University of California

Los Angeles

Discovery and Clinical Decision Support for Personalized Healthcare

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Science in Electrical Engineering

by

Jinsung Yoon

2016

Abstract of the Thesis

Discovery and Clinical Decision Support for Personalized Healthcare

by

Jinsung Yoon

Master of Science in Electrical Engineering

University of California, Los Angeles, 2016

Professor Mihaela van der Schaar, Chair

With the advent of electronic health records, more data is continuously collected for individual patients and more data is available for review from past patients. Despite this, it has not yet been possible to successfully use this data to systematically build clinical decision support systems that can produce personalized clinical recommendations to assist clinicians in providing individualized healthcare. In this paper, we present a novel approach, Discovery Engine (DE) that discovers which patient characteristics are most relevant for predicting the correct diagnosis and/or recommending the best treatment regimen for each patient. We demonstrate the performance of DE in two clinical settings: diagnosis of breast cancer as well as personalized recommendation for a specific chemotherapy regimen for breast cancer patients. For each distinct clinical recommendation, different patient features are relevant; DE can discover these different relevant features and use them to recommend personalized clinical decisions. The DE approach achieves a 16.6% improvement over existing state-of-the-art recommendation algorithms in terms of kappa coefficients for recommending the personalized chemotherapy regimens. For diagnostic predictions, the DE approach achieves a 2.18% and 4.20% improvement over existing state-of-the-art prediction algorithms in terms of pre-

diction error rate and false positive rate, respectively. We also demonstrate that the performance of our approach is robust against missing information, and that the relevant features discovered by DE are confirmed by clinical references.

The thesis of Jinsung Yoon is approved.

William Hsu

Kung Yao

Mihaela van der Schaar, Committee Chair

University of California, Los Angeles

2016

To my family, colleagues and love . . .

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am grateful to Camelia Davtyan, Ahmed M. Alaa, David Davtyan, Sung Wook Lee and Mihaela van der Schaar and all my Decisionlab members for their help and support which have significantly improved my thesis.

CHAPTER 1

Introduction

Clinicians are routinely faced with the practical challenge of integrating high-dimensional clinical data in order to recommend the most appropriate clinical decision for a given patient [Ric08]. As the understanding of complex diseases progresses, the types of available tests and treatments diversify and, as a result, the difficulty of recommending the optimal clinical decision for a particular patient increases as well. Current clinical decisions continue to rely on clinical practice guidelines which, in cases where scientific analysis and evidence is scarce, are largely based on clinical experience and opinion. Also, current clinical practice guidelines are aimed at a "representative" patient rather than an individual patient who may display other relevant characteristics. Such "representative" guidelines may thus miss the opportunity to consider personal traits when recommending clinical decisions [CJK12, RHW98]. For example, the American Cancer Society (ACS) recently issued new guidelines which suggested that women with an average risk of breast cancer should start having mammograms at an age of 45 (five years later than ACS had previously recommended) [Soc15]. However, women who have certain risk factors (family history of breast cancer, no children, etc.) have a higher risk of developing breast cancer, and they would benefit from having mammograms at an earlier age. In cases such as these, the ACS guidelines recommend that a high risk patient consults with her physician to determine an appropriate screening age and interval, which is based on that particular physician's experience and opinion. Moreover, statistics show that diagnostic errors

result in 10% of patient deaths and represent the most frequent type of medical malpractice claims in the United States [Fre15]. This reality highly underscores the urgent need for building smart clinical decision support systems (CDSS) and diagnosis decision support systems (DDSS) that can assist clinicians in making accurate, personalized clinical recommendations [Ber07]. It has been recently recognized [AFN15] that medical informatics tools and machine learning techniques can be successfully used to provide recommendations for personalized diagnosis and treatment.

The goal of this paper is to develop methods that will enable CDSS to personalize their recommendations based on individual patient characteristics. The wealth of information being routinely collected as part of the electronic health record (EHR) provides an unprecedented opportunity to discover appropriate clinical recommendations for patients given historical information about the clinical decisions administered to similar patients and their actual outcomes [SC13]. However, using this information is difficult precisely because there is so much of it. The solution is to extract only the relevant information for the particular patient and the relevant clinical decisions previously used for similar patients among the wealth of available information. Extracting only the relevant information is important because using irrelevant features can significantly hurt the performance of the system, unnecessarily increase its complexity, and decrease its learning/adaptation speed [BL97]. Furthermore, efficient discovery of relevant patient features can help clinicians focus on the relevant information available about the patient without having to sift through a large patient record.

1.1 Paper Contribution

In this paper, we present a novel approach called Discovery Engine which optimizes clinical recommendations by identifying the features in the patient record

that differentiate the individuals who receive a certain clinical decision and respond positively from those who do not. Our approach utilizes the available contextual information about patients and learns from the large quantities of observational clinical data to inform clinical recommendations and make better decisions by learning from similar patients. We show that our DE approach consistently outperforms existing state-of-the-art machine learning algorithms both in terms of matching individual patients to the optimal treatment regimens as well as diagnosis accuracy.

One of the biggest challenges faced by this class of recommendation systems is that the rewards/actual outcomes of clinical decisions (e.g. five-year recurrence free survival) are usually not available [BH13, LDC12]. Moreover, even if rewards/actual outcomes of the clinical decisions were available, the counterfactuals rewards/actual outcomes of alternative clinical decisions that were not used, are never available [LDC12]. What is available, however, is a large medical literature that reports the results of a wide range of clinical studies, including different types of patients, different patient characteristics, different types of clinical decisions, and the actual outcomes of these decisions. We use the results of these studies to construct transfer rewards, which we use as proxies for rewards. This allows us to train the DE algorithms as well as to evaluate their performance in comparison to existing methods when the actual outcomes cannot be achieved. The four primary contributions of this paper are as follows:

- We describe a novel approach for discovering the most relevant information from the EHR that distinguishes between patients that should receive one particular clinical decision and the patients who should receive another. For instance, premenopausal breast cancer patients are more likely to respond to a specific type of chemotherapy such as CEF [Lev01].

- Using the past records in the EHR and external knowledge from the medical

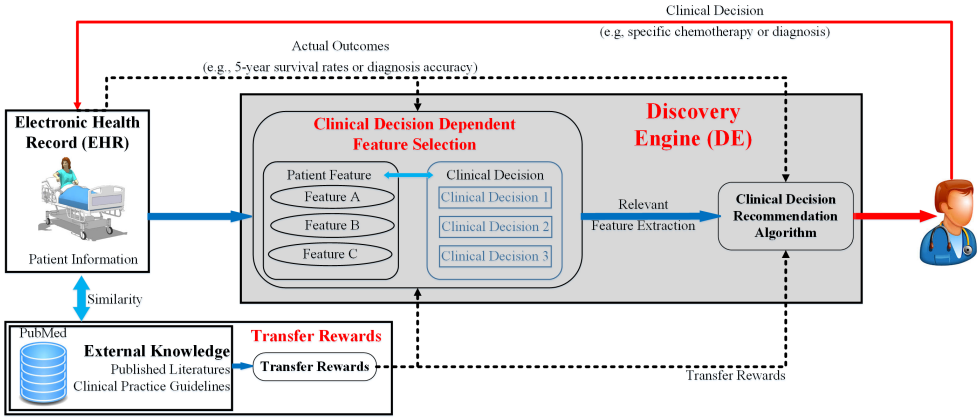

Figure 1.1: Personalized clinical decision support system using discovery engine (DE)

literature, our approach discovers the optimal personalized clinical decision based on the discovered relevant information (e.g. their clinical test results, treatment history, and outcomes).

- In lieu of having actual reward values associated with clinical decisions, we define the transfer rewards, a method for estimating actual outcomes described in external knowledge (published literature and clinical practice guidelines) based on their similarities to individual patients given reported characteristics.

- We apply DE to two medical applications: 1. Personalized treatment recommendations (chemotherapy regimens) for breast cancer patients and 2. Diagnosis of breast cancer. DE is used to discover which features are relevant to make a distinct clinical decision and then uses this knowledge to build a clinical decision recommendation system. We evaluate the performance of DE in the context of breast-cancer diagnosis and treatment, and show that it consistently and significantly outperforms state-of-the-art machine learning algorithms.

1.2 Related Work

1.2.1 Personalized Clinical Decision Support Systems

Current medical practice relies on manually curated systematic reviews of the available scientific evidence and clinical guidelines that provide recommendations for large groups of patients rather than personalized recommendations that are tailored to individual patients. Clinical decision support systems have been proposed before, but many of them do not consider the specific characteristics of patients and do not provide personalized clinical recommendations; hence, they are not very accurate and have only limited applicability in practice [Tsu98, XST14, CJK12, HC10, MPM15, RHW98]. Moreover, clinicians often refer to the medical literature available through Medline/PubMed, VisualDX, and UpToDate to help them associate observed finding with possible conditions and recommended decisions. However, these resources are also not customized to a specific patient's case.

Several CDSSs are currently implemented and used in clinical settings. For instance, WizOrder was developed to help reduce medical errors and support clinical decisions when entering orders [MWC05]. This system is now used in several neonatal intensive care units at Vanderbilt University Hospital. Assessment and Treatment of Hypertension: Evidence-based Automation (ATHENA) is a CDSS used to manage hypertension in primary care [GCT04]. It was developed to recommend drug therapies and assess and control their effect on blood pressure. ATHENA is now used in several clinics in Northern California. TherapyEdge-HIV is a web-based real time alerting system for the treatment of HIV which is used in over 42 clinical sites [BGS04]. While several CDSS are now implemented in clinical settings, their evaluations are often mixed. Some papers argue that CDSS significantly improve the effectiveness of clinical practice [KHB05, JSV11], while others conclude that CDSS do not affect mortality and may even moder-

ately increase morbidity outcomes [MKL14]. Furthermore, the cost-effectiveness of CDSS has yet to be demonstrated [BCP11].

The advent of Big Data has been identified as an opportunity to improve the performance of existing CDSS as well as catalyze the development of new CDSS [VHH15]. A large literature has used data-driven approaches to develop representative rather than personalized healthcare decision support systems [RZW15]. A smaller literature is dedicated to developing personalized CDSS. However, most existing papers in this strand of literature [DJ12, KTM10, DSB15, BDB11, SM06, MSA14, MWL12] only either just discuss opportunities rather than propose a concrete algorithm or they apply off-the-shelf machine learning techniques to the considered medical problem and do not address the unique characteristics and challenges of developing personalized CDSS. Diagnosis decision support systems have been developed for cardiovascular diseases and diabetes using ensemble learning [EKZ08], SVMs [CAT07], artificial neural networks [ZN15] or rule-based algorithms [Ano12]. Although some diagnosis decision support systems issue accurate diagnostic recommendations for specific diseases, most of them are based on a small number of manually selected features [PG07a, PG07b]. Whenever the number of features (contexts) is large, these methods fail to perform well [HMN05].

Most importantly, most of the proposed CDSS solely focus on diagnosis recommendations and do not provide solutions for the equally important problem of treatment recommendations. A small number of studies attempt to propose CDSS for treatment recommendations [KSE12]. However, these CDSS differ significantly from our DE. For instance, [TVD14] proposes antibiotic recommendation systems for representative patients but does not use a data-driven approach and [ZWH14] proposes a CDSS for personalized medicine recommendations, but which uses similarity information among drugs and patients that is specific to the study at hand and cannot be easily applied to other diseases - the similarity between patients is solely based on the ICD 9 codes and the similarity between drugs is based on

their chemical structure. Hence, the methods in [TVD14, ZWH14] are not widely applicable to a diverse set of patients and treatments.

In contrast, although our DE was only verified in the context of breast cancer so far, it is designed to operate in a variety of complex diseases such as breast cancer, lung cancer, prostate cancer etc. Furthermore, as we will show in our experimental section, it can perform well even when the number of features used to make a decision is large because it adopts a novel method to discover the most relevant features to consider when deciding on certain diagnosis or treatment options for a specific patient. For this, we developed a customized feature selection and decision making system which significantly outperforms existing off-the-shelf techniques. Importantly, based on the authors' knowledge, DE likely is the only one that considers features specific to a given set of clinical treatments and optimized for a specific patient, that are indicative of treatment success in a complex disease such as breast cancer.

Our work is also related to other works in the field of medical informatics dedicated to improving breast cancer diagnosis and treatment. However, our work is distinct from prior works, that they only analyzed the impact of specific patient features (such as genetic information or imaging information extracted from mammograms or other imaging) to improve the performance of CDSS [WLP14,KDE15,LPP14]. Our DE system is based on a novel set of machine learning methods developed especially for personalized diagnosis and treatment discovery which are shown in later sections to significantly outperform existing methods. Moreover, DE is not only applicable to breast cancer, but can be generally applied to discover personalized treatment for other complex diseases.

1.2.2 Relevant Feature Selection

Another strand of literature related to this work is relevant feature selection algorithms including Correlation Feature Selection (CFS) and Mutual Information Feature Selection (MIFS) [Hal99, PLD05]. These are related to our feature selection approach. However, our clinical decision dependent feature selection algorithm (CDFS) is very different from existing feature selection algorithms which focus on the patients' characteristics and not on how these characteristics distinctively impact different clinical decisions. Our approach is capable of discovering which different features are relevant to each different clinical decision. This makes CDFS similar to our prior work [TV14, TS15], the RELEAF algorithm. However, unlike RELEAF, which is very slow because it must compare all possible combinations of features, CDFS is able to discover the relevant features in a very fast and efficient manner because it adopts a sequential feature selection approach. This sequential approach significantly reduces the sample and computational complexity of the RELEAF algorithm.

1.2.3 Machine Learning Techniques

Our method also exhibits similarities to the contextual multi-armed bandit problem (MAB). However, contextual MABs are very inefficient when the number of contexts (in our case patient features) is large (see the experimental section of [TS15]). DE is able to successfully deal with the curse of dimensionality by discovering what information is relevant and making clinical decisions based only on relevant contexts rather than the entire set of contexts that can be extracted from the EHR, much of which is irrelevant to the decisions of whether to administer a specific treatment or not.

In Chapter 2, we present the detailed problem formulation under consideration. Chapters 3 and 4 present the details of the proposed algorithms for discovering rel-

evant features, recommending the optimal clinical decision and transfer learning, respectively. Chapter 5 and 6 presents the simulation results for the breast cancer patients for evaluating our proposed system against state-of-the-art machine learning algorithms and feature selection algortihms and Chapter 7 discusses the results and future works of the paper.

CHAPTER 2

System Model

In this chapter, we introduce our method - Discovery Engine (DE). The DE discovers/learns which features/characteristics of a patient are most informative in predicting the success of a clinical decision. For instance, the tumor grade may be found to be relevant for predicting the success of a certain type of chemotherapy in a patient, but not the success of another type of chemotherapy. Thus, different features may be discovered to be relevant for different decisions. Then, when a clinician requests the recommendation from DE for a specific patient, DE decides the best clinical recommendation for the patient which has the best estimated outcomes. The outcome of a decision is estimated based on the values of the relevant features (i.e. the features found to be relevant for that decision) of the patient. For instance, if the tumor grade was found to be relevant for a certain chemotherapy that chemotherapy will or will not be recommended to that patient depending on that patient's tumor grade.

Fig. 1.1 depicts the proposed system, which issues a personalized clinical recommendation to the physician about certain patients. The outcomes of certain clinical decisions are used as a reward to train the DE. While the proposed system is applicable in general, we illustrate its use in the context of breast cancer. (A nomenclature table summarizing the variables used and their definitions can be found in Appendix.)

Let $\mathbf{x} = \{x_1, x_2, ..., x_D\}$ denote the patient information where D is the total number of patient features such as age, tumor size, estrogen receptor information

etc.; $a \in A \triangleq \{a_1, a_2, ..., a_K\}$ denotes the clinical decision (e.g. chemotherapy regimens or breast cancer diagnosis) that is recommended to the patient. Each patient feature is denoted as $f \in F \triangleq \{f_1, f_2, ..., f_D\}$. The reward y is derived based on the actual patient outcomes (e.g. five-year survival rates or recurrence rates). Let $\mathbf{x}(n), a(n), y(n)$ be the patient information, clinical decision and reward of n-th patient and $\mathcal{H}_N = \{\mathbf{x}(n), a(n), y(n)\}_{n=1}^{N}$ be the information available for the N previously seen patients. This represents the training set.

The outcomes of a clinical decision a do not depend on all the features [PDK13]: we assume that the outcomes of a clinical decision a depend only on a subset of features $\mathcal{R}(a) \subseteq F$ which we call the relevant features. Let $\mathcal{R} = \bigcup_{a \in A} \mathcal{R}(a)$ be the set of all relevant features. The DE approach is capable of discovering different features that are relevant to different clinical decisions. We say that $\mathcal{R}(a)$ is relevant/informative for clinical decision a if the expected reward only depends on the information contained in $\mathcal{R}(a)$.

Our goal is to discover the relevant features of each clinical decision a (this may be different for each decision) and recommend the optimal clinical decision that corresponds to the discovered relevant patient information. The optimal recommended clinical decision is given by

$$a^*(\mathbf{x}_R) \triangleq \text{argmax}_a \mathbb{E}_{y|a,\mathbf{x}_{R(a)}}(y|a, \mathbf{x}_{R(a)}) \tag{2.1}$$

where $\mathbb{E}()$ is the expectation of the random variable. Therefore, $a^*(\mathbf{x}_R)$ is defined as the clinical decision that yields the best expected patient outcome for a patient characterized by the relevant features $\mathbf{x}_R$.

CHAPTER 3

Algorithms

Discovery engine (DE) consists of two algorithms: a clinical decision dependent feature selection algorithm (CDFS) and a clinical decision recommendation algorithm. As it can be seen in Fig. 1.1, DE discovers different relevant features for different clinical decisions using CDFS. The detailed steps of each algorithm are described in the following subsections and the pseudo-code of DE can be found in the Appendix.

3.1 Clinical Decision Dependent Feature Selection (CDFS)

To describe CDFS, we start by introducing a few notations. Let $\hat{y}_a$ and N_a be the sample mean rewards estimate and the number of patients who received the clinical decision a, respectively. Similarly let $\hat{y}_a^S(\mathbf{x}_S)$ and $N_a^S(\mathbf{x}_S)$ be the sample mean rewards estimate and the number of patients (whose feature information contains $\mathbf{x}_S$ and was provided clinical decision a), respectively. We formalize these variables as:

$$N_a = \sum_n \mathbb{I}\{a(n) = a\} \tag{3.1}$$

$$\hat{y}_a = \frac{1}{N_a} \sum_n \mathbb{I}\{a(n) = a\} \times y(n), \tag{3.2}$$

$$N_a^S(\mathbf{x}_S) = \sum_n \mathbb{I}\{\mathbf{x}_S \subset \mathbf{x}(n)\}\mathbb{I}\{a(n) = a\}, \tag{3.3}$$

$$\hat{y}_a^S(\mathbf{x}_S) = \frac{1}{(N_a^S(\mathbf{x}_s))} \sum_n \mathbb{I}\{\mathbf{x}_S \subset \mathbf{x}(n)\}\mathbb{I}\{a(n) = a\} \times y(n), \tag{3.4}$$

where $\mathbb{I}\{\}$ is the indicator function. We define the relevance metric $h_f^r(a)$ as the variance of the rewards for a certain action a if a given feature x_f is considered, when selecting the action: $|\hat{y}_a^f(x_f) - \hat{y}_a|$. This is weighted by $\frac{N_a^f(x_f)}{N_a}$ which represents the frequency with which feature x_f is present when action a is selected. We formalize this as:

$$h_f^r(a) \triangleq \sum_{x_f} \frac{N_a^f(x_f)}{N_a} |\hat{y}_a^f(x_f) - \hat{y}_a| \tag{3.5}$$

In addition, we define a redundancy metric $h_{f,s}^d(a)$ which measures how the expected reward made for a given patient is affected by considering an additional feature x_s when clinical decision a is recommended. We formalize this as:

$$h_{f,s}^d(a) = -\sum_{x_f, x_s} \frac{N_a^{f,s}(x_f, x_s)}{N_a} [\hat{y}_a^{f,s}(x_f, x_s) - \hat{y}_a^s(x_s)] \tag{3.6}$$

Then, we define $U_f(a)$ as the utility obtained if feature x_f is additionally selected as a relevant feature for clinical decision a. Let $\hat{\mathcal{R}}(a)$ be defined as the previously discovered relevant features set for clinical decision a. Then, the utility function $U_f(a)$ is defined as:

$$U_f(a) = h_f^r(a) - \frac{1}{|\hat{\mathcal{R}}(a)|} \sum_{s \epsilon \hat{\mathcal{R}}(a)} h_{f,s}^d(a) \tag{3.7}$$

where $1/|\hat{\mathcal{R}}(a)|$ is used as a normalization factor. The main steps of the CDFS are outlined below:

Step 1: For each clinical decision a, initialize $\hat{\mathcal{R}}(a)$ as the empty set (i.e. $\varnothing$) and its complementary set $(\hat{\mathcal{R}}^c(a))$ as the set of all features (i.e.F).

Step 2: The algorithm selects the first relevant feature which maximizes the

relevance metric $(h_f^r(a))$, i.e.,

$$G = \text{argmax}_{f \in \hat{\mathcal{R}}^c(a)} h_f^r(a)$$
$$\hat{\mathcal{R}}(a) \leftarrow \hat{\mathcal{R}}(a) \cup G$$

Step 3: The algorithm finds the subsequent relevant feature that maximizes utility function $(U_f(a))$, i.e.,

$$H = \text{argmax}_{f \in \hat{\mathcal{R}}^c(a)} U_f(a)$$
$$\hat{\mathcal{R}}(a) \leftarrow \hat{\mathcal{R}}(a) \cup H$$

Step 4: The algorithm iteratively runs Step 3 until the utility function $U_f(a)$ is less than threshold cost C, where C is an input parameter for the algorithm which can adjust the number of relevant features for handling the trade-off between the speed of convergence and the recommendation accuracy, i.e.,

$$\text{If } \max_{f \in \hat{\mathcal{R}}^c(a)} U_f(a) < C,$$
$$\text{then, } \hat{\mathcal{R}}(a) = \hat{\mathcal{R}}(a)$$

Discovered relevant features are used to recommend the optimal clinical decision in clinical decision recommendation algorithm.

3.2 Clinical Decision Recommendation Algorithm

The proposed clinical decision recommendation algorithm recommends the optimal clinical decision which maximizes the estimated patient outcome only based on the relevant contexts (features) discovered by CDFS. The main steps of the recommendation algorithm are outlined below:

Step 1: Find the set of unresolved clinical decisions (U) for the patient with

information vector $\mathbf{x}_{\hat{\mathcal{R}}(a)}$:

$$U = \{a \in A | N_a^{\hat{\mathcal{R}}(a)}(\mathbf{x}_{\hat{\mathcal{R}}(a)}) < C_{TH} \times \log(n)\} \tag{3.8}$$

where $C_{TH} \times \log(n)$ is a control function. C_{TH} is an input parameter which can adjust the trade-off between the confidence of the clinical recommendations and the learning speed of DE. If a set of unresolved clinical decisions exists ($U \neq \varnothing$), DE abstains from making clinical decision recommendations and only updates $N_a^{\hat{\mathcal{R}}(a)}(\mathbf{x}_{\hat{\mathcal{R}}(a)})$ and $\hat{y}_a^{\hat{\mathcal{R}}(a)}(\mathbf{x}_{\hat{\mathcal{R}}(a)})$ based on the obtained rewards. In other words, DE only issues recommendations when it is sufficiently confident about its clinical recommendations and it abstains otherwise.

Step 2: If there is no unresolved clinical decisions (i.e., $U = \varnothing$) for the patient with information vector $\mathbf{x}_{\hat{\mathcal{R}}(a)}$, the optimal clinical decision with respect to the relevant feature set $\hat{\mathcal{R}}(a)$ is determined as

$$\hat{a}(\mathbf{x}) = \text{argmax}_a \hat{y}_a^{\hat{\mathcal{R}}(a)}(\mathbf{x}_{\hat{\mathcal{R}}(a)}) \tag{3.9}$$

This optimization selects the clinical decision with the maximum estimated reward for the patient with relevant information vector $\mathbf{x}_{\hat{\mathcal{R}}(a)}$.

After the rewards of the recommended clinical decision are obtained, $N_{\hat{a}}^{\hat{\mathcal{R}}(a)}(\mathbf{x}_{\hat{\mathcal{R}}(\hat{a})})$ and $\hat{y}_{\hat{a}}^{\hat{\mathcal{R}}(a)}(\mathbf{x}_{\hat{\mathcal{R}}(\hat{a})})$ are updated as follows:

$$N_{\hat{a}}^{\hat{\mathcal{R}}(a)}(\mathbf{x}_{\hat{\mathcal{R}}(a)}) = \sum_n \mathbb{I}\{\mathbf{x}_{\hat{\mathcal{R}}(a)} \subset \mathbf{x}(n)\} \times \mathbb{I}\{a(n) = \hat{a}\} \tag{3.10}$$

$$\hat{y}_{\hat{a}}^{\hat{\mathcal{R}}(a)}(\mathbf{x}_{\hat{\mathcal{R}}(a)}) = \frac{y(n)}{N_a^{\hat{\mathcal{R}}(a)}(\mathbf{x}_s)} \sum_n \mathbb{I}\{\mathbf{x}_{\hat{\mathcal{R}}(a)} \subset \mathbf{x}(n)\} \times \mathbb{I}\{a(n) = \hat{a}\} \tag{3.11}$$

The computational complexity of DE is $O(ND^2)$; hence, DE has a relatively low run-time complexity with high dimensional datasets.

3.3 DE with Missing Information

Electronic health records, more often than not, may have missing information for some patients [LDC12]; hence, DE must be able to operate properly even with missing information.

Suppose that the dataset contains missing information. We can divide the feature information vector $\mathbf{x}$ into two components: the available features ($\mathbf{x}^{av}$) and the missing features ($\mathbf{x}^{m}$). Thus, $\mathbf{x} = \{\mathbf{x}^{av}, \mathbf{x}^{m}\}$. First, the relevance metric of CDFS is solely computed based on the available information:

$$h_f(a) \triangleq \sum_{x_f^{av}} \frac{N_a^f(x_f^{av})}{N_a} |\hat{y}_a^f(x_f^{av}) - \hat{y}_a|$$

Therefore, if the feature f is frequently missing, $h_f(a)$ decreases, and as a result the feature f is rarely selected as a relevant feature.
Second, it should also be noted that we can estimate the reward ($\hat{y}_a^{\hat{\mathcal{R}}(a)}(\mathbf{x}^{av}_{\hat{\mathcal{R}}(a)})$) with missing information based on a given patient's available relevant information, $\mathbf{x}^{av}_{\hat{\mathcal{R}}(a)}$, for each clinical decision a. More specifically, it can be estimated as:

$$\begin{aligned}
\hat{y}_a^{\hat{\mathcal{R}}(a)}(\mathbf{x}^{av}_{\hat{\mathcal{R}}(a)}) &= \mathbb{E}(\hat{y}_a^{\hat{\mathcal{R}}(a)}(\mathbf{x}^{av}_{\hat{\mathcal{R}}(a)}, \mathbf{x}^{m}_{\hat{\mathcal{R}}(a)}) | \mathbf{x}^{av}_{\hat{\mathcal{R}}(a)}) \\
&= \sum_{\mathbf{x}^{m}_{\hat{\mathcal{R}}(a)}} \hat{y}_a^{\hat{\mathcal{R}}(a)}(\mathbf{x}^{av}_{\hat{\mathcal{R}}(a)}, \mathbf{x}^{m}_{\hat{\mathcal{R}}(a)}) \times P(\mathbf{x}^{m}_{\hat{\mathcal{R}}(a)} | \mathbf{x}^{av}_{\hat{\mathcal{R}}(a)}) \\
&= \sum_{\mathbf{x}^{m}_{\hat{\mathcal{R}}(a)}} \hat{y}_a^{\hat{\mathcal{R}}(a)}(\mathbf{x}_{\hat{\mathcal{R}}(a)}) \times P(\mathbf{x}^{m}_{\hat{\mathcal{R}}(a)} | \mathbf{x}^{av}_{\hat{\mathcal{R}}(a)})
\end{aligned}$$

We can estimate the conditional probability, $P(\mathbf{x}^{m}_{\hat{\mathcal{R}}(a)} | \mathbf{x}^{av}_{\hat{\mathcal{R}}(a)})$, based on the probability distribution of the features in the training set. Based on this estimation rule, we can robustly identify the optimal clinical decision even if there is missing information.

CHAPTER 4

Transfer Rewards

As discussed in the introduction chapter, the most valuable rewards for most clinical decision support systems are, in theory, the actual patient outcomes (e.g. 5 year survival rates or recurrence rates in the case of breast cancer). However, these outcomes are very difficult to obtain in practice [LDC12]. Instead, we use a proxy for outcomes based on external knowledge which consists of published literature and clinical practice guidelines. We refer to all external knowledge simply as references in the remainder of the paper.

The idea is to match patients to appropriate relevant references. For each patient and each reference, we define the term similarity as the amount of information that reference provides about that patient. Similarity is computed by calculating the posterior probability of that patient feature belonging to the population demography from the reference. Then we aggregate the actual outcomes of certain clinical decisions for each reference according to the similarity (posterior probability) and use this as a transfer reward for that clinical decision when applied to that patient. The system model for transfer reward estimation is illustrated in Fig. 4.1.

To compute the transfer rewards, we first estimate the similarity between a patient and a reference. The first step of estimating this similarity is to find the relevant patient features for each reference; we do this using a sequential feature selection algorithm based on the mutual information in order to deal with population demography [PLD05]. The mutual information between the i-th reference

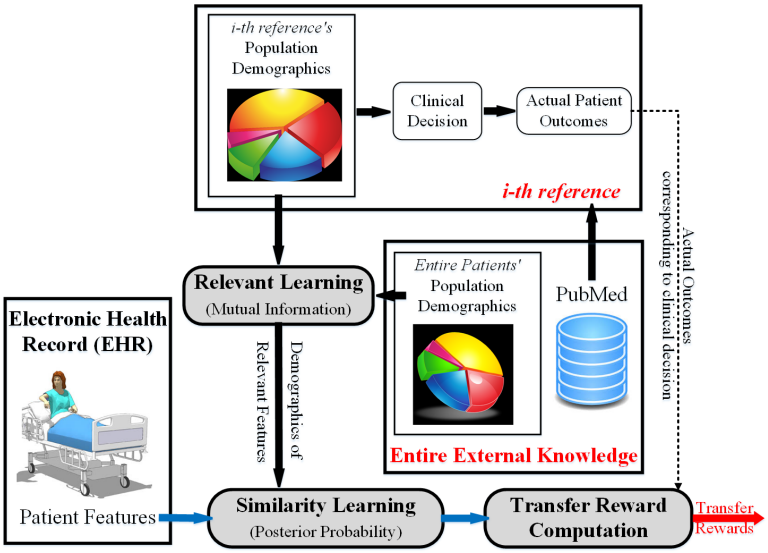

Figure 4.1: System model of transfer reward estimation

(E_i) and the k-th feature (f_k) is defined as:

$$I(f_k; E_i) = \sum_{x \in \chi_k} P(x|E_i) \times \log \frac{P(x|E_i)}{P(x)} \tag{4.1}$$

where $P(x|E_i)$ is the probability of feature x in reference i, $P(x)$ is the probability of feature x across the entire set of references, and χ_k is context space of f_k. Let $\hat{\mathcal{R}}(E_i)$ be defined as the discovered relevant feature set for reference E_i and the utility function $U_f(E_i)$ is determined as:

$$U_f(E_i) = I(x_f; E_i) - \frac{1}{|\hat{\mathcal{R}}(E_i)|} \sum_{s \in \hat{\mathcal{R}}(a)} I(x_f|E_i; x_s|E_i) \tag{4.2}$$

where $\frac{1}{|\hat{\mathcal{R}}(E_i)|}$ is used as a normalization factor. This utility function measures an increment of mutual information between relevant feature set and the reference when feature f is the additionally selected as relevant feature.

The algorithm selects the first relevant feature which maximizes the mutual

information with E_i, i.e.

$$G = \text{argmax}_{f \epsilon \hat{\mathcal{R}}^c(E_i)} I(x_f; E_i)$$
$$\hat{\mathcal{R}}(E_i) \leftarrow \hat{\mathcal{R}}(E_i) \cup G$$

Then, the algorithm finds the subsequent relevant feature that maximizes utility function $U_f(E_i)$, i.e.,

$$H = \text{argmax}_{f \epsilon \hat{\mathcal{R}}^c(E_i)} U_f(E_i)$$
$$\hat{\mathcal{R}}(E_i) \leftarrow \hat{\mathcal{R}}(E_i) \cup H$$

The algorithm iteratively adds new relevant features in $\hat{\mathcal{R}}(E_i)$ until the maximum utility function $U_f(a)$ becomes less than zero.

The second step is to compute a posterior probability of a patient feature set belonging to the population demography from the reference. Given the n-th patient, characterized by the feature vector $\mathbf{x}_n = \{x_1(n), ...x_D(n)\}$, we compute the posterior probability that the given patient belongs to the population demography of the reference; we express this value as $P(E_i|X_1 = x_1(n), ..., X_D = x_D(n))$. We compute this via Bayes rule; it is computationally convenient to take logarithms:

$$\begin{aligned}
&\log(P(E_i|X_1 = x_1(n), ..., X_D = x_D(n)) \\
&= \log \frac{P(X_1 = x_1(n), ...X_D = x_D(n)|E_i) \times P(E_i)}{P(X_1 = x_1(n), ..., X_D = x_D(n))} \\
&\approx \log P(E_i) + \sum_{l \epsilon \hat{\mathcal{R}}(E_i)} \log \frac{P(X_l = x_l(n)|E_i)}{P(X_l = x_l(n))}
\end{aligned}$$

where we write $P(E_i)$ as the probability of selecting the i-th reference as the best clinical decision for the entire population. We define this approximation of posterior probability as a similarity between n-th patient and i-th reference ($Sim_{E_i}(x_n)$). Second, we compute the estimated transfer reward of each clini-

cal decision a for n-th patient as a weighted sum of actual outcomes of clinical decisions in each reference according to the similarity ($Sim_{E_i}(x)$). i.e.,

$$\hat{Sim}_{a|E_i}(x) = \frac{Sim_{E_i}(x)}{\sum_{i:a\in A(E_i)} Sim_{E_i}(x)}$$

$$tr_a(x) = \sum_{i:a\in A(E_i)} \hat{Sim}_{a|E_i}(x) \times r(a|E_i)$$

where $r(a|E_i)$ is an actual patient outcome for clinical decision a in i-th reference, ($\hat{Sim}_{a|E_i}(x)$) is a normalized similarity for clinical decision a, and $A(E_i)$ is the set of clinical decisions considered in i-th reference. We define this estimated reward as the transfer reward ($tr_a(x)$). These provide a complete ranking of each clinical decision for each patient. The clinical decision with the highest transfer reward is the optimal clinical decision for the given patient. The pseudo-code for estimating transfer rewards is given in Appendix.

CHAPTER 5

Experiment1: Chemotherapy Recommendation for Breast Cancer Patients

While our DE algorithm can be applicable in general clinical decision support systems, in this chapter, we illustrate its use in the context of a personalized recommendation system of chemotherapy regimens for breast cancer patients in this section. Fig. 5.1. illustrates the system model of this application.

5.1 Data Description

From an initial set of 2,353 references (performing a narrow search of breast cancer chemotherapy regimens using PubMed Clinical Queries), 32 references were selected for further analysis. (The complete list of the 32 references is provided in Appendix.) The list was compiled based on the following two criteria: (1) References contain the clinical outcomes for at least one of the 6 standard chemotherapy regimens for breast cancer patients [BL97,BLS14,FGR15]; (2) References contain the demographic information of the breast cancer patients enrolled in the randomized trials. The sample size of reported references ranged from 50 to 3,934 individuals. There was no cross over of individual subjects between these references. A summary of the population demographics, chemotherapy regimens and actual outcomes in references is provided in Appendix. We evaluate our DE algorithm and benchmarks on the de-identified database of 10,000 breast cancer patients which was created based on the patients participating in the National

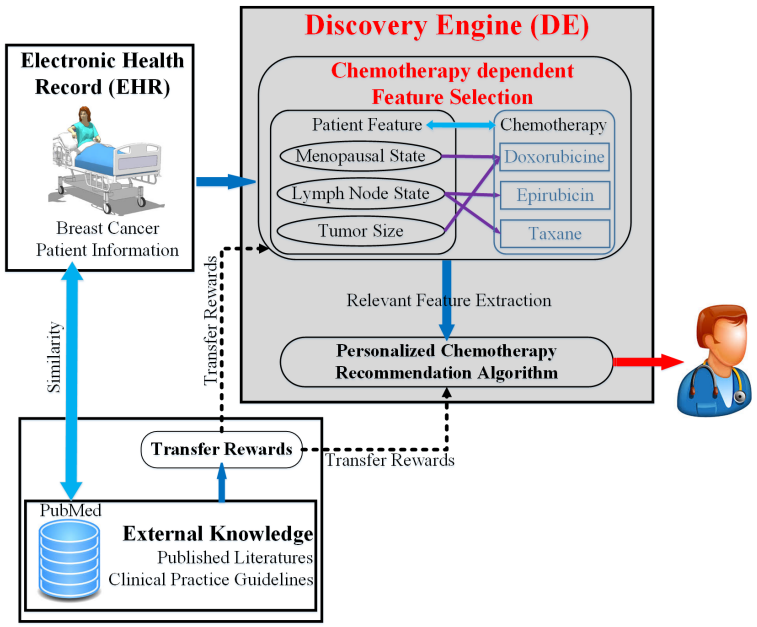

Figure 5.1: Personalized chemotherapy recommendation system for breast cancer patients using DE algorithm

Surgical Adjuvant Breast and Bowel Project (NSABP) in the United States and Canada. The patients were diagnosed with operable and palpable breast cancer by core needle biopsy or fine needle aspiration. The patient data is characterized by 15 features summarized in Table 5.2 and those are also corresponding with the patient features in 32 references.

Table 5.1: Code for each chemotherapy regimen

Code	**Specific Chemotherapy Regimen**
AC	Doxorubicin + Cyclophosphamide
ACT	Doxorubicin + Cyclophosphamide + Taxanes
AT	Doxorubicin + Taxanes
CAF	Cyclophosphamide + Doxorubicin + 5-Fluorouracil
CEF	Cyclophosphamide + Epirubicin + 5-Fluorouracil
CMF	Cyclophosphamide + Methotrexate + 5-Fluorouracil

We iteratively evaluated the performance of the algorithms based on 10 rounds with 10 different training sets and reported the average performance as a final performance of each algorithm. In each round, we used a randomly selected

training set of 4,000 patients among 10,000 entire patients and a disjoint testing set of 6,000 patients. In other words, no training data were used during testing of the model, but 10 different models were used to derive the average performance. We select 4,000 patients to be in the training set, since the performance of all algorithms (besides ACL) saturated beyond this number of patients.

Table 5.2: Summary of patient information features

Feature	**Range**	**Feature**	**Range**
Age	30s ~ 60+	PLNC*	0 ~ 10+
Menopausal	Pre/Post	Lymph Node Status	Pos/Neg
Race	White/Black/Other	WHO Score*	0 ~ 5
Estrogen Receptor	Pos/Neg	Surgery Type	BCT*/MRM*/ No
Progesterone Receptor	Pos/Neg	Prior Radiotherapy	Exp / No
HER2NEU*	Pos/Neg/Neu	Prior Chemotherapy	Exp / No
Tumor Stage	T1 ~ T4	Histology	Ductal / Mix / Lobular
Tumor Grade	G1 ~ G3		

*PLNC: positive axillary lymph node count
*BCT: breast conservative therapy. MRM: modified radical mastectomy
*HER2NEU: human epidermal growth factor receptor 2
*WHO score: Eastern Cooperative Oncology Group (ECOG) score which is widely used in publications by the World Health Organization (WHO)
*Features with categorical values are changed mutually exclusive binary indicator for the evaluation.

5.2 Benchmarks

We compare the performance of DE with 9 state-of-the-art classification algorithms, ensemble learning algorithms and feature selection algorithms which are widely used in CDSS and also commonly used benchmarks in medical and machine learning references:

- Correlation Feature Selection (CFS): a well-known feature selection algorithm [Hal99];
- All Contextual Learning (ACL): a well-known contextual learning algorithm which uses all features. This is a modified offline version of the contextual bandit algorithm of Slivkins [GOS08];
- Multivariate Logistic Regression (Logit);
- Linear Regression (Linear);
- Multivariate Support Vector Machines (SVM); we use a radial basis function (RBF) kernel SVM;
- Support Vector Machines with Feature Selection (SVMs-f) [WMC00];
- Adaptive Boosting (AdaBo);
- Classification Tree (CTree);
- Regularized Multivariate Logistic Regression using Lasso (ReLog);
- Regularized Linear Regression using Lasso (ReLin);

5.3 Success of the Optimal Chemotherapy Recommendation for Breast Cancer Patients

Given a patient, both our algorithm and the benchmark algorithm recommend a chemotherapy regimen corresponding to particular references. If the recommended chemotherapy has the highest estimated transfer reward for the patient among all six chemotherapy regimens, we regard the algorithm in question as making the correct recommendation for that patient; i.e. it has recommended the best course of treatment. (Notice that the best course of treatment may not promise a good

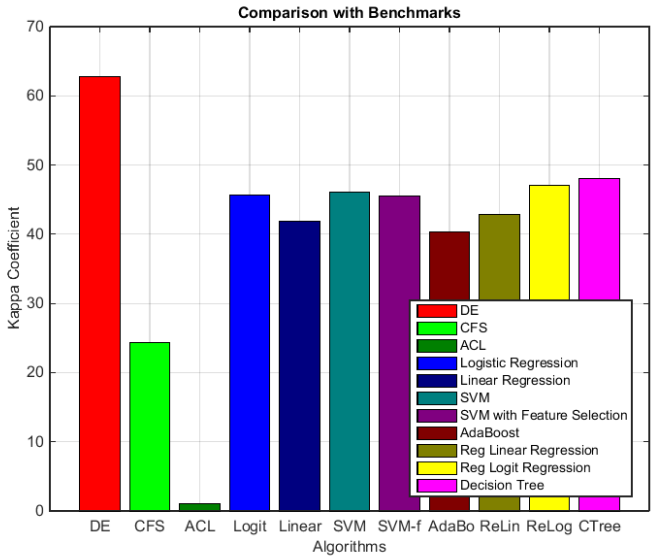

Figure 5.2: Performance analysis with benchmark algorithms in terms of Kappa coefficient

outcome: some cancers are not treatable.) We take the fraction/percentage of correct recommendations to be the success rate for the algorithm in question.

Given the success rate for the algorithm, we apply two performance metrics: the simple percent agreement and the Cohen's kappa coefficient [HJ15]. Simple percent agreement (p_0) is the success rate (the fraction of times the personalized treatment prediction coincides with the recommendation provided in the medical literature for the patient). Cohen's kappa coefficient (κ) is a metric which measures inter-rater agreement. It is usually considered a more robust measure than a simple percent agreement (p_0), because κ measures the improvement over chance agreements. If p_e is the probability of agreement by chance, then, kappa coefficient is defined as $\kappa = (p_0 - p_e)/(1 - p_e)$.

The bar graphs in Fig. 5.3(a) show that the first chemotherapy recommendation of DE is successful (as defined above) 73.4% of the time and Fig. 5.3(b) also shows that one of the first two recommendations is successful 88.4% of the time. This is 7.7% better than the second best approach (SVM) in terms of selecting

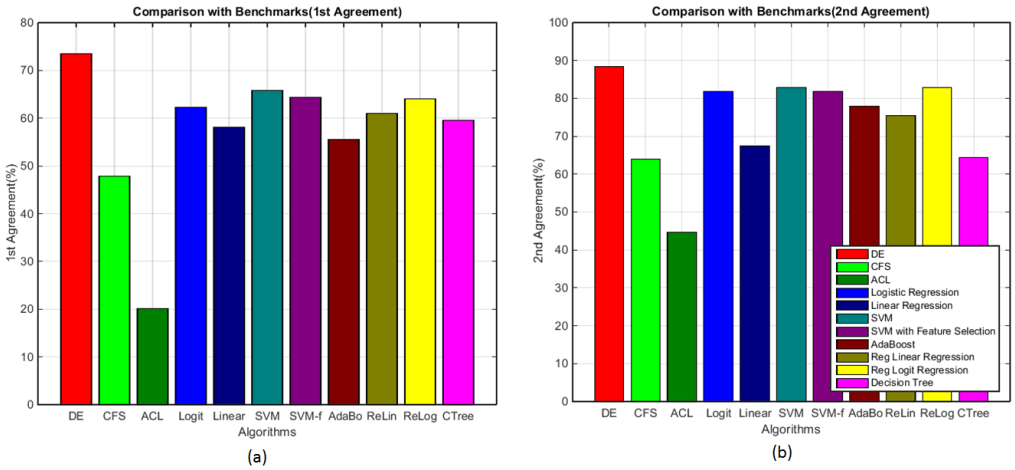

Figure 5.3: Performance analysis with benchmark algorithms (a) 1st simple percent agreement, (b) 2nd simple percent agreement

the optimal chemotherapy on its first choice (p-value < 0.01), and 5.6% better in terms of matching the optimal chemotherapy within the first 2 choices (p-value < 0.01). This is already a significant improvement; however, in terms of kappa coefficients, the improvement is even greater: DE works 16.6% better than SVM which is described in Fig. 5.2 (p-value < 0.01). This is because SVM indiscriminately recommends the popular chemotherapies and is not robust when classifying the less popular chemotherapies. Given robustness considerations, which are essential in medical treatment recommendations, kappa coefficients are more often used as a performance metric in medical informatics.

When comparing our algorithm with other algorithms that rely on feature selection, we again see a significant improvement. Again, note that while other algorithms use feature selection, they do not select relevant features for specific chemotherapies, and it is through this selection that our algorithm achieves improvement. CFS achieves only a 48% of simple percent agreement because it cannot use the efficacy of the chemotherapy to discriminate the relevant features and hence the technique is entirely unsupervised. ACL succumbs to the "curse of dimensionality" because there are 15 features with different ranges, resulting in

over 7 million combinations to explore. Logistic regression, linear regression, and SVM perform worse than DE because they do not consider the relevant features for selecting chemotherapies at all.

5.4 Relevant Features for Each Chemotherapy

Table 5.3 shows the top 4 ranked relevant features discovered by CDFS - tumor stage, positive axillary lymph node number (PLNC), estrogen receptor etc.- for recommending AC, ACT, AT, CAF, CEF and CMF chemotherapy. As it can be seen from Table 5.3, CDFS is able to discover the different relevant features that are relevant for different chemotherapy regimens.

Table 5.3: Discovered relevant feature for each chemotherapy

Chemotherapy Code	**1st Relevant Feature**	**2nd Relevant Feature**	**3rd Relevant Feature**	**4th Relevant Feature**
AC	PLNC	tumor Stage	estrogen Receptor	Age
ACT	Tumor Stage	Prior Chemotherapy	PLNC	Estrogen Receptor
AT	Prior Chemotherapy	PLNC	Surgery Type	Age
CAF	Surgery Type	Tumor Stage	Age	Tumor Grade
CEF	PLNC	Estrogen Receptor	Tumor Stage	Age
CMF	Estrogen Receptor	PLNC	Radiotherapy	Tumor Stage

It is important to note the features discovered by DE are indeed confirmed to be relevant by clinical studies. Firstly, note that the six considered chemotherapy regimens are commonly recommended to node positive breast cancer patients, i.e. patients where cancer has been found in the lymph nodes [FGR15]. It is extremely

important to know whether lymph nodes are positive or negative. PLNC tells us both the number of nodes and whether lymph nodes are positive or negative. For instance, zero PLNC implies node negative breast cancer, while otherwise indicates node positive breast cancer. Hence, PLNC is selected as a relevant feature by CDFS. Secondly, the menopausal status is considered important because medications affect cancer differently in premenopausal and postmenopausal women [Lev01]. More specifically, the CEF chemotherapy is only recommended to premenopausal women. Although the menopausal status is not included in this relevant feature set, women over the age of 50 are usually considered postmenopausal [Lev01]. Therefore, age was correctly identified by DE to be a discriminative feature for selecting among chemotherapy regimens. Thirdly, tumor stage is another important feature to consider when deciding among chemotherapy regimens as described in reference [BLS14]. Medications A(Doxorubicin), T(Taxotere), E(Epirubicin) are recommended for advanced breast cancer and our top six chemotherapy regimens include more than one of these medications. Therefore, DE has correctly discovered that the features that are relevant for these chemotherapy regimens contain tumor stage information. Finally, the medication T(Taxotere) is usually recommended to breast cancer patients who do not respond to their current chemotherapy. Thus, the prior chemotherapy information is correctly discovered by DE to be relevant for AT and ACT therapies.

5.5 Performance when Patient Information is Missing

As explained before, patient information is often missing from the EHR. Moreover, studies have shown that the missing information is often not random [Coh60]. For example, the age of the patient is easy to record and blood pressure is often verified several times by a nurse when a patient is seen by a clinician, so it is typically neither missing nor incorrect. However, HER2NEU may not be recorded

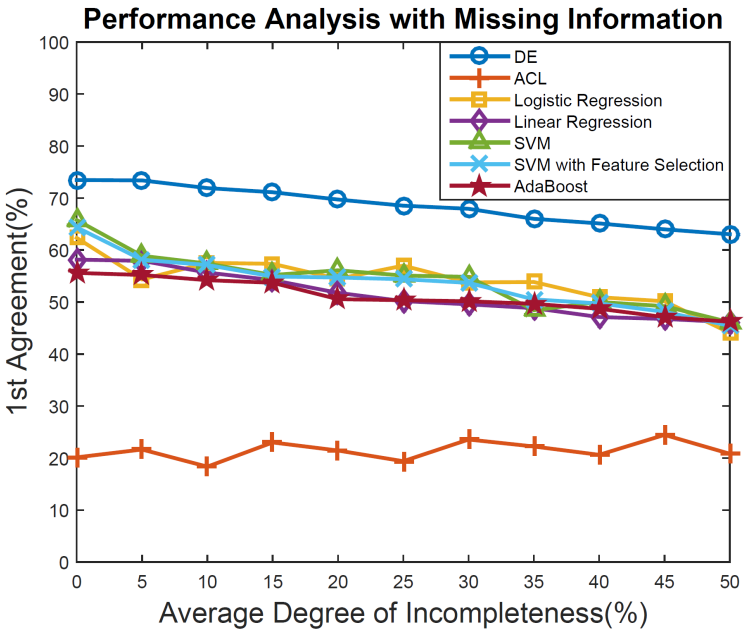

Figure 5.4: Performance analysis with missing information

depending on the diagnostic tests ordered and capabilities of a medical center.

Fig. 5.4. describes the performance of DE when features are missing with various rates. It shows the performance degradation of DE and of the benchmark algorithms as a function of the average degree of incompleteness [Coh60]. We did not use the percentage of missing features as a metric since the features are not randomly missing. The percentages of missing features corresponding to each degree of incompleteness are described in Appendix; these percentages were computed based on statistics extracted from medical records of patients. Fig. 5.4. shows that the performance of DE degrades from 73.4% to 63.0% (when the average degree of incompleteness is 50%). However, even with missing information, DE continues to outperform the other methods. DE discovers relevant features with low missing probability, and is able to estimate the missing feature information based on the available feature information. As a result, the impact of missing information is minimized. In fact, DE performs better than most other algorithms even when DE misses significant amounts of information from the EHR while the

other algorithms make their decision with full information. Hence we can indeed see that the performance of DE is robust even when information is missing.

CHAPTER 6

Experiment II: Diagnosis Decision Support System for Breast Cancer Patients

In this section, we illustrate how the DE algorithm can be used for breast cancer diagnosis. In this case, we can directly use patients' actual outcomes as the rewards. Fig. 6.1 describes the system model of this application.

6.1 Data Description

Table 6.1: Summary of the feature information

No	Information Type	Explanation
1	Radius	Mean of distance from center to points on the perimeter
2	Texture	Standard deviation of gray-scale values
3	Perimeter	The perimeter of tumor cell nucleus
4	Area	The area of tumor cell nucleus
5	Smoothness	Local variation in radius lengths
6	Compactness	$Perimeter^2/area - 1$
7	Concavity	Severity of concave portions of the contour
8	Concave Points	Number of concave portions of the contour
9	Symmetry	Symmetricity of tumor cell nucleus
10	Fractal Dimension	Coastline approximation – 1
*Features consist of mean, standard errors and worst of above 10 info types. *Each info type is computed real value feature for each tumor cell nucleus.		

In this section we evaluate the performance of DE for breast cancer diagnosis using the UCI Diagnostic Wisconsin Breast Cancer Database [BHC10]. The dataset contains 30 patient features extracted from needle biopsy features such as

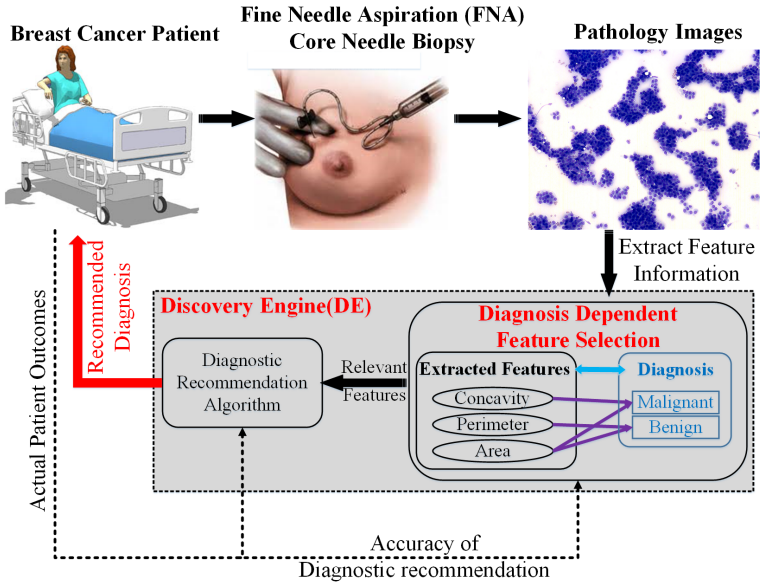

Figure 6.1: Personalized diagnostic recommendation system for breast cancer patients using DE algorithm

radius, compactness, or smoothness of tumor cell nucleus. Table 6.1 summarizes the details of 30 patient features. The number of instances in this dataset is 569 and the diagnosis (label) for each instance is either malignant or benign.

6.2 Benchmarks

We compare the performance of DE in this clinical setting with all the benchmarks describes in Section VI-B. In addition, we add 3 existing state-of-the-art feature selection algorithms as the benchmarks in order to compare the performance of our feature selection algorithm (CDFS) separately.

- Mutual Information Feature Selection (MIFS): a well-known feature selection algorithm based on mutual information [PLD05];
- Relevance Learning with Feedback (RELEAF): an action dependent relevance learning algorithm based on the expected rewards [TV14, TS15];

- Principal Component Analysis (PCA): a statistical procedure to discover linearly uncorrelated variables based on orthogonal transformation [AN07]

6.3 Experiments Setup

First, we compare our DE algorithm against state-of-the-art machine learning algorithms: logistic/linear regressions, rbf kernel SVM, adaptive boosting algorithms, classification tree and regularized logistic/linear regressions. We use 10-fold cross-validation in order to evaluate the performance of algorithms. We performed 10 independent cross validation runs and report the average performance of 10 runs.

To highlight the impacts of CDFS, we performed two additional sets of experiments. In the first set, we compared the performance of our DE system using CDFS with the performance of the DE system where CDFS was replaced with one of the four different feature selection algorithms: CFS, MIFS, RELEAF, and PCA. This comparison shows the impact of CDFS on the overall performance of the DE. Other experiment settings are exactly the same as the first experiment.

In the second set of additional experiments, we use our feature selection algorithm CDFS in conjunction with the diagnostic recommendation made by the benchmark algorithms - linear regression, logistic regression, SVM - to highlight the specific impact of our feature selection algorithm. Other experiment settings are exactly the same as the first experiment.

6.4 Performance of Diagnostic Recommendation for Breast Cancer Patients

Given a patient, DE and the other benchmark algorithms classify the tumors as malignant or benign. To quantify their performance, we apply three different per-

formance metrics: the prediction error rate (PER), the false positive rate (FPR), and the false negative rate (FNR). PER is defined as the fraction of times the recommended diagnosis of our algorithm is different from the actual label. FPR and FNR are defined as the diagnosis error rate for benign instances and for malignant instances, respectively. The main goal of DDSS is to minimize the FPR given an allowable threshold for FNR as selected by the clinicians. In practice, this is often set to be below 2% [PG07a]. Therefore, in this experiment, the FNR threshold is also set to be 2%. Using this threshold, we can characterize our performance metrics as follows.

$$\text{minimize FPR}$$

$$\text{subject to FNR} \leq 2\%$$

6.4.1 Comparison with Machine Learning Algorithms

Table 6.2: Comparison with state-of-the-art machine learning techniques

%	**DE**	**L**ogit	**Linear**	**SVMs**	**SVMs-f**	**Adabo**	**CTree**	**ReLog**	**ReLin**
PER	2.23	11.77	8.47	4.41	4.52	9.12	11.45	6.71	5.51
FPR	2.62	18.3	13.55	6.82	7.03	14.86	18.64	10.11	9.15
FNR	1.92	1.96	1.98	1.99	1.98	1.91	1.95	1.92	1.93

As the Table 6.2 shows, our DE algorithm has 2.23% prediction error rates and 2.62% false positive rates which is 2.18% and 4.20% better than the second best algorithm (SVMs) when the tolerable threshold of FNR is set to below 2% (p-value < 0.01). There are two reasons for the outstanding performance of the DE approach. First, our diagnostic recommendation algorithm yields high accuracy for classification, because it is able to provide personalized diagnosis, while other comparable algorithms apply the same model for all patients. Second, DE can discover different relevant features for different diagnosis based on CDFS, while the

other algorithms (Logistic/Linear Regression, SVMs, AdaBoost and Classification Tree) base their decisions on all the features without relevant feature discovery.

6.4.2 Comparison with Feature Selection Algorithms

In this subsection, we demonstrate the impact of the CDFS algorithm on the DE system. We compare the performance of the DE using CDFS with the performance of DE using different feature selection algorithms.

Table 6.3: CDFS Performance comparison with other feature selection algorithms

%	**CDFS**	**RELEAF**	**CFS**	**MIFS**	**PCA**
PER	2.23	18.37	2.76	3.19	3.94
FPR	2.62	24.11	3.90	3.99	6.44
FNR	1.92	1.96	1.98	1.90	1.94

As seen in Table 6.3, CDFS outperforms all other feature selection algorithms when the tolerable threshold of FNR is set to below 2%. This is because CDFS is capable of discovering diagnosis relevant features based on their impact on the expected diagnosis accuracies. Although RELEAF also considers the dependence between diagnosis accuracy and feature selection, it is extremely slow when the number of features is large, as is the case for this and many other medical datasets. Furthermore, combinatorial approach (RELEAF) requires a relatively large amount of training sets to accurately discover the relevant feature, which is not the case for the medical dataset available to us.

Next, we replace DE's recommendation algorithms with various existing machine learning algorithms in order to demonstrate the impact of the CDFS component of DE on the diagnostic decisions. As seen in Table 6.4, CDFS improves the performance of all benchmark algorithms because it is able to accurately discover and select the (different) features that are relevant for different diagnosis.

Table 6.4: Impact of the CDFS in conjunction with alternative machine learning algorithms

	PER (%)		**PER (%)**	
	CDFS	w/o CDFS	CDFS	w/o CDFS
Linear	5.49	8.47	8.88	13.55
Logit	7.84	11.77	12.70	18.30
SVMs	4.01	4.41	5.51	6.82

6.5 Relevant Features for Diagnostic Decision

Table 6.5 shows the top 5 ranked relevant features discovered by CDFS: worst perimeter, concave points concavity, radius and area etc. for diagnosing malignant or benign cancer among all of the features summarized in Table 6.1. As seen in Table 6.2, CDFS is able to discover the different features that are relevant for different diagnoses.

Table 6.5: Discovered relevant features for each diagnosis

	Malignant	**Benign**
1st Relevant Feature	Worst perimeter	Worst concave points
2nd Relevant Feature	Worst concave points	Mean concave points
3rd Relevant Feature	Worst radius	Worst perimeter
4th Relevant Feature	Mean concavity	Worst radius
5th Relevant Feature	Worst area	Mean concavity

It should be noted that the relevance of the features discovered by DE confirms findings of prior clinical studies. For instance, studies of breast biopsies [Per01] state that the 3 most important factors to diagnose tumor cell nuclei as malignant or benign are the relative size ratio between nucleus and cytoplasm, irregular shape, and irregular chromatin. However, because chromatin feature information is not available in our dataset, only the relative size and irregular shape can be potential candidates as the relevant features. The size related features are radius, perimeter, and area, and the shape related features are the concavity and concave

points [Per01]. The top 5 features found to be relevant by DE to classify malignant and benign tumor cell nuclei are all related to the tumor shape and relative size, which is in accordance to reference [Per01]. Features such as texture, smoothness, compactness, symmetry, and fractal dimension are not found to be relevant by DE and are not mentioned as important features in reference [Per01]. Hence, we can conclude that DE can discover the relevant features for making a correct diagnosis without prior medical knowledge.

CHAPTER 7

Discussion and Future Works

We proposed a novel approach that discovers the most relevant information from the EHR to determine which clinical decision to recommend for a patient. Furthermore, our approach uses this information to provide personalized recommendations to assist physicians in their decision making process. Our results demonstrate that DE outperforms existing machine learning, prediction, and feature selection methods in both CDSS and DDSS applications. These superior recommendations are extremely important because they have the potential to prevent medical errors and thus improve the quality of medical care. We also showed that our method is robust against missing information, which is important in numerous clinical settings.

The limitations of the thesis can be summarized as follows. First, we assume that the matching a patient to a population distribution in literatures will achieve similar clinical rewards for the same treatment. However, it is hard to distinguish in practice. Therefore, it would be interesting to further improve Chapter 4 using propensity score matching, similarity learning approach or off-policy estimation methods to estimate the counterfactual rewards. Second, even though our DE approach can be applicable in general, its performance is only validated on breast cancer patients. Therefore, it would be interesting to apply diverse complex diseases for validating the generality and the outperformance of DE approach.

Future work will consider that feature information (such as the tumor size, PLNC number and tumor radius in the case of a breast tumor) may change over

time. The time dependence may also influence the duration of a therapy and the selection of future therapies. Another DE extension will consider the global sequence of treatment decisions that optimizes long-term outcomes (e.g. overall survival rate or 5-year recurrence rate).

In conclusion, we believe that our proposed contextual learning approach demonstrates promise towards providing useful personalized clinical recommendations. As new types of treatment are evaluated and approved for medical use, clinicians will have an increasingly difficult time determining which clinical decisions are most effective for individual patients. DE provides a pathway towards providing computational methods for personalized clinical decision recommendations.

CHAPTER 8

Appendix

8.1 Nomenclature Table

Table 8.1: Nomenclature table

Notation	Interpretation	Notation	Interpretation
A	Entire set of clinical decisions	n	Current patient number
$A(E_i)$	Set of clinical decisions considered in reference i	$P(.)$	Probability of random variable
a	Clinical decision	$P(x\|E_i)$	Probability of feature x in i-th reference
$a^*(x)$	Optimal recommended clinical decision	p_0	Simple percent agreement
a_k	k-th clinical decision	p_e	The probability of agreement by chance
$\hat{a}(x)$	Recommended clinical decision by DE approach	$\mathcal{R}$	Relevant feature set for all a
C	Threshold cost	$\mathcal{R}(a)$	Relevant feature set for clinical decision a

Continued on next page

Table 8.1 – continued from previous page

Notation	Interpretation	Notation	Interpretation
C_{TH}	Coefficient of control function	$\hat{\mathcal{R}}(a)$	Discovered relevant feature set for clinical decision a
D	Total number of features	$\hat{\mathcal{R}}(E_i)$	Discovered relevant feature set for reference i
E_i	i-th reference	$r(a\|E_i)$	Actual patient outcome for a in i-th reference
$\mathbb{E}()$	Expectation of random variable	$Sim_(E_i)(x)$	Similarity between E_i and x
F	Entire set of patient features	$\hat{Sim}_(a\|E_i)(x)$	Normalized similarity for clinical decision a
f	Patient feature	$tr_a(x)$	Transfer reward of a given x
f_k	k-th patient feature	U_f	Utility function
H_N	Available information for N previous patients	x	Patient information vector
$h^d_{f,s}$	Redundancy metric	x_R	Relevant feature information
h^r_f	Relevance metric	x_k	Patient information of k-th feature
$I(f;E)$	Mutual information between f and E	x^{av}	Available patient features
κ	Cohen's kappa statistics	x^{miss}	Missing patient features
N	Total patient number	y	Rewards

Continued on next page

Table 8.1 – continued from previous page

Notation	Interpretation	Notation	Interpretation
N_a	Number of patients with clinical decision a	$\hat{y}_a^S(x_s)$	Sample mean reward estimator for patients contains x_s
$N_a^S(x_s)$	Number of patients contains x_s	$\hat{y}_a$	Sample mean reward estimator of clinical decision a

8.2 Pseudo-codes

Algorithm 1: Discovery Engine (DE)

Initialize: $\hat{\mathcal{R}}(a) = \varnothing, \hat{\mathcal{R}}^c(E_i) = \{f_1, f_2, ..., f_D\}$ for each a

while $n \geq 1$ **do**

(1) Clinical Decision Dependent Feature Selection

for each clinical decision a **do**

$G = \text{argmax}_{f \in \hat{\mathcal{R}}^c(a)} h_f^r(a)$

$\hat{\mathcal{R}}(a) = \hat{\mathcal{R}}(a) \cup G$

while $\max_{f \in \hat{\mathcal{R}}^c(a)} U_f(a) > 0$ **do**

$H = \text{argmax}_{f \in \hat{\mathcal{R}}^c(a)} U_f(a)$

$\hat{\mathcal{R}}(E_i) = \hat{\mathcal{R}}(E_i) \cup H$

end while

end for

(2) Optimal clinical decision recommendation

$U = \{a \in A | N_a^{\hat{\mathcal{R}}(a)}(x_{\hat{\mathcal{R}}(a)}) < C_{TH} \times \log(n)\}$

if $U = \varnothing$ **then**

$\hat{a}(x) = \text{argmax}_a \hat{y}_a^{\hat{\mathcal{R}}}(x_{\hat{\mathcal{R}}})$

end if

(3) Update

$N_{\hat{a}}^{\hat{\mathcal{R}}(a)}(x_{\hat{\mathcal{R}}(a)} = \sum_n \mathbb{I}\{x_{\hat{\mathcal{R}}(a)} \subset x(n)\} \times \mathbb{I}\{a(n) = \hat{a}\}$

$\hat{y}_a^{\hat{\mathcal{R}}(a)}(x_{\hat{\mathcal{R}}(a)}) = \frac{y(n)}{N_a^{\hat{\mathcal{R}}(a)}(x_s)} \sum_n \mathbb{I}\{x_{\hat{\mathcal{R}}(a)} \subset x(n)\} \times \mathbb{I}\{a(n) = \hat{a}\}$

end while

Figure 8.1: Pseudo-code I: Discovery engine (DE).

Algorithm 2: Transfer Reward Estimation

Initialize : $\hat{\mathcal{R}}(E_i) = \varnothing, \hat{\mathcal{R}}^c(E_i) = \{f_1, f_2, ..., f_D\}$ for each reference E_i

for each reference E_i **do**

$G = \text{argmax}_{f \in \hat{\mathcal{R}}^c(E_i)} I(x_f; E_i)$

$\hat{\mathcal{R}}(E_i) = \hat{\mathcal{R}}(E_i) \cup G$

while $\max_{f \in \hat{\mathcal{R}}^c(E_i)} U_f(E_i) > 0$ **do**

$H = \text{argmax}_{f \in \hat{\mathcal{R}}^c(E_i)} U_f(E_i)$

$\hat{\mathcal{R}}(E_i) = \hat{\mathcal{R}}(E_i) \cup H$

end while

$Sim_{E_i}(x) = \log P(E_i) + \sum_{l \in \hat{\mathcal{R}}(E_i)} \log \frac{P(X_l = x_l(n) | E_i}{P(X_l = x_l(n)}$

for each clinical decision a **do**

$\hat{Sim}_{a|E_i}(x) = \frac{Sim_{E_i}(x)}{\sum_{i:a \in A(E_i)} Sim_{E_i}(x)}$

$tr_a(x) = \sum_{i:a \in A(E_i)} \hat{Sim}_{a|E_i}(x) \times r(a|E_i)$

end for

end for

Figure 8.2: Pseudo-code II: Transfer reward estimation.

8.3 Summary of the Information in References

Table 8.2: Summary of the information in references

Medical Paper No		Paper 1	Paper 2	...	Paper 32	Entire Papers
Population Demo	Age	**30s:** 21% **40s:** 32% **50s:** 27% **60s:** 20%	**30s:** 15% **40s:** 35% **50s:** 30% **60s:** 20%	...	**30s:** 16% **40s:** 47% **50s:** 22% **60s:** 15%	**30s:** 16% **40s:** 40% **50s:** 27% **60s:** 16%
	Estrogen Receptor	**Pos:** 61% **Neg:** 22% **Miss:** 17%	Missing	...	**Pos:** 26% **Neg:** 74% **Miss:** 0%	**Pos:** 63% **Neg:** 33% **Miss:** 4%
	Tumor Grade	Missing	Missing	...	**G1:** 7% **G2:** 47% **G3:** 40% **Miss:** 6%	**G1:** 15% **G2:** 41% **G3:** 38% **Miss:** 6%
	...	...				
	Tumor Size	Missing	**T1:** 19% **T2:** 58% **T2:** 58% **T2:** 58% **T2:** 58%	...	**T1:** 41% **T2:** 51% **T2:** 51% **T2:** 51% **T2:** 51%	**T1:** 37% **T2:** 40% **T2:** 40% **T2:** 40% **T2:** 40%
	Lymph Node Status	**Pos:** 99% **Neg:** 0% **Miss:** 1%	Missing	...	**Pos:** 66% **Neg:** 36% **Miss:** 0%	**Pos:** 33% **Neg:** 66% **Miss:** 0%
Chemotherapy		AC CMF	AT CAF	...	CEF CMF	
Actual Outcomes (5 year survival rates)		AC: 95% CMF: 92%	AT: 72% CAF: 78%	...	CEF: 82% CMF: 81%	
*We can use various actual outcomes from references. (e.g. 5 year recurrence free survival rates)						

8.4 Full Lists of 32 Medical References for Experiment I

Table 8.3: Entire lists of 32 medical references for experiment 1

No	Reference
1	A. Di Leo, D. Gancberg, D. Larsimont, M. Tanner, T. Jarvinen, G. Rouas and M. J. Piccart, "HER-2 amplification and topoisomerase II α gene aberrations as predictive markers in node-positive breast cancer patients randomly treated either with an anthracycline-based therapy or with cyclophosphamide, methotrexate, and 5-fluorouracil," Clinical Cancer Research, vol. 8, no. 5, pp. 1107-1116, 2002.
2	P. Therasse, L. Mauriac, M. Welnicka-Jaskiewicz, P. Bruning, T. Cufer, H. Bonnefoi and M. J. Piccart, "Final results of a randomized phase III trial comparing cyclophosphamide, epirubicin, and fluorouracil with a dose-intensified epirubicin and cyclophosphamide+ filgrastim as neoadjuvant treatment in locally advanced breast cancer: an EORTC-NCIC-SAKK multicenter study," Journal of clinical oncology, vol. 21, no. 5, pp. 843-850, 2003.
3	S. M. Bang, D. S. Heo, K. H.. Lee, J. H. Byun, H. M. Chang, D. Y. Noh and N. K. Kim, "Adjuvant doxorubicin and cyclophosphamide versus cyclophosphamide, methotrexate, and 5 fluorouracil chemotherapy in premenopausal women with axillary lymph node positive breast carcinoma," Cancer, vol. 89, no. 12, pp. 2521-2526, 2000.
Continued on next page	

Table 8.3 – continued from previous page

No	Reference
4	B. Fisher, S. Anderson, E. Tan-Chiu, N. Wolmark, D. L. Wickerham, E. R. Fisher and W. B. Farrar, "Tamoxifen and chemotherapy for axillary node-negative, estrogen receptor negative breast cancer: findings from National Surgical Adjuvant Breast and Bowel Project B-23," Journal of clinical oncology, vol. 19, no. 4, pp. 931-942, 2001.
5	L. J. Goldstein, A. O'Neill, J. A. Sparano, E. A. Perez, L. N. Shulman, S. Martino and N. E. Davidson, "Concurrent doxorubicin plus docetaxel is not more effective than concurrent doxorubicin plus cyclophosphamide in operable breast cancer with 0 to 3 positive axillary nodes: North American Breast Cancer Intergroup Trial E 2197," Journal of Clinical Oncology, vol. 26, no. 25, pp. 4092-4099, 2008.
6	J. Jassem, T. Pienkowski, A. Piuzanska, S. Jelic, V. Gorbunova, Z. Mrsic-Krmpotic and C. Weil, "Doxorubicin and paclitaxel versus fluorouracil, doxorubicin, and cyclophosphamide as first-line therapy for women with metastatic breast cancer: final results of a randomized phase III multicenter trial," Journal of Clinical Oncology, vol. 19, no. 6, pp. 1707-1715, 2001.
7	P. Rastogi, S. J. Anderson, H. D. Bear, C. E. Geyer, M. S. Kahlenberg, A. Robidoux and N. Wolmark, "Preoperative chemotherapy: updates of national surgical adjuvant breast and bowel project protocols B-18 and B-27," Journal of Clinical Oncology, vol. 26, no. 5, pp. 778-785, 2008.

Continued on next page

Table 8.3 – continued from previous page

No	Reference
8	H. D. Bear, S. Anderson, A. Brown, R. Smith, E. P. Mamounas, B. Fisher and N. Wolmark, "The effect on tumor response of adding sequential preoperative docetaxel to preoperative doxorubicin and cyclophosphamide: preliminary results from National Surgical Adjuvant Breast and Bowel Project Protocol B-27," Journal of Clinical Oncology, vol. 21, no. 22, pp. 4165-4174, 2003.
9	S. Jones, F. A. Holmes, J. O'Shaughnessy, J. L. Blum, S. J. Vukelja, K. J. McIntyre and M. A. Savin, "Docetaxel with cyclophosphamide is associated with an overall survival benefit compared with doxorubicin and cyclophosphamide: 7-year follow-up of US Oncology Research Trial 9735," Journal of Clinical Oncology, vol. 27, no. 8, pp. 1177-1183, 2009.
10	M. Martin, A. Lluch, M. A. Segui, A. Ruiz, M. Ramos, E. Adrover, and J. R. Mel, "Toxicity and health-related quality of life in breast cancer patients receiving adjuvant docetaxel, doxorubicin, cyclophosphamide (TAC) or 5-fluorouracil, doxorubicin and cyclophosphamide (FAC): impact of adding primary prophylactic granulocyte-colony stimulating factor to the TAC regimen," Annals of oncology, vol. 17, no. 8, pp. 1205-1212, 2006.
11	M. Martin, T. Pienkowski, J. Mackey, M. Pawlicki, J. P. Guastalla, C. Weaver and C. Vogel, "Adjuvant docetaxel for node-positive breast cancer," New England Journal of Medicine, vol. 352,no. 22, pp. 2302-2313, 2005.

Continued on next page

Table 8.3 – continued from previous page

No	Reference
12	J. M. Nabholtz, J. R. Mackey, M. Smylie, A. Paterson, D. R. Noel, T. Al-Tweigeri, and A. Riva, "Phase II study of docetaxel, doxorubicin, and cyclophosphamide as first-line chemotherapy for metastatic breast cancer," Journal of clinical oncology, vol. 19, no. 2, pp. 314-321, 2001.
13	A. Rody, T. Karn, R. Gatje, A. Ahr, C. Solbach, K. Kourtis, and M. Kaufmann, "Gene expression profiling of breast cancer patients treated with docetaxel, doxorubicin, and cyclophosphamide within the GEPARTRIO trial: HER-2, but not topoisomerase II α and microtubule-associated protein tau, is highly predictive of tumor response," The Breast, vol. 16, no. 1, pp. 86-93, 2007.
14	M. Martin, A. Villar, A. Sole-Calvo, R. Gonzalez, B. Massuti, J. Lizon and E. Diaz-Rubio, "Doxorubicin in combination with fluorouracil and cyclophosphamide (iv FAC regimen, day 1, 21) versus methotrexate in combination with fluorouracil and cyclophosphamide (iv CMF regimen, day 1, 21) as adjuvant chemotherapy for operable breast cancer: a study by the GEICAM group," Annals of Oncology, vol. 14, no. 6, pp. 833-842, 2003.
15	K. S. Albain, W. E. Barlow, S. Shak, G. N. Hortobagyi, R. B. Livingston, I. T. Yeh and Breast Cancer Intergroup of North America, "Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial," The lancet oncology, vol. 11, no. 1, pp. 55-65, 2010.

Continued on next page

Table 8.3 – continued from previous page

No	Reference
16	J. L. Misset, M. di Palma, M. Delgado, R. Plagne, P. Chollet, P. Fumoleau and G. Mathe, "Adjuvant treatment of node-positive breast cancer with cyclophosphamide, doxorubicin, fluorouracil, and vincristine versus cyclophosphamide, methotrexate, and fluorouracil: final report after a 16-year median follow-up duration," Journal of clinical oncology, vol. 14, no. 4, pp. 1136-1145, 1996.
17	M. N. Levine, V. H. Bramwell, K. L. Pritchard, B. D. Norris, L. E. Shepherd, H. Abu-Zahra and D. Tu, "Randomized trial of intensive cyclophosphamide, epirubicin, and fluorouracil chemotherapy compared with cyclophosphamide, methotrexate, and fluorouracil in premenopausal women with node-positive breast cancer. National Cancer Institute of Canada Clinical Trials Group," Journal of Clinical Oncology, vol. 16, no. 8, pp. 2651-2658, 1998.
18	A. S. Knoop, H. Knudsen, E. Balslev, B. B. Rasmussen, J. Overgaard, K. V. Nielsen and B. Ejlertsen, "Retrospective analysis of topoisomerase II α amplifications and deletions as predictive markers in primary breast cancer patients randomly assigned to cyclophosphamide, methotrexate, and fluorouracil or cyclophosphamide, epirubicin, and fluorouracil: Danish Breast Cancer Cooperative Group," Journal of Clinical Oncology, vol. 23, no. 30, pp. 7483-7490, 2005.
19	K. I. Pritchard, L. E. Shepherd, F. P. O'Malley, I. L. Andrulis, D. Tu, V. H. Bramwell and M. N. Levine, "HER2 and responsiveness of breast cancer to adjuvant chemotherapy," New England Journal of Medicine, vol. 354, no. 20, pp. 2103-2111, 2006.

Continued on next page

Table 8.3 – continued from previous page

No	Reference
20	A. Webb, D. Cunningham, J. H. Scarffe, P. Harper, A. Norman, J. K. Joffe and M. Meehan, "Randomized trial comparing epirubicin, cisplatin, and fluorouracil versus fluorouracil, doxorubicin, and methotrexate in advanced esophagogastric cancer," Journal of Clinical Oncology, vol. 15, no. 1, pp. 261-267, 1997.
21	B. Fisher, J. Dignam, E. P. Mamounas, J. P. Costantino, D. L. Wickerham, C. Redmond and R. G. Margolese, "Sequential methotrexate and fluorouracil for the treatment of node-negative breast cancer patients with estrogen receptor-negative tumors: eight-year results from National Surgical Adjuvant Breast and Bowel Project (NSABP) B-13 and first report of findings from NSABP B-19 comparing methotrexate and fluorouracil with conventional cyclophosphamide, methotrexate, and fluorouracil," Journal of Clinical Oncology, vol. 14, no. 7, pp. 1982-1992, 1996.
22	R. C. Coombes, J. M. Bliss, J. Wils, F. Morvan, M. Espie, D. Amadori and M. Marty, "Adjuvant cyclophosphamide, methotrexate, and fluorouracil versus fluorouracil, epirubicin, and cyclophosphamide chemotherapy in premenopausal women with axillary node-positive operable breast cancer: results of a randomized trial. The International Collaborative Cancer Group," Journal of clinical oncology, vol. 14, no. 1, pp. 35-45, 1996.

Continued on next page

Table 8.3 – continued from previous page

No	Reference
23	M. Colozza, A. Sidoni, A. M. Mosconi, A. Cavaliere, G. Bisagni, S. Gori and Italian Oncology Group for Clinical Research, "HER2 overexpression as a predictive marker in a randomized trial comparing adjuvant cyclophosphamide/methotrexate/5-fluorouracil with epirubicin in patients with stage I/II breast cancer: long-term results," Clinical breast cancer, vol. 6, no. 3, pp. 253-259, 2005.
24	G. Bonadonna, E. Brusamolino, P. Valagussa, A. Rossi, L. Brugnatelli, C. Brambilla and U. Veronesi, U, "Combination chemotherapy as an adjuvant treatment in operable breast cancer," New England Journal of Medicine, vol. 294, no. 8, pp. 405-410, 1976.
25	G. Bonadonna, P. Valagussa, A. Moliterni, M. Zambetti and C. Brambilla, "Adjuvant cyclophosphamide, methotrexate, and fluorouracil in node-positive breast cancer - the results of 20 years of follow-up," New England Journal of Medicine, vol. 332, no. 14, pp. 901-906, 1995.
26	F. JÃ€nicke, A. Prechtl, C. Thomssen, N. Harbeck, C. Meisner, M. Untch and German Chemo N0 Study Group, "Randomized adjuvant chemotherapy trial in high-risk, lymph node - negative breast cancer patients identified by urokinase-type plasminogen activator and plasminogen activator inhibitor type 1," Journal of the National Cancer Institute, vol. 93, no. 12, pp. 913-920, 2001.

Continued on next page

Table 8.3 – continued from previous page

No	Reference
27	L. Mauriac, M. Durand, J. Chauvergne, J. M. Diluydy and F. Bonichon, "Randomized trial of adjuvant chemotherapy for operable breast cancer comparing iv CMF to an epirubicin-containing regimen," Annals of oncology, vol. 3, no. 6, pp. 439-443, 1992.
28	S. Menard, P. Valagussa, S. Pilotti, L. Gianni, E. Biganzoli, P. Boracchi and G. Bonadonna, "Response to cyclophosphamide, methotrexate, and fluorouracil in lymph node - positive breast cancer according to HER2 overexpression and other tumor biologic variables," Journal of clinical oncology, vol. 19, no. 2, pp. 329-335, 2001.
29	J. M. Nabholtz, C. Falkson, D. Campos, J. Szanto, M. Martin, S. Chan and TAX 306 Study Group, "Docetaxel and doxorubicin compared with doxorubicin and cyclophosphamide as first-line chemotherapy for metastatic breast cancer: results of a randomized, multicenter, phase III trial," Journal of Clinical Oncology, vol. 21, no. 6, pp. 968-975, 2003.
30	M. Bontenbal, G. J. Creemers, H. J. Braun, A. C. de Boer, J. T. Janssen, R. B. Leys and C. Seynaeve, "Phase II to III study comparing doxorubicin and docetaxel with fluorouracil, doxorubicin, and cyclophosphamide as first-line chemotherapy in patients with metastatic breast cancer: results of a Dutch Community Setting Trial for the Clinical Trial Group of the Comprehensive Cancer Centre," Journal of clinical oncology, vol. 23, no. 28, pp. 7081-7088, 2005.

Continued on next page

Table 8.3 – continued from previous page

No	Reference
31	T. J. Evans, A. Yellowlees, E. Foster, H. Earl, D. A. Cameron, A. W. Hutcheon and J. L. Mansi, "Phase III randomized trial of doxorubicin and docetaxel versus doxorubicin and cyclophosphamide as primary medical therapy in women with breast cancer: an anglo-celtic cooperative oncology group study," Journal of Clinical Oncology, vol. 23, no. 13, pp. 2988-2995, 2005.
32	J. A. Sparano, A. N. Makhson, V. F. Semiglazov, S. A. Tjulandin, O. I. Balashova, I. N. Bondarenko and W. R. Rackoff, "Pegylated liposomal doxorubicin plus docetaxel significantly improves time to progression without additive cardiotoxicity compared with docetaxel monotherapy in patients with advanced breast cancer previously treated with neoadjuvant-adjuvant anthracycline therapy: results from a randomized phase III study," Journal of Clinical Oncology, vol. 27, no. 27, pp. 4522-4529, 2009.

8.5 Degree of Incompleteness for Experiment II

Table 8.4: Degree of incompleteness for experiment II

		Corresponding degree of incompletness				
Average Degree of Incompleteness		**10%**	**20%**	**30%**	**40%**	**50%**
Feature	Age	0.43	0.87	1.3	1.73	2.16
	Menopausal	10.53	21.06	31.59	42.12	52.65
	Race	9.12	18.23	27.35	36.47	45.58
	Estrogen Receptor	5.9	11.81	17.71	23.61	29.51
	Progesterone Receptor	8.63	17.26	25.89	34.52	43.15
	HER2NUE	13.72	27.44	41.16	54.87	68.59
	Tumor Stage	5.25	10.5	15.75	20.99	26.24
	Tumor Grade	12.5	25.01	37.51	50.02	62.52
	PLNC	8.29	16.57	24.86	33.14	41.43
	Lymph Node Status	10.17	20.35	30.52	40.7	50.87
	WHO Score	15.38	30.76	46.14	61.52	76.9
	Surgery Type	6.62	13.24	19.86	26.48	33.1
	Prior Radiotherapy	13.57	27.14	40.71	54.28	67.84
	Prior Chemotherapy	14.21	28.43	42.64	56.86	71.07
	Histology	15.67	31.35	47.02	62.7	78.37

References

[AFN15] Metin Akay, Dimitrios I Fotiadis, Konstantina S Nikita, and Robert W Williams. "Guest Editorial: Biomedical Informatics in Clinical Environments." *Biomedical and Health Informatics, IEEE Journal of*, **19**(1):149–150, 2015.

[AN07] Arthur Asuncion and David Newman. "UCI machine learning repository.", 2007.

[Ano12] PK Anooj. "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules." *Journal of King Saud University-Computer and Information Sciences*, **24**(1):27–40, 2012.

[BCP11] Ashly D Black, Josip Car, Claudia Pagliari, Chantelle Anandan, Kathrin Cresswell, Tomislav Bokun, Brian McKinstry, Rob Procter, Azeem Majeed, and Aziz Sheikh. "The impact of eHealth on the quality and safety of health care: a systematic overview." *PLoS Med*, **8**(1):e1000387, 2011.

[BDB11] Casey Bennett, Thomas Doub, April Bragg, Jason Luellen, Christina Van Regenmorter, Jennifer Lockman, and Randall Reiserer. "Data mining session-based patient reported outcomes (PROs) in a mental health setting: toward data-driven clinical decision support and personalized treatment." In *Healthcare Informatics, Imaging and Systems Biology (HISB), 2011 First IEEE International Conference on*, pp. 229–236. IEEE, 2011.

[Ber07] Eta S Berner. *Clinical decision support systems*. Springer, 2007.

[BGS04] R Boulme, D Gonzalez, and JC Schmit. "Storing genotypic resistance data and linking to other clinical information." In *Poster presented at the XV International AIDS Conference, Bangkok, Thailand*, 2004.

[BH13] Casey C Bennett and Kris Hauser. "Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach." *Artificial intelligence in medicine*, **57**(1):9–19, 2013.

[BHC10] Taxiarchis Botsis, Gunnar Hartvigsen, Fei Chen, and Chunhua Weng. "Secondary use of EHR: data quality issues and informatics opportunities." *AMIA Summits Transl Sci Proc*, **2010**:1–5, 2010.

[BL97] Avrim L Blum and Pat Langley. "Selection of relevant features and examples in machine learning." *Artificial intelligence*, **97**(1):245–271, 1997.

[BLS14] Ashwinkumar Badanidiyuru, John Langford, and Aleksandrs Slivkins. "Resourceful contextual bandits." *arXiv preprint arXiv:1402.6779*, 2014.

[CAT07] Emre Çomak, Ahmet Arslan, and İbrahim Türkoğlu. "A decision support system based on support vector machines for diagnosis of the heart valve diseases." *Computers in Biology and Medicine*, **37**(1):21–27, 2007.

[CJK12] Sang-Hoon Cho, Jongsu Jeon, and Seung Il Kim. "Personalized medicine in breast cancer: a systematic review." *Journal of breast cancer*, **15**(3):265–272, 2012.

[Coh60] J Cohen. "A coefficient of agreement for nominal scales. Educational and Psychosocial Measurement, 20, 37-46.", 1960.

[DJ12] Nassim Douali and M Jaulent. "Genomic and personalized medicine decision support system." In *Complex Systems (ICCS), 2012 International Conference on*, pp. 1–4. IEEE, 2012.

[DSB15] Klaus Donsa, Stephan Spat, Peter Beck, Thomas R Pieber, and Andreas Holzinger. "Towards personalization of diabetes therapy using computerized decision support and machine learning: some open problems and challenges." In *Smart Health*, pp. 237–260. Springer, 2015.

[EKZ08] Jae-Hong Eom, Sung-Chun Kim, and Byoung-Tak Zhang. "AptaCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction." *Expert Systems with Applications*, **34**(4):2465–2479, 2008.

[FGR15] F. M. Fleegler, J. Griggs, B. Reiner, B. Reville, S. F. Schnall, M. C. Weiss, and L. Weissmann. "Chemotherapy Medicine.", 2015.

[Fre15] M. Frellick. "Landmark Report Urges Reform to Avert Diagnostic Errors.", 2015.

[GCT04] Mary K Goldstein, Robert W Coleman, Samson W Tu, Ravi D Shankar, Martin J O'Connor, Mark A Musen, Susana B Martins, Philip W Lavori, Michael G Shlipak, Eugene Oddone, et al. "Translating research into practice: organizational issues in implementing automated decision support for hypertension in three medical centers." *Journal of the American Medical Informatics Association*, **11**(5):368–376, 2004.

[GOS08] Lori J Goldstein, Anne O'Neill, Joseph A Sparano, Edith A Perez, Lawrence N Shulman, Silvana Martino, and Nancy E Davidson. "Concurrent doxorubicin plus docetaxel is not more effective than concurrent

doxorubicin plus cyclophosphamide in operable breast cancer with 0 to 3 positive axillary nodes: North American Breast Cancer Intergroup Trial E 2197." *Journal of Clinical Oncology*, **26**(25):4092–4099, 2008.

[Hal99] Mark A Hall. *Correlation-based feature selection for machine learning*. PhD thesis, The University of Waikato, 1999.

[HC10] Margaret A Hamburg and Francis S Collins. "The path to personalized medicine." *New England Journal of Medicine*, **363**(4):301–304, 2010.

[HJ15] P. J. Hamel and P. Johnson. "Chemotherapy Regimen for Breast Cancer.", 2015.

[HMN05] Peter Hall, JS Marron, and Amnon Neeman. "Geometric representation of high dimension, low sample size data." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**(3):427–444, 2005.

[JSV11] Monique WM Jaspers, Marian Smeulers, Hester Vermeulen, and Linda W Peute. "Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings." *Journal of the American Medical Informatics Association*, **18**(3):327–334, 2011.

[KDE15] Finn Kuusisto, Inês Dutra, Mai Elezaby, Eneida A Mendonça, Jude Shavlik, and Elizabeth S Burnside. "Leveraging Expert Knowledge to Improve Machine-Learned Decision Support Systems." *AMIA Summits on Translational Science Proceedings*, **2015**:87, 2015.

[KHB05] Kensaku Kawamoto, Caitlin A Houlihan, E Andrew Balas, and David F Lobach. "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success." *Bmj*, **330**(7494):765, 2005.

[KSE12] Evaggelos C Karvounis, Kostas Stefanou, Themis P Exarchos, Alexandros T Tzallas, Markos Tsipouras, and Dimitrios I Fotiadis. "A treatment decision support system for patients receiving Ventricular Assist Device (VAD) therapy." In *Biomedical and Health Informatics (BHI), 2012 IEEE-EMBS International Conference on*, pp. 695–698. IEEE, 2012.

[KTM10] Ioannis Kouris, Charalampos Tsirmpas, Stavroula G Mougiakakou, Dimitra Iliopoulou, and Dimitris Koutsouris. "E-Health towards ecumenical framework for personalized medicine via Decision Support System." In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pp. 2881–2885. IEEE, 2010.

[LDC12] Roderick J Little, Ralph D'Agostino, Michael L Cohen, Kay Dickersin, Scott S Emerson, John T Farrar, Constantine Frangakis, Joseph W Hogan, Geert Molenberghs, Susan A Murphy, et al. "The prevention and treatment of missing data in clinical trials." *New England Journal of Medicine*, **367**(14):1355–1360, 2012.

[Lev01] Mark Levine et al. "Clinical practice guidelines for the care and treatment of breast cancer: adjuvant systemic therapy for node-positive breast cancer (summary of the 2001 update)." *Canadian Medical Association Journal*, **164**(5):644–646, 2001.

[LPP14] Jie Liu, David Page, Peggy Peissig, Catherine McCarty, Adedayo A Onitilo, Amy Trentham-Dietz, and Elizabeth Burnside. "New genetic variants improve personalized breast cancer diagnosis." *AMIA Summits on Translational Science Proceedings*, **2014**:83, 2014.

[MKL14] Lorenzo Moja, Koren H Kwag, Theodore Lytras, Lorenzo Bertizzolo, Linn Brandt, Valentina Pecoraro, Giulio Rigon, Alberto Vaona, Francesca Ruggiero, Massimo Mangia, et al. "Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis." *American journal of public health*, **104**(12):e12–e22, 2014.

[MPM15] Sabrina Molinaro, Stefania Pieroni, Fabio Mariani, and Michael N Liebman. "Personalized medicine: Moving from correlation to causality in breast cancer." *New Horizons in Translational Medicine*, **2**(2):59, 2015.

[MSA14] Mary McNamara, Karthik Sarma, Denise R Aberle, Alex AT Bui, and Corey Arnold. "Data Model for Personalized Patient Health Guidelines: An Exploratory Study." In *AMIA Annual Symposium Proceedings*, volume 2014, p. 1835. American Medical Informatics Association, 2014.

[MWC05] Randolph A Miller, Lemuel R Waitman, Sutin Chen, and S Trent Rosenbloom. "The anatomy of decision support during inpatient care provider order entry (CPOE): empirical observations from a decade of CPOE experience at Vanderbilt." *Journal of biomedical informatics*, **38**(6):469–485, 2005.

[MWL12] Allison B McCoy, Lemuel R Waitman, Julia B Lewis, Julie A Wright, David P Choma, Randolph A Miller, and Josh F Peterson. "A framework for evaluating the appropriateness of clinical decision support alerts and responses." *Journal of the American Medical Informatics Association*, **19**(3):346–352, 2012.

[PDK13] Thomas A Peterson, Emily Doughty, and Maricel G Kann. "Towards precision medicine: advances in computational approaches for the anal-

ysis of human variants." *Journal of molecular biology*, **425**(21):4047–4063, 2013.

[Per01] K Person. "On Lines and Planes of Closest Fit to System of Points in Space. Philiosophical Magazine, 2, 559-572.", 1901.

[PG07a] Kemal Polat and Salih Güneş. "Breast cancer diagnosis using least square support vector machine." *Digital Signal Processing*, **17**(4):694–701, 2007.

[PG07b] Kemal Polat and Salih Güneş. "Breast cancer diagnosis using least square support vector machine." *Digital Signal Processing*, **17**(4):694–701, 2007.

[PLD05] Hanchuan Peng, Fuhui Long, and Chris Ding. "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy." *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **27**(8):1226–1238, 2005.

[RHW98] Robert D Rosenberg, William C Hunt, Michael R Williamson, Frank D Gilliland, Philip W Wiest, Charles A Kelsey, Charles R Key, and Michael N Linver. "Effects of age, breast density, ethnicity, and estrogen replacement therapy on screening mammographic sensitivity and cancer stage at diagnosis: review of 183,134 screening mammograms in Albuquerque, New Mexico." *Radiology*, **209**(2):511–518, 1998.

[Ric08] Terri D Richmond. "The current status and future potential of personalized diagnostics: Streamlining a customized process." *Biotechnology annual review*, **14**:411–422, 2008.

[RZW15] Sudha Ram, Wenli Zhang, Max Williams, and Yolande Pengetnze. "Predicting asthma-related emergency department visits using big data." *Biomedical and Health Informatics, IEEE Journal of*, **19**(4):1216–1223, 2015.

[SC13] Edward H Shortliffe and James J Cimino. *Biomedical informatics: computer applications in health care and biomedicine*. Springer Science & Business Media, 2013.

[SM06] Hyunjin Shin and Mia K Markey. "A machine learning perspective on the development of clinical decision support systems utilizing mass spectra of blood samples." *Journal of Biomedical Informatics*, **39**(2):227–248, 2006.

[Soc15] American Cancer Society. "American Cancer Society Guidelines for the Early Detection of Cancer.", 2015.

[TS15] Cem Tekin and Mihaela van der Schaar. "RELEAF: An algorithm for learning and exploiting relevance." *Selected Topics in Signal Processing, IEEE Journal of*, **9**(4):716–727, 2015.

[Tsu98] Shusaku Tsumoto. "Automated extraction of medical expert system rules from clinical databases based on rough set theory." *Information sciences*, **112**(1):67–84, 1998.

[TV14] Cem Tekin and Mihaela Van Der Schaar. "Discovering, learning and exploiting relevance." In *Advances in Neural Information Processing Systems*, pp. 1233–1241, 2014.

[TVD14] R Tsopra, A Venot, and C Duclos. "An Algorithm Using Twelve Properties of Antibiotics to Find the Recommended Antibiotics, as in CPGs." In *AMIA Annual Symposium Proceedings*, volume 2014, p. 1115. American Medical Informatics Association, 2014.

[VHH15] Marco Viceconti, Peter Hunter, and Rod Hose. "Big Data, Big Knowledge: Big Data for Personalized Healthcare." *Biomedical and Health Informatics, IEEE Journal of*, **19**(4):1209–1215, 2015.

[WLP14] Yirong Wu, Jie Liu, David Page, Peggy Peissig, Catherine McCarty, Adedayo A Onitilo, and Elizabeth S Burnside. "Comparing the value of mammographic features and genetic variants in breast cancer risk prediction." In *AMIA Annual Symposium Proceedings*, volume 2014, p. 1228. American Medical Informatics Association, 2014.

[WMC00] Jason Weston, Sayan Mukherjee, Olivier Chapelle, Massimiliano Pontil, Tomaso Poggio, and Vladimir Vapnik. "Feature selection for SVMs." In *NIPS*, volume 12, pp. 668–674. Citeseer, 2000.

[XST14] Jie Xu, Daby Sow, Deepak Turaga, and Mihaela van der Schaar. "Online Transfer Learning for Differential Diagnosis Determination." In *AAAI Workshop on the World Wide Web and Public Health Intelligence*, 2014.

[ZN15] Konstantia Zarkogianni and Konstantina S Nikita. "Personal health systems for diabetes management, early diagnosis and prevention." *Handbook of Research on Trends in the Diagnosis and Treatment of Chronic Conditions*, p. 465, 2015.

[ZWH14] Ping Zhang, Fei Wang, Jianying Hu, and Robert Sorrentino. "Towards personalized medicine: leveraging patient similarity and drug similarity analytics." *AMIA Summits on Translational Science Proceedings*, **2014**:132, 2014.