IEEE international symposium on circuits and systems, pp 1–4. <https://doi.org/10.1109/ISCAS.2017.8050303>
- Afshar A, Perros I, Papalexakis EE, Searles E, Ho J, Sun J (2018) COPA: constrained PARAFAC2 for sparse and large datasets. In: The 27th ACM international conference on information and knowledge management (CIKM '18). <https://doi.org/10.1145/3269206.3271775>
- Alaa AM, van der Schaar M (2018) Forecasting individualized disease trajectories using interpretable deep learning. [arXiv:1810.10489](https://arxiv.org/abs/1810.10489)
- Allardet-Servent J, Forel JM, Roch A, Guervilly C, Chiche L, Castanier M, Embriaco N, Gainnier M, Papazian L (2009) FiO₂ and acute respiratory distress syndrome definition during lung protective ventilation. *Crit Care Med* 37(1):202–207. <https://doi.org/10.1097/CCM.0b013e31819261db>
- Bagnall A, Bostrom A, Large J, Lines J (2016) The great time series classification bake off: an experimental evaluation of recently proposed algorithms. Extended version [arXiv:1602.01711](https://arxiv.org/abs/1602.01711)
- Che C, Xiao C, Liang J, Jin B, Zho J, Wang F (2017) An RNN architecture with dynamic temporal matching for personalized predictions of Parkinson's disease. In: Proceedings of the 2017 SIAM international conference on data mining, pp 198–206. <https://doi.org/10.1137/1.9781611974973.23>
- Chollet F (2015) Keras. <https://keras.io>
- Dau HA, Silva DF, Petitjean F, Forestier G, Bagnall A, Keogh E (2017) Judicious setting of Dynamic Time Warping's window width allows more accurate classification of time series. In: Proceedings—2017 IEEE international conference on big data, big data 2017. <https://doi.org/10.1109/BigData.2017.8258009>
- Fanaee-T H, Oliveira M, Gama J, Malinowski S, Morla R (2013) Event and anomaly detection using tucker3 decomposition. In: Proceedings of 20th European conference on artificial intelligence (ECAI'2013)-ubiquitous data mining workshop, vol 1, pp 8–12. [arXiv:1406.3266v1](https://arxiv.org/abs/1406.3266v1)
- Filho RR, Rocha LL, Correa TD, Pessoa CMS, Colombo G, Assuncao MSC (2016) Blood lactate levels cutoff and mortality prediction in sepsis—time for a reappraisal? A retrospective cohort study. *Shock* 46(5):480–485. <https://doi.org/10.1097/SHK.0000000000000667>

- Geler Z, Kurbalija V, Ivanovic M, Radovanovic M, Dai W (2019) Dynamic time warping: Itakura vs Sakoe–Chiba. In: IEEE international symposium on innovations in intelligent systems and applications, INISTA 2019—Proceedings. <https://doi.org/10.1109/INISTA.2019.8778300>
- Ghassemi M, Naumann T, Schulam P, Beam AL, Ranganath R (2018) Opportunities in machine learning for healthcare. [arXiv:1806.00388](https://arxiv.org/abs/1806.00388)
- Giorgino T (2009) Computing and visualizing dynamic time warping alignments in R: the dtw package. *J Stat Softw* 31(7):1–24. <https://doi.org/10.18637/jss.v031.i07>
- Guo C, Lu M, Chen J (2020a) An evaluation of time series summary statistics as features for clinical prediction tasks. *BMC Med Inform Decis Mak* 20(1):1–20. <https://doi.org/10.1186/s12911-020-1063-x>
- Guo D, Duan G, Yu Y, Li Y, Wu FX (2020b) A disease inference method based on symptom extraction and bidirectional Long Short Term Memory networks. *Methods* 173(April 2019):75–82. <https://doi.org/10.1016/j.ymeth.2019.07.009>
- Harutyunyan H, Khachatryan H, Kale DC, Steeg GV, Galstyan A (2018) Multitask learning and benchmarking with clinical time series data. [arXiv:1703.07771](https://arxiv.org/abs/1703.07771)
- Henderson J, Ho JC, Kho AN, Denny JC, Malin BA, Sun J, Ghosh J (2017) Granite: diversified. Sparse tensor factorization for electronic health record-based phenotyping. In: IEEE international conference on healthcare informatics (ICHI). <https://doi.org/10.1109/ICHI.2017.61>
- Henderson J, Malin BA, Ho JC (2018) PIVETed-granite: computational phenotypes through constrained tensor factorization. [arXiv:1808.02602v1](https://arxiv.org/abs/1808.02602v1)
- Ho J, Ghosh J, Steinhilber SR, Stewart WF, Denny JC, Malin BA, Sun J (2014a) Limestone: high-throughput candidate phenotype generation via tensor factorization. *J Biomed Inform* 52:199–211. <https://doi.org/10.1016/j.jbi.2014.07.001>
- Ho J, Ghosh J, Sun J (2014b) Marble: high-throughput phenotyping from electronic health records via sparse nonnegative tensor factorization. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 115–124. <https://doi.org/10.1145/2623330.2623658>
- Itakura F (1975) Minimum prediction residual principle applied to speech recognition. *IEEE Trans Acoust Speech Signal Process* 23(1):67–72. <https://doi.org/10.1109/TASSP.1975.1162641>
- Johnson AE, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, Moody B, Szolovits P, Celi AL, Mark RG (2016) MIMIC-III, a freely accessible critical care database. *Sci Data* 3:160035. <https://doi.org/10.1038/sdata.2016.35>. <https://www.nature.com/articles/sdata201635>
- Johnson AEW, Pollard TJ, Mark RG (2017) Reproducibility in critical care: a mortality prediction case study. In: 2nd machine learning for healthcare conference, PMLR, vol 68. <http://proceedings.mlr.press/v68/johnson17a.html>
- Kaji DA, Zech JR, Kim JS, Cho SK, Dangayach NS, Costa AB, Oermann EK (2019) An attention based deep learning model of clinical events in the intensive care unit. *PLoS ONE* 14(2):1–17. <https://doi.org/10.1371/journal.pone.0211057>
- Keogh EJ, Pazzani MJ (1999) Scaling up dynamic time warping to massive datasets. *Princ Data Min Knowl Discov* 1704(Derriere):1–11. https://doi.org/10.1007/978-3-540-48247-5_1
- Kiers HAL (2000) Towards a standardized notation and terminology in multiway analysis. *J Chemom* 14:105–122
- Kolda TG, Bader BW (2009) Tensor decompositions and applications. *SIAM Rev* 51(3):455–500. <https://doi.org/10.1137/07070111X>
- Kruse CS, Goswamy R, Raval Y, Marawi S (2016) Challenges and opportunities of big data in health care: a systematic review. *JMIR Med Inform* 4(4):e38. <https://doi.org/10.2196/medinform.5359>
- Le Gall JR, Lemeshow S, Saulnier F (1993) Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study. *JAMA* 270(24):2957–2963
- Legrand M, Payen D (2011) Understanding urine output in critically ill patients. *Ann Intensive Care* 1(1):13. <https://doi.org/10.1186/2110-5820-1-13>. <http://www.annalsintensivecare.com/content/1/1/13>
- Lei L, Zhou Y, Zhai J, Zhang L, Fang Z, He P, Gao J (2018) An effective patient representation learning for time-series prediction tasks based on EHRs. In: Proceedings—2018 IEEE international conference on bioinformatics and biomedicine, BIBM 2018. <https://doi.org/10.1109/BIBM.2018.8621542>
- Li Y, Chaiteerakij R, Kwon JH, Jang JW, Lee HL, Cha S, Ding XW, Thongprayoon C, Ha FS, Nie CY, Zhang Q, Yang Z, Giama NH, Roberts LR, Han T (2018) A model predicting short-term mortality in patients with advanced liver cirrhosis and concomitant infection. *Medicine* 97(41):e12758

- Lin YW, Zhou Y, Faghri F, Shaw M, Campbell R (2019) Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. *PLoS ONE* 14(7):e0218942. <https://doi.org/10.1371/journal.pone.0218942>
- Lipton ZC (2016) The mythos of model interpretability. *arXiv:1606.03490*
- Lipton ZC, Kale DC, Elkan C, Wetzel R (2016) Learning to diagnose with LSTM recurrent neural networks. In: 4th international conference on learning representations, ICLR 2016—conference track proceedings, pp 1–18. *arXiv:1511.03677*
- Luo Y, Xin Y, Joshi R, Celi L, Szolovits P (2016) Predicting ICU mortality risk by grouping temporal trends from a multivariate panel of physiologic measurements. In: 30th AAAI conference on artificial intelligence, AAAI 2016, pp 42–50
- Moor M, Horn M, Rieck B, Roqueiro D, Borgwardt K (2019) Early recognition of sepsis with Gaussian process temporal convolutional networks and dynamic time warping. *arXiv:1902.01659*
- Muller M (2007) Dynamic time warping. In: *Information retrieval for music and motion*, Springer, Berlin, Heidelberg, chap 4, pp 69–84
- Murali AR, Devarbhavi H, Venkatachala PR, Singh R, Sheth KA (2014) Factors that predict 1-month mortality in patients with pregnancy-specific liver disease. *Clin Gastroenterol Hepatol* 12(1):109–113. <https://doi.org/10.1016/j.cgh.2013.06.018>
- Niennattrakul V, Ratanamahatana CA (2009) Learning DTW global constraint for time series classification. *arXiv:0903.0041*
- Park BS, Yoon JS, Moon JS, Won KC, Lee HW (2013) Predicting mortality of critically ill patients by blood glucose levels. *Diabetes Metab J* 37:385–390
- Perros I, Papalexakis EE, Wang F, Vuduc R, Searles E, Thompson M, Sun J (2017) SPARTan: scalable PARAFAC2 for large and sparse data. In: *KDD*. <https://doi.org/10.1145/3097983.3098014>
- Purushotham S, Meng C, Che Z, Liu Y (2018) Benchmarking deep learning models on large healthcare datasets. *J Biomed Inform* 83:112–134. <https://doi.org/10.1016/j.jbi.2018.04.007>
- Rabanser S, Shchur O, Günnemann S (2017) Introduction to tensor decompositions and their applications in machine learning, pp 1–13. *arXiv:1711.10781*
- Ratanamahatana CA, Keogh E (2004) Making time-series classification more accurate using learned constraints. In: *SIAM proceedings series*, pp 11–22. <https://doi.org/10.1137/1.9781611972740.2>
- Reimers N, Gurevych I (2017) Optimal hyperparameters for deep LSTM-networks for sequence labeling tasks. *arXiv:1707.06799*
- Ribas Ripoll VJ, Vellido A, Romero E, Ruiz-Rodríguez JC (2014) Sepsis mortality prediction with the quotient basis kernel. *Artif Intell Med* 61(1):45–52. <https://doi.org/10.1016/j.artmed.2014.03.004>
- Ruffini M, Gavaldà R, Limón E (2017) Clustering patients with tensor decomposition 68. <https://doi.org/10.1002/dei>. *arXiv:1708.08994*
- Sakoe H, Chiba S (1978) Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans Acoust Speech Signal Process ASSP* 26(1):43–49
- Salvador S, Chan P (2007) FastDTW: toward accurate dynamic time warping in linear time and space. *Intell Data Anal* 11(5):561–580. <https://doi.org/10.3233/ida-2007-11508>
- Sanderson M, Chikhani M, Blyth E, Wood S, Moppett IK, McKeever T, Simmonds MJR (2018) Predicting 30-day mortality in patients with sepsis: an exploratory analysis of process of care and patient characteristics. *J Intensive Care Soc* 19(4):299–304. <https://doi.org/10.1177/1751143718758975>
- Scherpf M, Gräßer F, Malberg H, Zaunseder S (2019) Predicting sepsis with a recurrent neural network using the MIMIC III database. *Comput Biol Med* 113(June):103395. <https://doi.org/10.1016/j.combiomed.2019.103395>
- Shokoohi-Yekta M, Hu B, Jin H, Wang J, Keogh E (2017) Generalizing DTW to the multi-dimensional case requires an adaptive approach. *Data Min Knowl Disc* 31(1):1–31. <https://doi.org/10.1007/s10618-016-0455-0>
- Sidiropoulos ND, De Lathauwer L, Fu X, Huang K, Papalexakis EE, Faloutsos C (2017) Tensor decomposition for signal processing and machine learning. *IEEE Trans Signal Process* 65(13):3551–3582. <https://doi.org/10.1109/TSP.2017.2690524>. *arXiv:1607.01668*
- Song H, Rajan D, Thiagarajan JJ, Spanias A (2018) Attend and diagnose: clinical time series analysis using attention models. In: 32nd AAAI conference on artificial intelligence, AAAI 2018, pp 4091–4098. *arXiv:1711.03905*
- Suresh H, Gong JJ, Guttaj J (2018) Learning tasks for multitask learning: heterogeneous patient populations in the ICU. In: *KDD*. <https://doi.org/10.1145/3219819.3219930>. *arXiv:1806.02878*

- Tan CW, Petitjean F, Webb GI (2019) FastEE: fast ensembles of elastic distances for time series classification. *Data Min Knowl Discov*. <https://doi.org/10.1007/s10618-019-00663-x>
- Ting H, Chen M, Hsieh Y, Chan C (2010) Good mortality prediction by Glasgow Coma scale for neuro-surgical patients. *J Chin Med Assoc* 73(3):139–143. [https://doi.org/10.1016/S1726-4901\(10\)70028-9](https://doi.org/10.1016/S1726-4901(10)70028-9)
- Trzeciak S, Dellinger RP, Chansky ME, Arnold RC, Schorr C, Milcarek B, Hollenberg SM, Parrillo JE (2007) Serum lactate as a predictor of mortality in patients with infection. *Intensive Care Med* 33:970–977. <https://doi.org/10.1007/s00134-007-0563-9>
- Vervliet N, Debals O, Sorber L, Van Barel M, De Lathauwer L (2016) Tensorlab 3.0
- Xiao C, Choi E, Sun J (2018) Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *J Am Med Inform Assoc* 25(10):1419–1428. <https://doi.org/10.1093/jamia/ocy068>
- Yu K, Zhang M, Cui T, Hauskrecht M (2020) Monitoring ICU mortality risk with a long short-term memory recurrent neural network. *Pac Symp Biocomput* 25:103–114. https://doi.org/10.1142/9789811215636_0010
- Zhang Z, Xu X, Ni H, Deng H (2014) Urine output on ICU entry is associated with hospital mortality in unselected critically ill patients. *J Nephrol* 27:65–71. <https://doi.org/10.1007/s40620-013-0024-1>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Chi Zhang¹  · Hadi Fanaee-T² · Magne Thoresen¹

Chi Zhang
chi.zhang@medisin.uio.no

Hadi Fanaee-T
hadi.fanaee@hh.se

Magne Thoresen
magne.thoresen@medisin.uio.no

¹ Department of Biostatistics, Domus Medica, University of Oslo, Oslo, Norway

² Center for Applied Intelligent Systems Research, Halmstad University, Halmstad, Sweden

Paper II

Effects of patient-level risk factors, departmental allocation and seasonality on intrahospital patient transfer patterns - network analysis applied on a Norwegian single-centre dataset

Chi Zhang, Torsten Eken, Silje Bakken Jørgensen, Magne Thoresen, Signe Søvik

Accepted for publication in: *BMJ Open* (2022), 0:e054545. DOI: 10.1136/bmjopen-2021-054545.



Effects of patient-level risk factors, departmental allocation and seasonality on intrahospital patient transfer patterns: Network analysis applied on a Norwegian single-centre dataset

Chi Zhang¹, Torsten Eken^{2,3}, Silje Bakken Jørgensen⁴, Magne Thoresen¹, Signe Søvik^{3,5}

¹Oslo Centre for Biostatistics and Epidemiology, Department of Biostatistics, University of Oslo, Oslo, Norway

²Department of Anaesthesia and Intensive care medicine, Division of Emergencies and Critical Care, Oslo University Hospital Ullevål, Oslo, Norway

³Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

⁴Department of Microbiology and Infection Control, Akershus University Hospital, Lørenskog, Norway

⁵Department of Anaesthesia and Intensive care medicine, Akershus University Hospital, Lørenskog, Norway

Corresponding author Signe Søvik

Address: Department of Anaesthesia and Intensive Care
Akershus University Hospital, NO-1478 Lørenskog, Norway
Phone: +47 915 73 914 e-mail: signe.sovik@medisin.uio.no

Clinical trial number and registry URL: Not applicable

Prior Presentations: None

Acknowledgements: Haldor Husby at the Akershus University Hospital's Department for Data Extraction and Analysis provided invaluable assistance

Word counts: Abstract (max 300): 302
Main text: 4407

Number of Figures: 3

Number of Tables: 3

Abbreviated title: Intra-hospital patient transfer network analysis

Funding: Financial support was provided by the authors' institutions. Chi Zhang was funded by the Norwegian Research Council (BigInsight project, grant #237718)

Conflicts of interests: None declared

Abstract:

Objectives: Describe patient transfer patterns within a large Norwegian hospital. Identify risk factors associated with a high number of transfers. Develop methods to monitor intrahospital patient flows to support capacity management and infection control.

Design: Retrospective observational study of linked clinical data from electronic health records.

Setting: Tertiary care university hospital in the Greater Oslo Region, Norway

Participants: All adult (≥ 18 y) admissions to the Gastroenterology, Gastrointestinal surgery, Neurology and Orthopaedics departments at Akershus University Hospital, June 2018-May 2019.

Methods: Network analysis and graph theory. Poisson regression analysis.

Outcome measures: Primary outcome was network characteristics at the departmental level. We describe location-to-location transfers using unweighted, undirected networks for a full-year study period. Weekly networks reveal changes in network size, density and key categories of transfers over time. Secondary outcome was transfer trajectories at the individual patient level. We describe the distribution of transfer trajectories in the cohort and associate number of transfers with patient clinical characteristics.

Results: The cohort comprised 17,198 hospital stays. Network analysis demonstrated marked heterogeneity across departments and throughout the year. The Orthopaedics department had largest transfer network size and density and greatest temporal variation. More transfers occurred during weekdays than weekends. Summer holiday affected transfers of different types (ED-any location / bed ward-bed ward / to-from technical wards) differently.

Over 75% of transferred patients followed one of twenty common intrahospital trajectories, involving 1–3 transfers. Higher number of intrahospital transfers was associated with emergency admission (transfer rate ratio RR=1.827), non-prophylactic antibiotics (RR=1.108), surgical procedure (RR=2.939), and stay in ICU or high-dependency unit (RR=2.098). Additionally, gastrological (RR=1.211), orthopaedic (RR=1.295) and neurological (RR=1.114) patients had higher risk of many transfers than gastroenterology patients (all effects: $p < 0.001$).

Conclusions: Network and transfer chain analysis applied on patient location data revealed logistic and clinical associations highly relevant for hospital capacity management and infection control.

Strengths and limitations of this study

- Strengths of this study include its comprehensive dataset with time-stamped patient-level information on intrahospital transfers, admission type, demographics, physiological derangement and antibiotic use
- Both static and temporal network analysis methods were applied to capture different aspects of patient flow
- Regression techniques complemented the network analyses, assessing associations between patient-level risk factors and longer intrahospital transfer trajectories
- Limitations of the study include a twelve-month dataset, hampering robust analysis of yearly seasonality
- Data were limited to patients in four hospital departments, precluding network analysis at an all-hospital or inter-hospital level

1. Introduction

Hospitals are complex systems which must run smoothly to ensure treatment quality and patient safety. Transfer of patients between specialised departments is a key part of hospital operation, and optimisation of patient flows is crucial for hospital capacity management and infection control.

Patients' journeys through the hospital may be analysed from different viewpoints. At a *systemic* level, assessment of overall transfer patterns makes it possible to identify logistic problems and make adequate adjustments. Improved bed utilisation to counter unanticipated under- and over-capacity in different hospital wards may reduce variability of workload and stress for hospital personnel. However, placement of patients in inappropriate settings may reduce quality of care and increase the risk of errors.¹ Physically moving patients also carries a risk of introducing infectious agents to new staff and hospital areas. Certain locations and the transfers between them may be crucial for hospital operation, and some pathways may be especially vulnerable to e.g., understaffing or closure during an infection outbreak.² At an *individual* level, each transfer requires handover of medical information, its quality and completeness being essential for error avoidance and continuity of care. Hospitalised patients who are frequently transferred have increased risk of falls, delirium, prolonged hospital length of stay, healthcare-associated infections and mortality.³⁻⁶ Characterisation of transfer patterns both at a systemic and an individual level may thus be relevant for understanding and revising healthcare service use. For organisational planning, smoothed, long-term transfer data showing major patient flows and seasonality is key. In contrast, real-time data reveals the extremes and allows for immediate intervention, e.g., when monitoring detects excessive staff workload or single patients subjected to transfer pathways known to carry unacceptable risk.

Intrahospital patient transfers have been studied using graph theory and network analysis. Construction of a weighted, directed transfer network describing the emergency surgical services in a UK hospital identified potential hubs and bottlenecks in the system.² Static and temporal transfer networks of patient flow in two UK acute care hospitals were evaluated and related to emergency department performance.⁷ Inter-departmental patient transfer networks for five European hospitals were constructed and used in a simulation study of infection spread among high- and low-risk patient groups.⁸ Network analysis has also been applied on national,⁹ regional¹⁰ and simulated¹¹ patient transfer data to elucidate spread of resistant microbes.

The above studies almost exclusively evaluated transfer networks for entire hospital systems. Most studies analysed networks as static entities, without attention to possible temporal changes in size or connectivity. Furthermore, network analysis alone is insufficient to describe individual patient trajectories, since in this method each patient's intrahospital journey is broken up into a number of separate location-to-location moves. All transfers are then analysed collectively without regard to their sequence.

In this study we examined patient transfers in a large Norwegian hospital using electronic health record data. Our primary objective was to describe intrahospital transfer patterns at a systemic level. To this end we applied network analysis on all transfers in four hospital departments, highlighting the heterogeneity of transfer patterns across departments and over time. Our secondary objective was to evaluate transfers on an individual level. We identified typical and atypical transfer trajectories and assessed whether patient characteristics, including admission type, age, gender, surgery, antibiotic usage and physiological derangement were associated with a higher number of intrahospital transfers.

2. Methods

Hospital characteristics

This retrospective, observational cohort study analysed data from Akershus University Hospital (AUH), Norway, a tertiary hospital serving a population of 560,000 within the Greater Oslo Region. Patients who need cardiac surgery and neurosurgery or suffer from major trauma are referred elsewhere. In 2019, AUH had 763 somatic (non-psychiatric) beds, 66,280 somatic admissions, 33,886 day cases, and 366,858 somatic ambulatory consultations. The Emergency Department (ED) is an integrated division of the hospital and predominantly receives urgent cases arriving by ambulance and pre-screened patients from Local Emergency Medical centres. Approximately 75% of patients presenting in the ED are transferred to other hospital wards. Two surgical suites together provide 22 operating rooms (ORs). Two mixed medical-surgical intensive care units (ICUs), one cardiac high-dependency unit (HDU) and one mixed postoperative care unit / HDU together provide 14 invasive and 8 non-invasive ventilator beds.

Data collection

The study was approved by the Regional Committee for Medical and Health Research Ethics (REK SørØst C Ref.no. 33192) and was considered exempt from patient consent requirements by the institutional Data Protection Officer (Ref.no. 2019/56). Pseudonymized data was extracted on 6 Dec 2019 by the AUH Department for Data Extraction and Analysis and stored and processed within the Service for Sensitive Data at the University of Oslo. The study period was 365 days starting at a Monday in June 2018; the exact week number was not released to the authors for privacy protection reasons. For time-stamped data, the granularity of time is hourly.

All adult (≥ 18 years) admissions to any ward in the departments of Gastroenterology, Gastrointestinal surgery, Neurology or Orthopaedics in the study period were included. The four departments, two medical and two surgical, each containing 1–3 wards (Table 1 & S1), were selected because they treat defined patient groups that were expected to differ from each other. Study sample size was not pre-defined. The cohort only contained admissions occurring after the study start time and hence excluded patients who were already admitted. Stays with incomplete or erroneous data (e.g., missing length of stay, negative time durations) were excluded. For patients with multiple hospital stays within the study period, each stay was treated as a unique event.

We extracted time stamped patient location data throughout each stay to construct individual intrahospital transfer trajectories. The following information was also extracted from AUH electronic health records: demographics (age, gender), admission type (elective or emergency), time of hospital admission, physiological derangement measured as National Early Warning Score 2 (NEWS2¹²), administration of antibiotics (excluding surgical prophylaxis), and whether the patient underwent a surgical procedure. The term “patient record” refers to all data collected during a stay at the hospital.

Key variable definitions

At AUH, NEWS2 is routinely scored in the ED and three times daily in bed wards. The maximum and mean NEWS2 for each patient during their first 48 hours of stay were used in analyses, to utilise NEWS2 as a marker of physiologically deranged state around hospital admission. Any non-prophylactic antibiotics usage during a stay was coded as a binary *yes-no* variable. Length of hospital stay was converted to days. ICUs, HDUs and ORs were collectively denoted *Technical* wards. A surgical procedure was assumed for stays with OR or Day Surgery Unit in the location log. In line with Norwegian hospital routine, the transition from one hospital day to the next was defined to occur at 07:00 hours. Weekdays and weekends were Monday–Friday and Saturday–Sunday, respectively.

Intrahospital transfer, or transfer for brevity, is a patient movement from one physical location (ward, ED, etc.) to another. For perioperative transfers, we chose to combine multiple *consecutive* patient movements between the pre- and post-operative HDU and the OR into a single location, ORBLOCK, to avoid inflating the number of transfers.

Transfer patterns

We describe intrahospital transfer patterns at hospital departmental level, using networks, and at individual patient level using transfer chains. In network analysis and graph theory,^{13,14} a network is a graph that contains two types of elements: *vertices* (or nodes) and *edges*. A vertex represents the elementary unit of the system, and an edge captures the interaction between two different units. The edge can be directed or undirected. If two vertices are connected more than once, a weight can be assigned to the edge between them. Network *density* is defined as the ratio of the number of existing edges over the sum of all possible edges for all vertices. *Degree* of a vertex is the number

of other vertices it is connected with. Taking into account whether each connected vertex is on the ‘from’ or ‘to’ side of the edge, *out-degree* and *in-degree* for a vertex can be computed.

In this study, vertices were hospital locations patients had visited, and edges were the transfers between any two locations. Imagine an emergency patient who needs surgery and therefore is transferred from the ED to a bed ward, then to the OR, then back to the same bed ward, and eventually discharged home. This transfer history can be constructed into a network of three vertices and three edges, if we ignore the final discharge to home. For each of the four departments studied we first constructed an unweighted, undirected network to explore global connectivity, disregarding timing, frequency and type of transfers. We further constructed more detailed networks by letting them change with time, from study week 1 through 52, and from weekday to weekend. Finally, we examined the temporal frequency of specific types of transfers. Here, we categorised all edges in the networks into three broad transfer groups: *ED–Any* (transfers from the ED to any other ward), *Bed ward–Bed ward* (transfers not involving technical wards), and *Technical* (transfers involving technical wards, i.e., ICUs, HDUs and ORs).

Network analysis only captures the grand total of location-to-location transfers and is insufficient to examine individual patients’ transfer trajectories. We therefore extracted the transfer chain for each stay, keeping the sequence of locations. Variables of interest were the actual transfer sequences themselves and the length of the chains, i.e., the number of transfers.

Network and statistical analysis

Network size was quantified by number of vertices and edges (unique locations and transfers). For the 52 weekly networks in the temporal network analyses, we report mean and standard deviation (SD) of weekly vertex and edge counts. Network density was computed for the undirected, unweighted networks. In- and out-degrees for vertices were computed for directed, unweighted networks in the weekday–weekend network comparison. Descriptive statistics of patient cohort characteristics are provided as counts, percentages or medians (10th–90th percentile) as appropriate. Frequencies of the various transfer chains were examined and the most common types of chains listed.

Two multivariate Poisson regression models were used to identify risk factors associated with higher number of intrahospital transfers. Explanatory variables used in both models were age

(categorised as 18–39, 40–64, 65–84, and 85+ years), gender, admission type (elective vs. emergency), departmental allocation, physiological derangement (mean first 48-h NEWS2, categorised as 0–2, 3–4, 5–6, and 7+), and whether non-prophylactic antibiotics were administered during the stay. In the second model we also included variables indicating treatment (having undergone surgery, having a stay in an HDU or ICU). Interaction terms between departmental allocation, surgery, and antibiotics use were modelled. Results are reported as transfer rate ratios (RR) with 95% confidence intervals (CI). P-values less than 0.01 are considered statistically significant.

All analyses were implemented in the statistical software R (version 3.4.2). Network analyses and visualisation were conducted using packages `igraph`¹⁵ (<https://igraph.org>) and `ggraph`.¹⁶

3. Results

Patient cohort and hospital locations

After processing, the cohort contained 17,198 unique records. Table 1 summarizes cohort demographics and locations visited, stratified by whether the stay involved surgery.

The Gastroenterology department had fewest admissions (N=1712, 10%); the other three departments had between 4788 and 5522 admissions. Surgical procedures were rare for stays in the Neurology department (1%) and common in the Orthopaedics department (61%). Overall, across departments, 63%–83% of patient stays were non-elective, i.e., emergency admissions starting in the ED.

The proportion of stays with non-prophylactic antibiotics administered varied from 11% (neurological patients not undergoing surgery) to 57% (gastroenterological patients undergoing surgery). Antibiotics use was more common for stays with surgery, irrespective of department. LOS and maximum NEWS2 during the first 48 hours of stay were higher in stays with surgery, except for in the Neurology department. Stays with surgery also on average comprised two more unique intrahospital locations and more transfers than stays without surgery (median 3 times versus 1). Overall, 0.5% of patients experienced eight or more transfers. Maximum transfer count varied markedly between the medical departments (eight) and the surgical departments (twenty-three).

Transfer networks

A total of 1940 (11%) stays comprised only one intrahospital location and were excluded from network analysis. In total, 35,001 location-to-location transfers were found for the remaining 15,258 patient stays. Figure 1 displays department-wise static networks. Vertex colours indicate ward types. In general, the ED, OR, HDUs and ICUs had many connections with wards in all studied departments. Many emergency patients ultimately allocated to one of the four studied departments were initially transferred from the ED to the OR, an HDU, or one of a wide range of surgical and medical wards belonging to other departments. The Orthopaedic department network was most densely connected, comprising 28 locations and 155 unique location-to-location transfer pathways, giving a network density of 0.410. The Neurology department network was the least densely connected, with 20 vertices, 55 edges and network density 0.288. Despite a much larger

cohort size (4788 vs. 1712 stays), the Gastroenterology department had network size and density very similar to Neurology.

Figure 2A visualises week-by-week edge (location) and vertex (unique transfer pathway) counts throughout the study period, stratified by hospital department. The Gastrosurgery and Orthopaedics networks contained many more transfer pathways than the other two departments. The Orthopaedics network also displayed marked temporal variations over the year. A dip in connectivity around study week 3–10 could have been due to less elective surgery and closure of wards during summer holidays.

Figure 2B displays week-by-week number of transfers, stratified by transfer type and department. In Figure 2C these data are normalised by number of admissions during that week in the corresponding department. *ED–Any* type transfers denote emergency hospital admissions and were relatively constant over time for all departments. The neurology department had fewest elective admissions, thus its normalised *ED–Any* was close to 1. Counts of *Bed ward–Bed ward* transfers also were rather constant and did not constitute much of the traffic. In contrast, *Technical* type transfers, involving transfers to and from ICUs, HDUs and ORs, showed distinct temporal variation and lower activity during the summer holidays.

Network connectivity varied during the week (Figure 3). On average, networks included more locations (vertices) and almost twice as many unique location-to-location transfer pathways (edges) on weekdays as during weekends. A majority of hospital locations visited by our patient cohort received patients from more locations (higher in-degree) and transferred patients to more locations (higher out-degree) on weekdays than on weekends. Adjusted for number of patients present (bed occupancy was higher on weekdays than weekends), number of transfers was still higher on weekdays. In contrast, the number of unique locations used by patients ultimately allocated to one of the four studied departments was higher on weekends.

The ED had a very large out-degree but zero in-degree, as this ward feeds patients to many locations but receives no patients from other hospital wards. Conversely, the “home” wards for our patient cohort (Gastroenterology, Gastrosurgery A/B, Neurology A/B, Orthopaedic A/B/C) received their patients from more locations than they transferred patients to. Home wards thus

“assembled” patients from the ED, OR, HDUs and any “ad hoc” wards, ultimately for patient discharge to home.

Patient transfer chain analysis

The 15,258 patient stays comprising more than one intrahospital location followed 1118 unique transfer chains, i.e., sequences of locations. Chain utilisation was highly skewed: 75% of transferred patients followed one of the top twenty (1.8%) transfer chains (Table 2). The three most common transfer chains, from the ED to one of the two neurological bed wards or the ED observation unit, together represented one third of transferred patients. Ten out of the twenty most common transfer chains involved only one transfer and started in the ED. The sub-pattern *Bed ward–ORBLOCK–Bed ward* occurred in nine out of the twenty most common transfer chains.

In contrast, the majority of unique transfer chains occurred infrequently: 10% of patient stays (1505 out of 15,258) followed one of 976 uncommon patterns (87% of all types), each occurring ≤ 7 times over the one-year period. Compared to the majority, in this 10%-subgroup patients stayed at a higher number of unique hospital locations (4 (2–5) vs. 3 (2–3)) and more often underwent surgery (64% vs. 49%) and advanced treatment in an HDU or ICU (42% vs. 23%). Also, in this 10%-subgroup, 8% of patient stays from the Gastrosurgery or Orthopaedics department also involved stays in medical bed wards, as opposed to 0% among the remaining surgical patient stays.

Two multivariate Poisson regression models identified risk factors associated with a higher number of intrahospital transfers (Table 3). In the first model, older age was negatively associated with more transfers, while higher first-48h mean NEWS2 was associated with more transfers. The effect increased from NEWS2 score 0–2 via 3–4 to 5–6, from where it levelled off. The effects of age and NEWS2 were no longer significant when treatment in the OR or HDU/ICU was adjusted for (Model 2 in Table 3). Gender did not contribute significantly in either model.

Emergency hospital admission and antibiotics use were associated with increased risk of undergoing more transfers, as was treatment in the OR or an HDU/ICU. Although much of the increased risk was explained by these factors, admission to surgical departments (Gastrosurgery and Orthopaedics) in itself increased the risks of more transfers. Modelled interactions between departmental allocation, surgery and antibiotics use were not significant (Table S2, Supplementary material).

4. Discussion

The main finding in this retrospective study applying network analysis on patient location data in four hospital departments was a marked heterogeneity in patient transfer patterns. Departments differed markedly regarding network size and density, transfer types, and temporal changes over the week and year.

Why network analysis of patient location data?

Given the range of health services offered to different patient populations, patient flows within hospitals would be expected to vary widely between departments and even wards. Optimisation of patient logistics is key to reduce delays and overcrowding, and thus time and health care costs. Availability of beds in wards specialised for each patient's medical condition likely reduces errors and improves quality of care. Detailed knowledge of highly connected hospital hubs and patient trajectories is also important for prevention and control of hospital infections.^{2,3}

Heterogeneity in size and connectivity of transfer networks

In all departments studied, a majority of stays were emergency admissions via the ED. As in previous work¹, the number of emergency admissions was relatively constant over time (Figure 2B & C). The ED acted as a hub feeding patients to their allocated department's "home" wards (Figure 1). However, networks revealed that Gastrointestinal and Orthopaedic patients also to a large degree were treated in surgical wards in other surgical departments. "Home" and "non-home" wards alike transferred patients to and from the OR. Likely, the large proportion of emergency admissions at AUH intermittently caused patient surges, overcrowding, and patients being placed in any suitable ward with a free bed and only later transferred to a ward in their allocated department. Gastrointestinal and Orthopaedic patients, many of whom are multimorbid, also had stays in a number of medical wards (Figure 1). This resulted in the two surgical departments having larger and 2.5–3 times more densely connected transfer networks than those of the two medical departments, which may have treated more homogenous patient populations. Although Neurology had many more admissions than Gastroenterology, the two networks were very similar in number of locations and connectivity.

These findings illustrate that patient flows in one department may be heavily affected by logistic changes implemented in seemingly unconnected departments. Weighted and directed networks would provide important additional information, useful for real-time monitoring of patient flows.

Temporal variation in patient transfer networks

Monitoring of temporal changes in patient transfer networks is relevant for capacity planning, but in-depth organisational knowledge of studied departments is required for interpretation of findings to be reliable. We saw marked heterogeneity across hospital departments regarding temporal variability (Figure 2A). Week-to-week variation in number of transfers was much larger in the Orthopaedics department than in Gastrosurgery, despite the two networks having similar edge and vertex counts when averaged over the year. The contrast between the two surgical departments and Neurology was pronounced.

Higher temporal variability in the Orthopaedics department seemed to reside in transfers involving ORs, ICUs and HDUs (Figure 2B) and partially reflected the weekly number of admitted patients (Figure 2C). Both surgical departments had a drop in transfers during summer holidays, when fewer elective surgical procedures are performed. In the Gastroenterology department, some change in patient logistics must have been implemented around Christmas, i.e., study week 26. Similar effects of organisational changes have been reported in UK acute care data.⁷

Network connectivity also changed over the week. On average, studied hospital wards were connected to almost twice as many locations during the week than on weekends (Figure 3). The “assembling” function of “home” wards, i.e., wards belonging to the four studied departments (higher in-degree than out-degree) also was less marked on weekends. Admissions occurring on weekends have been shown to more often result in transfer to the ICU and to be associated with increased adjusted mortality rates.^{17,18}

Individual patient transfer trajectories

Standardised patient trajectories facilitate hospital logistics and specialised treatment. Network analysis however examines the total number of transfers and does not capture their sequence in individual patients.^{3,4} Moreover, in some health care systems the format of patient location data does not facilitate analysis of entire patient trajectories, and data validity may be poor.¹

Core hospital pathways manage a majority of patients.⁷ We found that 11% of stays involved only one location. A further 67% of stays followed one of twenty common patient transfer chains, half of which started in the ED and involved only one transfer (Table 2).

In contrast, a substantial minority of patient stays represented a very high number of uncommon, non-standardised hospital location sequences. These uncommon transfer chains included more locations and more often multiple OR visits and ICU / HDU stays. Among the stays following the 10% least common transfer chains, 8% of patients allocated to one of the two surgical departments (Orthopaedics or Gastrosurgery) also had stays in medical bed wards. In contrast, surgical patients following the 90% most common transfer chains had no medical ward stays. Multimorbidity thus seemed to predispose for non-standard needs, which again is known to carry higher risk of unwanted outcomes.³⁻⁶

Although not necessarily causal, the factors associated with higher number of intrahospital transfers in our regression analysis are clinically recognisable as proxy variables for more complex hospital stays. Caution must be used when interpreting effect sizes, since there could have been interdependence between variables. The regression model controlling for age and gender showed an increase in number of transfers with increasingly deranged physiological state early in the hospital stay, quantified as mean NEWS2 during the first 48 h. The effect levelled out at NEWS2 of five or higher, values often associated with transfer to more advanced care¹². In the model that also adjusted for treatment in an HDU, ICU or OR, the statistical contribution of NEWS2 was no longer detectable. The effect of these two variables could thus not be disentangled by our analysis.

Non-prophylactic antibiotics use was associated with more patient transfers. This variable could have acted as proxy for e.g., bowel anastomosis leakage or postoperative wound infections needing repeated surgical treatment, and postoperative pneumonia needing advanced monitoring or mechanical ventilatory support. Interestingly, stays in surgical departments were associated with increased number of intrahospital transfers even after statistical adjustment for clinical risk factors, OR and ICU treatment.

Implications for clinicians and policymakers

Analysis of intrahospital patient transfer networks is relevant for design of new hospital buildings and allocation of hospital areas for essential units, e.g., those acting as hubs. Proximity between

wards frequently connected by transfers may increase efficiency. In wards with known high connectivity, planning more isolation beds might be prudent to shield vulnerable individuals and prevent outbreaks.

Ongoing monitoring of the connectivity (in-degree and out-degree) of individual hospital wards is highly relevant for infection prevention and control. When new pathogens emerge simultaneously in different wards with no apparent linkage, network and transfer sequence analysis may reveal possible transmission routes that can be controlled. To limit a hospital outbreak, it may be useful to identify units so frequently connected by transfers that they should be regarded as equally exposed to an infectious agent.

Patient transfer is often necessary for diagnostics or specialised treatment, but intrahospital transfers may also result from foreseeable and preventable factors such as seasonal overcrowding and staffing shortages, construction work, or wards being closed during infection outbreaks. Evaluation of factors resulting in transfer peaks might motivate improved institutional preparedness. Placing patients in inappropriate specialty areas increases the risk of medical errors when staff are exposed to unfamiliar medical conditions, treatments, or devices. Real-time transfer analysis may identify and warn hospital managers about unusual, potentially high-risk transfer sequences.

The methods applied in this study could be used to monitor patient flows, predict likely logistic problems and routes of infection spread, and develop plans for optimising placement of patients deemed at risk for long and complicated hospital stays. There is a need for standardised indicators of patient flow logistics to facilitate comparison between institutions and health systems.¹

Strengths and limitations

A one-year study period prevented analysis of long-term trends. We only had data for adult patients allocated to four selected hospital departments. Short-term patient movement, e.g., for medical imaging or diagnostic procedures, was not studied. Generalisability of our findings may be limited to similar health care systems.

Strengths of this study include that complete, high time-resolution datasets comprising both elective and emergency admissions were evaluated on a departmental and ward level. Transfers

were categorised by subtype, and individual transfer trajectories were associated with key clinical patient characteristics. Our methodological approaches should be transferable to new settings.

Conclusion

Temporal network analysis applied on departmental and ward level provides insight into the heterogeneity of intrahospital patient transfers. The method is a potential tool for continuous, automated monitoring of patient flows. Analysis of typical and atypical patient transfer trajectories is a useful supplement. Obvious areas of benefit are hospital capacity management across wards and departments, and infection prevention and control.

Areas remaining for future research include patient and systemic factors that may predict and prevent extremely long transfer trajectories. Frequent changes of intrahospital location may negatively affect important aspects of patients' experience of their care, such as quality and consistency of medical information given and confidence in hospital staff.¹⁹ Intrahospital transfer patterns should therefore be studied also in view of patient-reported outcome measures.

6. Authors' contributions

MT and TE: Devised the overall project. TE: Defined the dataset, oversaw data extraction and interpretation, contributed to data restructuring. SS and SBJ: Conceptualised the study. ZC: Designed data analysis plan, cleaned and analysed data, made all figures, created tables. ZC and SS: Interpreted results and drafted the manuscript. MT: Oversaw analysis and interpretation. All authors critically evaluated and discussed ongoing analyses, critically revised the manuscript, and approved the final version.

7. Data sharing statement

The data that support the findings of this study are de-identified patient records available from Akershus University Hospital (AUH) Department for Data Extraction and Analysis, but due to Norwegian legislation restrictions apply to their availability. The data was used under license for the current study and so are not publicly available. Data are however available from the authors upon reasonable request provided that permission is obtained from the South-East Norway Regional Committee for Medical and Health Research Ethics and the AUH Privacy Ombudsman for Research.

8. License statement

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in BMJ Open and any other BMJ products and to exploit all rights, as set out in our licence.

9. Patient and public involvement

A representative from the AUH user involvement committee evaluated and approved of the project prior to data release.

10. References

1. Villa S, Prenestini A, Giusepi I. A framework to analyze hospital-wide patient flow logistics: Evidence from an Italian comparative study. *Health Policy* 2014; **115**: 196–205.
2. Kohler K, Ercole A. Can network science reveal structure in a complex healthcare system? A network analysis using data from emergency surgical services. *BMJ Open* 2020; **10**: e034265.
3. Boncea EE, Expert P, Honeyford K, *et al.* Association between intrahospital transfer and hospital-acquired infection in the elderly: A retrospective case–control study in a UK hospital network. *BMJ Qual Saf* 2021; doi: 10.1136/bmjqs-2020-012124
4. Bristol AA, Schneider CE, Lin S-Y, Brody AA. A Systematic review of clinical outcomes associated with intrahospital transitions. *J Healthc Qual* 2019; **42**: 175–187.
5. Bush K, Barbosa H, Farooq S, *et al.* Predicting hospital-onset *Clostridium difficile* using patient mobility data: A network approach. *Infect Control Hosp Epidemiology* 2019; **40**: 1380–1386.
6. McHaney-Lindstrom M, Hebert C, Flaherty J, Mangino JE, Moffatt-Bruce S, Root ED. Analysis of intra-hospital transfers and hospital-onset *Clostridium difficile* infection. *J Hosp Infect* 2018; **102**: 168–169.
7. Bean DM, Stringer C, Beeknoo N, Teo J, Dobson RJB. Network analysis of patient flow in two UK acute care hospitals identifies key sub-networks for A&E performance. *PLoS One* 2017; **12**: e0185912.
8. Tahir H, López-Cortés LE, Kola A, *et al.* Relevance of intra-hospital patient movements for the spread of healthcare-associated infections within hospitals - a mathematical modeling study. *PLoS Comput Biol* 2021; **17**: e1008600.
9. Nekkab N, Crépey P, Astagneau P, Opatowski L, Temime L. Assessing the role of inter-facility patient transfer in the spread of carbapenemase-producing Enterobacteriaceae: The case of France between 2012 and 2015. *Sci Rep* 2020; **10**: 14910.
10. Piotrowska MJ, Sakowski K, Lonc A, Tahir H, Kretzschmar ME. Impact of inter-hospital transfers on the prevalence of resistant pathogens in a hospital–community system. *Epidemics* 2020; **33**: 100408.
11. Vilches TN, Bonesso MF, Guerra HM, Fortaleza CMCB, Park AW, Ferreira CP. The role of intra and inter-hospital patient transfer in the dissemination of healthcare-associated multidrug-resistant pathogens. *Epidemics* 2019; **26**: 104–15.

12. Royal College of Physicians. National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS. Updated report of a working party. 2017. <https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2> (accessed May 31, 2021).
13. Barabasi A-L. Network Science. Cambridge University Press, 2016.
14. Newman MEJ. The Structure and Function of Complex Networks. *Siam Rev* 2003; **45**: 167–256.
15. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal Complex Systems* 2006. http://interjournal.org/manuscript_abstract.php?361100992.
16. Pedersen TL. ggraph: An implementation of grammar of graphics for graphs and networks. 2021. <https://cran.r-project.org/package=ggraph> (accessed May 31, 2021).
17. Vest-Hansen B, Riis AH, Sørensen HT, Christiansen CF. Out-of-hours and weekend admissions to Danish medical departments: Admission rates and 30-day mortality for 20 common medical conditions. *BMJ Open* 2015; **5**: e006731.
18. Duvald I, Moellekaer A, Boysen MA, Vest-Hansen B. Linking the severity of illness and the weekend effect: A cohort study examining emergency department visits. *Scand J Trauma Resusc Emerg Medicine* 2018; **26**: 72.
19. Raleigh VS, Frosini F, Sizmur S, Graham C. Do some trusts deliver a consistently better experience for patients? An analysis of patient experience across acute care surveys in English NHS trusts. *BMJ Qual Saf* 2012; **21**: 381.

11. Legends

Figure 1 Patient transfer networks in four hospital departments

Unweighted, undirected transfer networks for four hospital departments over a one-year period. Vertex (location) colours distinguish different functionality, i.e., ED, ORs, ICUs, medical and surgical wards. Vertex size is proportional to its degree (number of other locations connected to it). Network sizes are given as edge and vertex counts (E, V) and density. PHDU, Postoperative HDU; EDOU, Emergency Department observation unit; GI S, Gastrointestinal surgery; Day S, Day surgery; Mix S, Mixed surgery; TCVS, Thoracic and Cardiovascular surgery. Complete list of abbreviations in Table S1 (Supplementary materials).

Figure 2 Temporal changes in network size by hospital department and transfer type

A: Weekly network sizes in terms of transfer pathway (edge) and location (vertex) counts. **B:** Weekly sum of transfers, split by transfer type. **C:** Weekly sum of transfers by type, normalised by number of patient admissions in the corresponding department that week. Study week is counted from a Monday in June 2018; hence study weeks 1–13 denote June-August, and so forth. *ED–Any*, transfers from the ED to any other ward; *Bed ward–Bed ward*, transfers between regular wards; *To–from Technical*, transfers involving technical wards, i.e., ICUs, HDUs and ORs.

Figure 3 Transfer network connectivity on weekdays and weekends

For thirty hospital wards, daily average number of hospital locations the ward received patients from (In degree, green dots) and sent patients to (Out degree, amber dots). Data for all stays allocated to any of the four studied departments, split by weekday / weekend. Full-year network size (all four departments) reported as mean (SD) of edge (E) and vertex (V) counts.

Table 1 Clinical and location characteristics of study cohort

Department			Gastroenterology		Gastrosurgery		Neurology		Orthopaedics	
Admissions (N)			1712		5522		4788		5176	
Surgery			Yes	No	Yes	No	Yes	No	Yes	No
N, p1%			69 4.0	1643 96.0	1942 35.2	3580 64.8	46 1.0	4742 99.0	3171 61.3	2005 38.7
Emergency admission (N, p1%, p2%)			49 2.9 71.0	1059 61.9 64.4	1029 18.6 53.0	3043 55.1 85.0	45 9.3 97.8	3944 82.4 83.2	1643 31.7 51.8	1605 31.0 80.0
Antibiotic use (N, p1%, p2%)			39 2.3 56.5	337 19.7 20.5	663 12.0 34.1	1176 21.3 32.8	9 1.9 19.6	527 11.0 11.1	1272 24.6 40.1	317 6.1 15.8
Cohort characteristics										
Age			65 36–81	65 30–85	58 28–80	62 31–83	51 35–75	64 32–84	68 40–85	69 35–88
NEWS2 score			4 1.0-7.7	2 0-6	3 1-6	2 0-6	1 0-3	2 0-5	3 1-6	2 0-6
LOS (days)			5.3 1.2–13 44	2.0 0.5–7.9 86	3.9 1.1–13 184	2.0 0.5–7.1 90	2.2 0.8–7.2 49	2.8 0.7–10 113	4.3 1.3–12 84	1.2 0.3–5.8 43
Unique wards			4 3–5 6	2 1–3 4	4 3–5 7	2 1–3 5	4 3–5 6	2 1–2 6	4 3–5 9	2 1–3 5
Individual transfers			3 1–4 7	1 0–2 6	3 2–4 21	1 0–2 9	2 2–3 4	1 0–1 8	3 2–5 23	1 0–2 6
Ward	Type	N								
Emergency Dept	ED	12370	1101		4058		3980		3231	
Operating Room	OR	5032	67		1828		45		3092	
Day Surgery		235	2		119		1		113	
Postop. HDU	Tech	5444	144		2018		99		3183	
General ICU		134	8		75		24		27	
Medical ICU		201	86		45		47		23	
Cardiac HDU		2					1		1	
ED Obs. Unit		2036	531		747		54		704	
Hemodialysis		8	2						6	
Orthopedic A	Surg	2144			11		2		2131	
Orthopedic B		1849			13				1836	
Orthopedic C		603	1		9				593	
Gastrosurgery A		2328	5		2301		2		20	
Gastrosurgery B		2488	4		2462		1		21	
Mixed Surgery		197	2		101		1		93	
Urology		534	4		386		2		142	
ThoracoVascular		496			166		1		329	
Neurology A	Med	2521	2		1		2516		2	
Neurology B		2245	5		1		2237		2	
Neuro Rehab		306					306			
Gastroenterology		1263	1246		15				2	

Palliation A	8		8		
Geriatrics	5	1	2		2
Palliation B	11				11
Infection/Hema	14	8	4		2
Infection A	20	8	8	1	3
Cardiac	9		3	1	5
Cardiac/Renal	8	3	3	1	1
Pulmonary A	2		1		1
Pulmonary B	1				1

Upper panel: Cohort summary for four hospital departments, stratified by whether patient stay involved surgery. Antibiotics usage excludes surgical antimicrobial prophylaxis. N: number of patient stays. p1%: percentage of all patient stays in this department. p2%: percentage of patient stays in this department with same surgery status. **Mid panel:** Patient characteristics in each sub-cohort. NEWS2, National Early Warning Score 2, maximum value in first 48 h; LOS, Length of hospital stay; Unique wards, Number of unique wards visited during each patient stay. Age and NEWS2 are reported as median and 10th-90th percentiles. LOS, Unique wards and number of transfers are reported as median, 10th-90th percentiles and maximum. **Lower panel:** Number of visits to each of 30 observed wards, by patients' allocated department.

Table 2 The twenty most common intrahospital transfer trajectories

Location sequence	N	%	Cum %
ED – Neurology A	2015	13.3	13.3
ED – ED Observation Unit	1544	10.1	23.4
ED – Neurology B	1508	9.9	33.3
ED – Gastroscopy A	872	5.7	39.0
ED – Gastroscopy B	866	5.7	44.7
ED – Orthopaedic A – ORBLOCK – Orthopaedic A	474	3.1	47.8
ED – Gastroenterology A	470	3.1	50.9
Orthopaedic C – ORBLOCK – Orthopaedic C	429	2.8	53.7
ED – Orthopaedic B – ORBLOCK – Orthopaedic B	413	2.7	56.4
ED – Orthopaedic B	391	2.6	59.0
Orthopaedic B – ORBLOCK – Orthopaedic B	370	2.4	61.4
ED – Gastroscopy B – ORBLOCK – Gastroscopy B	349	2.3	63.7
Gastroscopy B – ORBLOCK – Gastroscopy B	325	2.1	65.8
ED – Orthopaedic A	324	2.1	67.9
Gastroscopy A – ORBLOCK – Gastroscopy A	309	2.0	69.9
Orthopaedic A – ORBLOCK – Orthopaedic A	293	1.9	71.8
ED – Gastroscopy A – ORBLOCK – Gastroscopy A	180	1.2	73.0
ED – Urology	154	1.0	74.0
ED – Neurology A – Neurology B	153	1.0	75.0
ED – ThoracoVascular	118	0.8	75.8
Total of 15258 patients	11557	75.8	75.8

The twenty most common out of a total of 1118 transfer chains observed in all 15258 patient stays in the departments of Gastroenterology, Gastrointestinal surgery, Neurology, and Orthopaedic surgery over a one-year study period. ED, Emergency department; ORBLOCK, pre- / postoperative High Dependency Unit in combination with OR treatment.

Table 3 Poisson regression analysis on number of intrahospital transfers per stay

Risk factors	Model 1			Model 2			
	RR	95% CI	P value	RR	95% CI	P value	
Age	18 – 39	Reference		Reference			
	40 – 64	0.984	0.949 – 1.021	0.405	1.017	0.980 – 1.055	0.382
	65 – 84	0.925	0.892 – 0.959	<0.001	0.982	0.947 – 1.019	0.344
	85+	0.835	0.793 – 0.880	<0.001	0.960	0.911 – 1.011	0.125
NEWS2 ¹	0 – 2	Reference		Reference			
	3 – 4	1.071	1.027 – 1.117	0.001	0.984	0.943 – 1.027	0.470
	5 – 6	1.138	1.051 – 1.231	0.001	0.956	0.882 – 1.034	0.270
	7+	1.132	0.988 – 1.289	0.068	0.801	0.699 – 0.914	0.001
Gender	Female	Reference		Reference			
	Male	0.984	0.961 – 1.009	0.205	0.997	0.973 – 1.021	0.786
Dept.	Gastroenterology	Reference		Reference			
	Gastrosurgery	1.679	1.590 – 1.773	<0.001	1.210	1.144 – 1.280	<0.001
	Neurology	1.039	0.980 – 1.102	0.199	1.117	1.053 – 1.184	<0.001
	Orthopaedics	2.406	2.281 – 2.540	<0.001	1.294	1.222 – 1.372	<0.001
Admission	Elective	Reference		Reference			
	Emergency	1.388	1.347 – 1.440	<0.001	1.834	1.778 – 1.892	<0.001
Antibiotics ²	No	Reference		Reference			
	Yes	1.372	1.336 – 1.409	<0.001	1.107	1.077 – 1.138	<0.001
Been to OR ³	No	Reference		Reference			
	Yes				2.936	2.846 – 3.029	<0.001
Been to ICU ⁴	No	Reference		Reference			
	Yes				2.106	2.025 – 2.189	<0.001

RR, patient transfer rate ratio. ¹Mean first 48-h NEWS2 score. ²Use of any non-prophylactic antibiotics.

³Indicates a surgical procedure. ⁴Stayed in an ICU or HDU, indicates a severe patient condition.

Figure 1 Patient transfer networks in four hospital departments

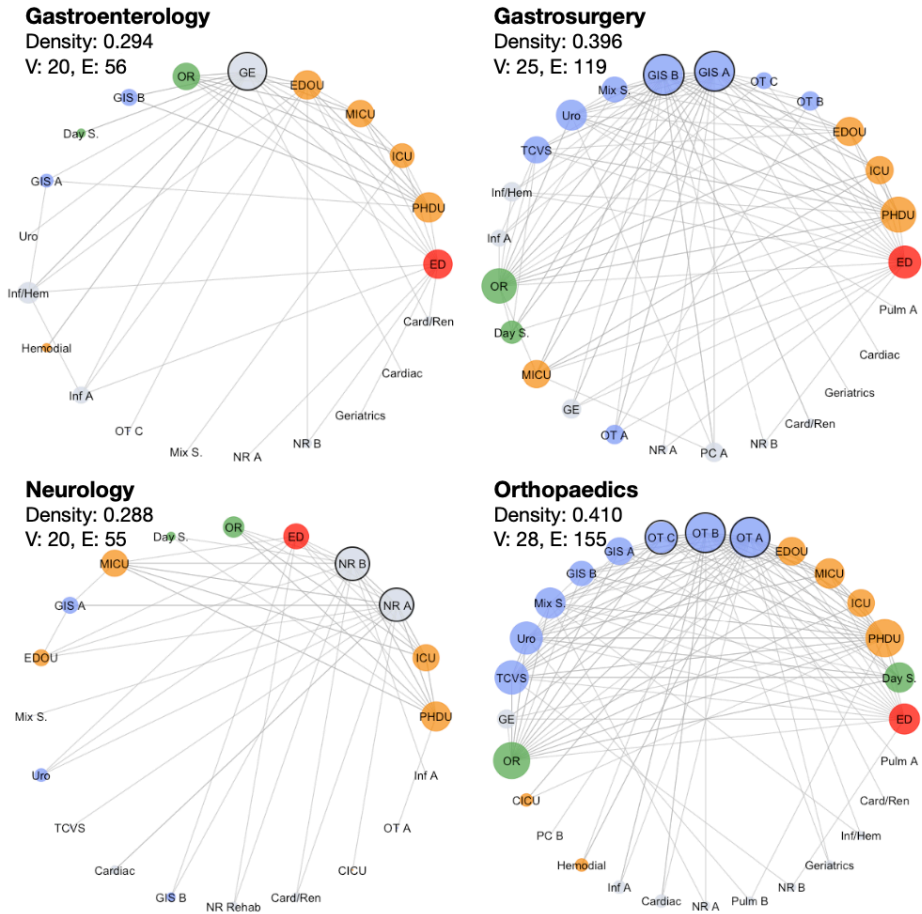


Figure 2 Temporal changes in network size by hospital department and transfer type

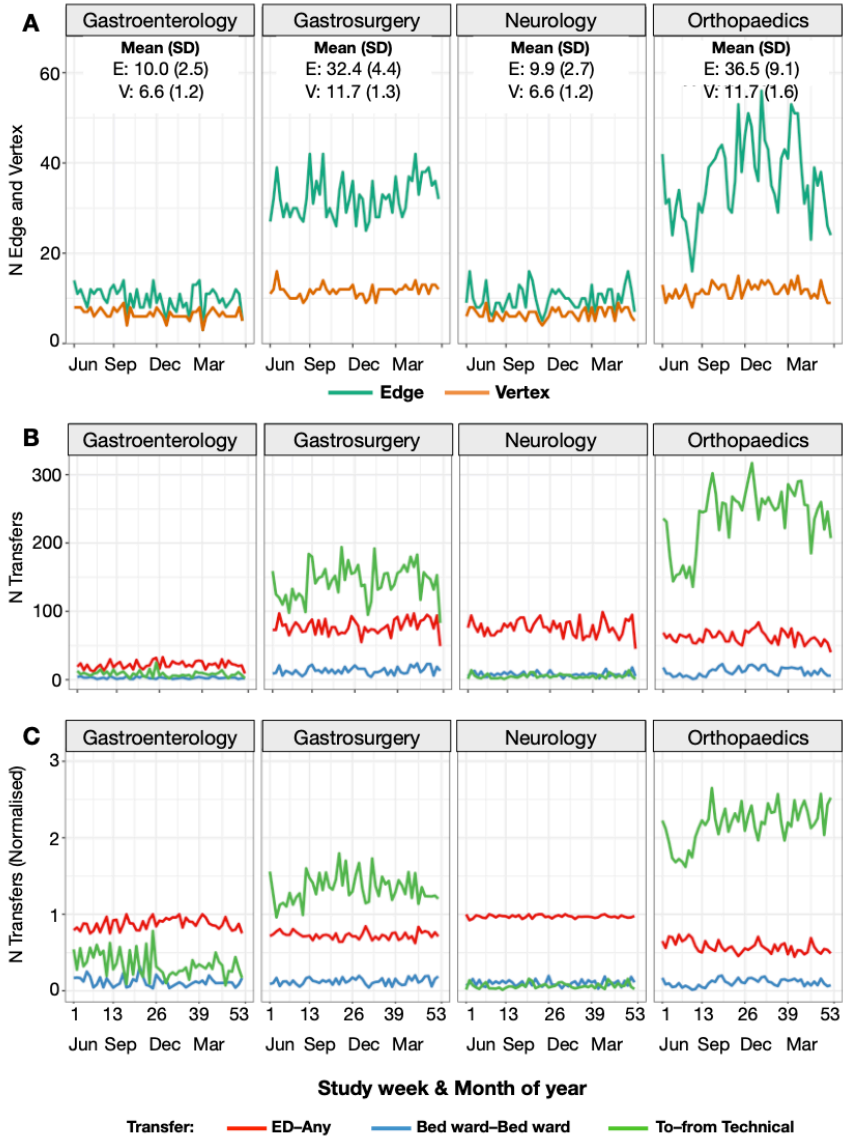
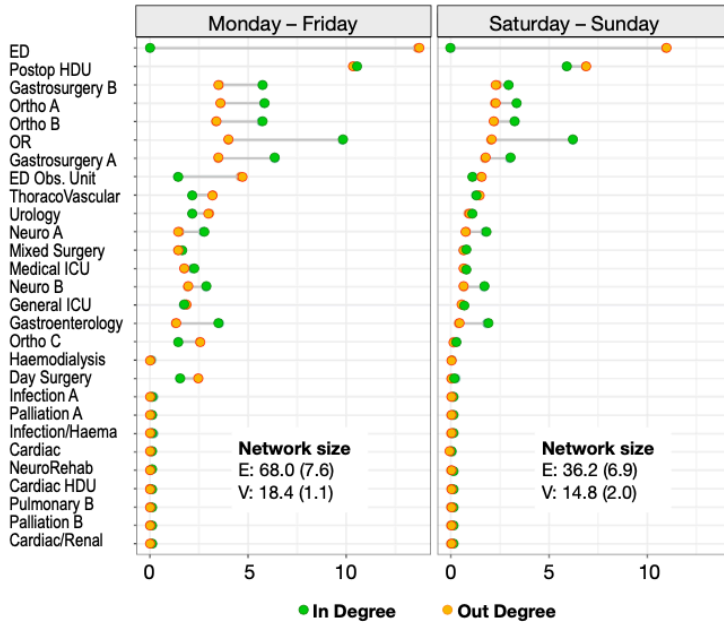


Figure 3 Transfer network connectivity on weekdays and weekends



Supplementary Tables

Heterogeneity of intrahospital patient transfer patterns – network characteristics, temporal changes and patient-level risk factors

by Chi Zhang, Torsten Eken, Silje Bakken Jørgensen, Magne Thoresen, Signe Søvik

Table S1 Hospital ward names, types and abbreviations

Type	Ward	Abbreviations used in Fig. 1	Department link for “home” wards in current study
ED	Emergency Dept	ED	
OR	Operating Room Suite Day Surgery Unit	OR Day S	
Technical	Postoperative HDU General ICU Medical ICU Cardiac HDU ED Observation Unit Haemodialysis	PHDU ICU MICU CICU EDOU Haemodial	
Surgical	Orthopaedic A Orthopaedic B Orthopaedic C	OT A OT B OT C	Orthopaedics Department
	Gastrosurgery A Gastrosurgery B	GIS A GIS B	Gastrosurgery Department
	Mixed Surgery Urology ThoracoVascular	Mix S Uro TCVS	
Medical	Neurology A Neurology B Neuro Rehab	NR A NR B NR Rehab	Neurology Department
	Gastroenterology	GE	Gastroenterology Department
	Palliation A Geriatrics Palliation B Infection/Haema Infection A Cardiac Cardiac/Renal Pulmonary A Pulmonary B	PC A Geriatrics PC B Inf/Haem Inf A Cardiac Card/Ren Pulm A Pulm B	

Table S2 Poisson regression analysis on number of intrahospital transfers, with modelled interactions between departments, OR stay, and antibiotics use

Risk factors		Model S1			Model S2		
		RR	95% CI	P value	RR	95% CI	P value
Age	18 – 39	Reference			Reference		
	40 – 64	1.016	0.980 – 1.055	0.386	1.020	0.983 – 1.058	0.304
	65 – 84	0.980	0.945 – 1.017	0.284	0.983	0.948 – 1.020	0.361
	85+	0.959	0.910 – 1.010	0.115	0.961	0.911 – 1.012	0.132
NEWS2 score ¹	0 – 2	Reference			Reference		
	3 – 4	0.986	0.945 – 1.029	0.523	0.989	0.948 – 1.032	0.626
	5 – 6	0.963	0.889 – 1.042	0.352	0.970	0.895 – 1.050	0.456
	7+	0.821	0.716 – 0.937	0.004	0.829	0.722 – 0.947	0.006
Gender	Female	Reference			Reference		
	Male	0.998	0.974 – 1.023	0.876	0.998	0.947 – 1.022	0.854
Department	Gastroenterology	Reference			Reference		
	Gastrosurgery	1.194	1.123 – 1.271	<0.001	1.182	1.100 – 1.271	<0.001
	Neurology	1.094	1.029 – 1.163	0.004	1.091	1.019 – 1.169	0.013
	Orthopaedics	1.179	1.101 – 1.263	<0.001	1.163	1.076 – 1.257	<0.001
Admission	Elective	Reference			Reference		
	Non-elective	1.844	1.788 – 1.902	<0.001	1.845	1.789 – 1.904	<0.001
Antibiotics ²	No	Reference			Reference		
	Yes	1.102	1.072 – 1.132	<0.001	1.073	0.947 – 1.213	0.264
Been to OR ³	No	Reference			Reference		
	Yes	2.307	1.953 – 2.707	<0.001	2.242	1.717 – 2.871	<0.001
Been to HDU ⁴	No	Reference			Reference		
	Yes	2.118	2.037 – 2.202	<0.001	2.118	2.037 – 2.201	<0.001
Interactions	GS * OR	1.240	1.051 – 1.472	0.012	1.319	1.024 – 1.731	0.039
	NR * OR	0.925	0.713 – 1.195	0.556	0.952	0.675 – 1.346	0.778
	OT * OR	1.380	1.167 – 1.640	<0.001	1.411	1.094 – 1.853	0.013
	GS * Antibiotics				1.038	0.903 – 1.194	0.603
	NR * Antibiotics				0.998	0.856 – 1.165	0.982
	OT * Antibiotics				1.079	0.915 – 1.274	0.403
	Antibiotics * OR				1.061	0.758 – 1.495	0.799

RR, patient transfer rate ratio. ¹Mean first 48-h NEWS2 score. ²Use of any non-prophylactic antibiotics. ³Indicates a surgical procedure. ⁴Stayed in an ICU or HDU, indicates a severe patient condition. GS, Gastrosurgery; NR, Neurology; OT, Orthopaedics.