Artificial Intelligence In Medicine 101 (2019) 101750

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

A predictive analytics framework for identifying patients at risk of developing multiple medical complications caused by chronic diseases

Amir Talaei-Khoei[a,b], Madjid Tavana[c,d,*], James M. Wilson[e]

[a] Department of Information Systems, University of Nevada, Reno, USA
[b] School of Software, University of Technology Sydney, Australia
[c] Business Systems and Analytics Department, Distinguished Chair of Business Analytics, La Salle University, Philadelphia, USA
[d] Business Information Systems Department, Faculty of Business Administration and Economics, University of Paderborn, Paderborn, Germany
[e] School of Community Health Sciences, University of Nevada, Reno, USA

ARTICLE INFO

Keywords:
Predictive analytics
Chronic disease
Artificial neural networks
Multi-Task learning
Regression.

ABSTRACT

Chronic diseases often cause several medical complications. This paper aims to predict multiple complications among patients with a chronic disease. The literature uses single-task learning algorithms to predict complications independently and assumes no correlation among complications of chronic diseases. We propose two methods (*independent prediction of complications with single-task learning* and *concurrent prediction of complications with multi-task learning*) and show that medical complications of chronic diseases can be correlated. We use a case study and compare the performance of these two methods by predicting complications of hypertrophic cardiomyopathy on 106 predictors in 1078 electronic medical records from April 2009-April 2017, inclusive. The methods are implemented using logistic regression, artificial neural networks, decision trees, and support vector machines. The results show multi-task learning with logistic regression improves the performance of predictions in terms of both discrimination and calibration.

1. Introduction

There is a body of literature in healthcare analytics that looks at the patterns in electronic medical records to predict medical outcomes [1–10]. Recent developments revealed that chronic diseases cause several medical complications [11]. For example, diabetes type 2 increases the risk for medical complications such as numbness in feet, kidney conditions, high blood pressure, loss of vision and stroke. The advancements in predictive analytics are now being used for data mining of electronic medical records in order to predict patients at risk of developing the complications caused by a chronic disease [12–15,16,17–21]. Most of the literature in this area predicts multiple medical complications of chronic diseases by independently predicting each complication. For instance, Choi et al. [13] use machine learning methods to predict four medical complications caused by diabetes type 2. In this paper, each of these four complications were predicted separately, regardless of their inter-relationships. Similarly, Sangi et al. [22] utilize different predictive models to predict the development of T2D complications among diabetes patients. Sangi et al. [22] assumes the complications have no interrelationships.

Despite the approach taken in the healthcare predictive analytics literature to ignore the relationship among multiple complications caused by chronic diseases, the literature noted the dependencies among complications of chronic diseases [11]. For example, Maron [23] systematically reviews the complications for hypertrophic cardiomyopathy and found many of these complications are related to each other. For instance, in order to identify diabetes patients at risk of vision loss, Piri et al. [8] consider the co-existence of other diabetes-related complications as predictors and implement single-task learning. However, in practice, the complications of chronic diseases may develop concurrently, and at the time of prediction analysis, we may not necessarily have information about a particular complication to set it as a predictor. This paper aims to predict multiple complications caused by chronic diseases when they are interrelated. This study extends the literature in this area by setting all multiple complications of a chronic disease that are in the scope of analysis as predicted variables and not predictors.

Despite hypothesis-testing methods that investigate the effect of a variable or a group of variables on developing complications of chronic diseases, the machine learning algorithms extract non-trial patterns in

* Corresponding author at: Business Systems and Analytics Department, Distinguished Chair of Business Analytics, La Salle University, Philadelphia, PA 19141, United States.
E-mail addresses: atalaeikhoei@unr.edu (A. Talaei-Khoei), tavana@lasalle.edu (M. Tavana), jamesmwilson@unr.edu (J.M. Wilson).

https://doi.org/10.1016/j.artmed.2019.101750
Received 14 July 2018; Received in revised form 7 July 2019; Accepted 30 October 2019

medical datasets or electronic medical records. Furthermore, the machine learning algorithms classify observations [24]. Therefore, the use of machine learning minimizes the insufficiency of using hypothesis testing, which requires a hypothesis based on previous literature [5]. Machine learning algorithms such as logistic regression (LR), neural networks, Support Vector Machine (SVM), and decision trees (DT) have been used extensively in healthcare [19,25–32]. The applications of machine learning in healthcare include, but are not limited to, predicting hospital readmissions [2,33], predicting survival of a medical procedure [5], early detection of chronic diseases [7,13,14,34], and predicting survival of chronic disease [35,36]. Another growing application of the machine learning algorithms is predicting the complications caused by chronic diseases. For instance, machine learning is used to predict vision loss caused by diabetes [8,37]; Kothari et al. [38] predicts heart disease and stroke in patients with type 2 diabetes; Pahl et al. [39] utilizes machine learning to predict heart failure resulting in death among children with cardiomyopathy.

As mentioned earlier, the literature on using machine learning to predict complications of chronic diseases only focuses on one specific complication and develops predictive models that can capture the occurrence of complications. This contrasts with the fact that among patients with chronic diseases, multiple complications are commonly observed and these medical complications are proven to be interrelated [11]. The studies in this area predict multiple clinical variables independently and assume that there is no relationship among different clinical variables e.g. [13]. However, the complications of chronic diseases are often related. For example, heart failure, problems with heart valves, cardiac arrest and sudden death are complications associated with hypertrophic cardiomyopathy [40], which have been shown to be interrelated [23]. Recent work by Piri et al. [8] tried to address this issue when predicting vision loss among diabetes patients by setting related complications as predictors. However, the complications may develop concurrently and the patient may not have these complications, but may be at risk of developing them. Therefore, the complications must be considered prediction outcomes.

Despite the literature that uses single-task learning (STL) to independently predict patients at risk of developing multiple medical complications of a chronic disease, this study aims to use multi-task learning (MTL) [41] by taking the dependencies of complications into account. MTL has been widely used in the literature of healthcare analytics [42,43] to identify patients at risk of developing a condition when the risk factors are related. For instance, MTL has been used to predict the progression of Alzheimer's disease based on interacting clinical variables [44,45]; Razavian et al. [46] utilizes MTL to predict chronic kidney disease progression. Unfortunately, these studies use MTL in order to predict one particular medical condition such as Alzheimer's or chronic kidney disease. To the best of our knowledge, the development of MLT-based models to predict multiple medical complications of a chronic condition has not yet been addressed. As such, this paper is an attempt to answer the following research question (RQ):

> Does concurrent modeling of medical complications through the utilization of their intrinsic correlation improve the predication of said complications when compared with the independent prediction of multiple complications for a chronic disease?

The main difference between the current work and the literature in healthcare analytics is related to the utilization of MTL for the prediction of multiple complications of chronic diseases. Despite the existing research, which considers only a single medical complication or does not consider intrinsic correlations among complications, the current study extends the said literature by providing an MTL-based model that predicts multiple complications of chronic diseases through the utilization of the dependencies in relation to the complications. The proposed models are compared when predicting heart failure, problems with heart valves, cardiac arrest and sudden death as commonly experienced complications of hypertrophic cardiomyopathy.

The remainder of this paper is organized as follows. Section 2 presents the method suggested in this study; section 3 explains the evaluation of the proposed method; section 4 presents the analysis of the evaluation results; and section 5 discusses the findings and concludes the paper.

2. Proposed methods

Our objective is to identify patients with a particular chronic disease, hereafter to be referred to as "patients", who are at risk of developing multiple medical complications related to their chronic disease. In this section, we first set our definition and build our analysis framework, then we propose two methods: (1) independent prediction of multiple complications that assumes no correlation among different complications and uses STL, and (2) concurrent prediction of multiple complications that relaxes the assumption and uses MTL.

2.1. Analysis framework

If we have electronic medical records (EMRs) for n different patients, we define the point of time t, $0 \le t$ for the electronic medical records of the patient p, $0 \le p \le n$. If there exists m medical complications for the chronic disease under the study, we are interested in predicting whether the patient p is going to develop the complication $0 \le comp < m$ in $0 \le y$ years. Therefore, we predict for the y th year when the time is $t_{predict} = t + y$. We define the vector of a predictors of patient p as $predictor_p = [predictor_{p,1}, predictor_{p,2}, ..., predictor_{p,a}]^\intercal$. The predictors' vector presents the EMR items at the given time t. Whether the patient p is going to develop the complication *comp* in y years or not is defined as $dev_p^{comp,y} \in \{0, 1\}$. The expected value of $dev_i^{comp,y}$ is a function of the probability of the occurrence of the complication *comp* in y years for the patient p, which can be represented by $\rho_p^{comp,y}$. Therefore, $dev_p^{comp,y} = 1$, if patient p develops complication *comp* in y years, otherwise $dev_p^{comp,y} = 0$. The probability distribution of $dev_p^{comp,y} \in \{0, 1\}$ as a Boolean function can be represented by a Bernoulli distribution that has two possible outcomes labelled by 0 and 1. Therefore, the probability distribution function (PDF) of $dev_p^{comp,y}$ can be defined as Eq. (1). The Bernoulli distribution has been widely used in the literature such as [47,48] for representing Boolean functions.

$$PDF(dev_p^{comp,y}) = \begin{cases} \rho_p^{comp,y}, & dev_p^{comp,y} = 1 \\ \rho_p^{comp,y}, & dev_p^{comp,y} = 0 \end{cases} \tag{1}$$

Having established the analysis framework, we then propose two approaches: STL and MTL. Although the proposed methods are flexible enough to be used by the majority of predictive models, the description below is based on logistic regression as the predictive model used in these approaches. In Section 2.4, we discuss the generalizability of these proposed methods to be used with alternative predictive models such as DT, support vector machine (SVM) and/or artificial neural network (ANN).

2.2. STL method: independent prediction of multiple complications (IPMC)

As the name suggests, this method assumes that there is no association among multiple future complications within the study and predicts each complication independently from the others.

Assuming that we have chosen logistic regression as our predictive model, we deploy Eq. (2) to model the association of $\rho_p^{comp,y}$ and $predictor_p$ for m complications.

$\forall\ 0 \le comp < m$:

$$Logit(\rho_p^{comp,y}) = \alpha^{comp,y} + \sum_{i=1}^{a} \beta_i^{comp,y} predictor_p \tag{2}$$

In the above equation, $\alpha^{comp,y}$ is an intercept and $\beta_i^{comp,y}$ is a

coefficient; both are considered specific to a particular complication. In the prediction modeled above, whether a patient develops a complication *comp* out of m complications in y years is based on $\rho_p^{comp,y}$ and is independent of other complications. Therefore, the learning practice of $Logit(\rho_p^{comp,y})$ will be conducted independently for each complication.

2.3. MTL method: concurrent prediction of multiple complications (CPMC)

The approach taken in IPMC is to assume that each future medical complication of a chronic disease will be independent. Given the interrelated nature of multiple complications [11], a real world prediction of future complications for chronic diseases must include multiple correlated tasks. Therefore, we deploy MTL in order to more efficiently learn several related complications concurrently as an inductive bias [49].

There are multiple learning tasks in MTL, a subset of which might be interrelated. For example, in prediction of patients at risk of developing complications of chronic diseases, these complications are assumed to be interrelated [11]. Learning multiple related tasks concurrently has been empirically [50,51] as well as theoretically [52] shown to often significantly improve performance relative to learning each task independently. Therefore, MTL aims to improve the performance of multiple tasks when they are interrelated. MTL improves the performance by learning tasks concurrently while using a shared representation; what is learned for each task can help other tasks be learned better. By using MTL, we increase performance by forcing the model to learn a more generalized representation as it learns not just for one specific task but for all the tasks.

Therefore, as opposed to IPMC, which uses STL, our proposed method is called Concurrent Prediction of Multiple Complications (CPMC) and uses MTL. MTL has demonstrated success in a broad spectrum of applications designed to more efficiently learn in a shared information space [53–55]. In this study, the shared information space is $predictor_p = [predictor_{p,1}, predictor_{p,2}, ..., predictor_{p,a}]^\tau$. In MTL, as opposed to STL, the learner can be trained in an environment of related learning tasks. Within such an environment, the learner, by unifying the training process, can sample from multiple tasks and search for a model that would have higher generalizability in comparison to multiple models generated in the STL process [56,57].

Within this study, let tasks be defined as the prediction of multiple medical complications related to the chronic disease under the study. In order to deploy MTL using logistic regression, we reconstruct Eq. (2) where $\rho_p^{comp,y}$ (i.e. the probability of the patient p develops the complication *comp* among m complications in y years) can be defined by intercepts i.e. $\alpha^{comp,y}$ and coefficients i.e. $\beta_i^{comp,y}$. Therefore, we have a set of m logistic regression models that predict complications$0 \le comp < m$.

Let all tasks (i.e. predictions of multiple complications) have the same a predictors, i.e. $predictor_p = [predictor_{p,1}, predictor_{p,2}, ..., predictor_{p,a}]^\tau$ that can be extracted from EMRs. Following MTL, to model the dependency of m complications, the correlation of coefficients for the i th predictor $predictor_{p,i}$ is considered across different predictions of multiple complications. To achieve this, Table 1 demonstrates the correlation matrix for the ith predictor $corMatrix_i$. The matrix has been built for each of the regression coefficients corresponding to a particular predictor. For instance, the regression coefficient for the ith predictor to predict in y year is modelled as the vector $\beta_i^y = [\beta_i^{1,y}, \beta_i^{2,y}, ..., \beta_i^{m,y}]^\tau$ and will be used for all patients p, $0 \le p \le n$. In the correlation matrix $corMatrix_i$, the regression coefficients for the ith predictor to predict in y year (β_i^y) includes information about the all complications. Therefore, the training of β_i^y has information about the complication *comp* as well as information about the other complications. Through this, we can address the dependencies among multiple, different complications.

Table 1
Building correlation matrix.

Predictions for Different Complications	β_1^y	β_2^y	...	β_i^y
Prediction of Complication 1 in y years.	$\beta_1^{1,y}$	$\beta_2^{1,y}$	...	$\beta_i^{1,y}$
Prediction of Complication 2 in y years.	$\beta_1^{2,y}$	$\beta_2^{2,y}$	...	$\beta_i^{2,y}$
...	...	...		...
Prediction of Complication m in y years.	$\beta_1^{m,y}$	$\beta_2^{m,y}$	...	$\beta_i^{m,y}$

Following the above described approach to implement MTL, Eq. (3) updates the logistic regression model of IPMC. As discussed earlier, β_i^y includes information about the complication *comp* as well as other complications. Therefore, despite Eq. (2), the regression model presented in Eq. (3) includes correlations among complications.

$\forall\ 0 \le comp < m$:

$$Logit(\rho_p^{comp,y}) = \alpha^{comp,y} + \sum_{i=1}^{a} \beta_i^y predictor_p \tag{3}$$

We first set all non-binary $predictor_{p,i}$ to have the mean of zero and standard deviation of 0.5 and then place weakly, informative prior distributions – explained in [58] – for the model variables. Table 2 presents the operationalization of all model variables in CPMC.

2.4. Generalizability

IPMC and CPMC were demonstrated using the implantation of logistic regression. In following, we discuss how these two approaches can be implemented using alternative predictive models.

The construction of IPMC followed the usual STL approach using predictive models and did not commit to any assumption; therefore, IPMC can be easily implemented by any predictive models such as DT, SVM and ANN. However, in the construction of CPMC, we enabled MTL by deploying a hierarchical correlation structure as a channel to transfer information across predictions of different complications. Although we used BLR to implement MTL, we made no assumption on the predictive model except that the model requires a multivariate Gaussian distribution. Most predictive models have been implemented in the literature in their multivariate Gaussian distribution representations, such as DT [66], SVM [67] and ANN [68].

Based on the above discussion, both IPMC and CPMC are generalizable in terms of using different predictive models. However, for CPMC, the predictive model needs to have a multivariate Gaussian distribution.

3. Evaluation

This section explains how IPMC and CPMC were evaluated and compared to each other in terms of their performance for predicting medical complications of chronic diseases. IPMC and CPMC are general methods independent of what chronic disease is under study. However, for the purpose of evaluation, the present study compares them in the context of hypertrophic cardiomyopathy and its common complications: heart failure, problems with heart valves, and cardiac arrest [40]. Fig. 1 presents the evaluation methodology.

IPMC and CPMC are evaluated to predict the complications of hypertrophic cardiomyopathy. Cardiomyopathy is used for a range of chronic cardiovascular conditions for heart muscles. While there are several types of cardiomyopathy, the most common type is hypertrophic cardiomyopathy [23]. Hypertrophic cardiomyopathy causes several serious complications. The present study focuses only on the three most commonly experienced complications associated with hypertrophic cardiomyopathy: heart failure, problems with heart valves, and cardiac arrest [40], which are proven to be correlated [23]. Therefore, this study can evaluate the effect of MTL by comparing the performance of IPMC and CPMC in predicting hypertrophic cardiomyopathy patients at risk of developing any of these three

Table 2
Variables and their operationalization.

Model Variable	Operationalization	Description
Coefficients: β_i^y	Across different m complications, the regression coefficients for the ith predictor are represented by $\beta_i^y = [\beta_i^{1,y}, \beta_i^{2,y}, ..., \beta_i^{m,y}]^\tau$. Following the recommendations of Jammalamadaka et al. [59] and Melie-Garcia et al. [60], in order to model the association of multiple m complications, β_i^y is designed to follow a multivariate Gaussian distribution with mean of $\mu = 0$ and covariance matrix $\Sigma = r_i^2 \Sigma_i$. We define the shrinkage scalar r_i and the original covariance matrix Σ_i later. $\beta_i^y = [\beta_i^{1,y}, \beta_i^{2,y}, ..., \beta_i^{m,y}]^\tau$ is said to follow a multivariate Gaussian distribution with mean of μ and covariance matrix of Σ if β_i^y can be expressed as $\beta_i^y = AZ + \mu$ where $\Sigma = AA^\tau$ (i.e. the modeling association of multiple complications) and $Z = [Z_1, Z_2..., Z_m]$ is a random vector $N(0,1)$. We let the mean of $\mu = 0$ which implies $\beta_i^y = AZ$. Therefore, the mean of $\mu = 0$ demonstrates no prior knowledge about the effect of predictors on the associations of multiple complications. To model the association of multiple complications, we incorporate the effect of $\beta_i^{1,y}, \beta_i^{2,y}, ...$ on $\beta_i^{comp,y}$ into a covariance matrix as a scaled covariance matrix $\Sigma = AA^\tau = r_i^2 \Sigma_i$.	To predict the y^{th} year, for each predictor, a coefficient is defined. For instance, β_i^y is defined as the coefficient of the i^{th} predictor to predict for the y^{th} year. In multitask learning, because the coefficients may be correlated with each other, $\beta_i^y = [\beta_i^{1,y}, \beta_i^{2,y}, ..., \beta_i^{m,y}]^\tau$ is defined to follow the multivariate Gaussian distribution.
Intercepts: $\propto^{comp,y}$	The model variable $\propto^{comp,y}$ follows the default prior of the Lorentz distribution ($x_0 = 0,\ \gamma = 10$) suggested by Gelman et al. [58]. The setting improves the robustness of the model by accepting $\frac{1}{10^9}$ for the very poor chance for chronic patients developing the complication and 1- $\frac{1}{10^9}$ for the opposite.	One intercept is defined for each complication to be predicted in the y^{th} year. The Lorentz distribution is selected for intercepts. The distribution has the center of zero ($x_0 = 0$) and scale of ten ($\gamma = 10$). It follows a bell-shape density function with a longer tail than the normal distribution. This allows both very likely and very unlikely chances of chronic patients developing the complication.
Correlation Matrix: $corMatrix_i$	The success of MTL is about how well we can model the correlation among coefficients. One issue in doing so is that the off-diagonal cannot guaranteed to be positive semidefinite [61]. Therefore, as suggested by the literature [62], $corMatrix_i$ follow the distribution proposed by Lewandowski et al. [63] with two parameters e.g. the first parameter is the number of tasks with m complications and the second parameter is the degree that we want the correlation matrix to shrink toward the identity matrix. As in this study, we have no prior information about the correlation of coefficients, we choose 1 for the second parameter of this distribution, which shows that the prior density for correlation matrices is uniform.	Correlation matrix refers to Table 1. To develop such a matrix, the challenge is that the value of diagonal elements would always be 1. Therefore, we use off-diagonal elements, which cannot guarantee to be positive semidefinite. Therefore, we generate a random matrix that follows Lewandowski's distribution. In order to design our correlation matrix with the Lewandowski's distribution, we chose m complications as the number of tasks and assume 1 for the correlation of coefficients as no prior information is available. For each predictor, one correlation matrix is defined.
Shrinkage Scalar: r_i	EMRs include a large number of variables that can be potentially used for prediction of chronic complications. Shrinkage is a solution to this dilemma to choose between a pragmatic model with a better performance and a more complex one fitting the data [64]. De Mol et al. [65] suggests using $r_i = \omega \times \varphi_i$ where ω is indicator of all predictors and φ_i is the indicator of $predictor_{p,i}$. These two indicators of ω and φ_i follow the truncated Lorentz distribution ($x_0 = 0,\ \gamma = 10$) suggested by Gelman et al. [58]. This makes it robust with densities only on positive real numbers but also with thicker tails than the normal distribution to be adaptive to large values.	In order to select the predictors, we define shrinkage for each EMR variable that generates more efficient model. Similar to coefficients, Lorentz distribution ($x_0 = 0,\ \gamma = 10$) is used to select the predictors. Therefore, for each predictor, there exists a shrinkage.
Covariance Matrix: Σ_i	$\sum_i = diagnalmatrix(sd_i) \times corMatrix_i \times diagnalmatrix(sd_i)$. ([59]; [60]).	One covariance matrix is defined for each predictor. Covariance matrix is a matrix whose element is the covariance with the elements of a random vector. In this study, the multivariate Gaussian distribution of the j^{th} coefficients across tasks is selected for the random vector. The decomposition structure suggested by literature ([59]; [60]) an used in this study as $\Sigma_i = diagnal\ matrix\ (sd_i) \times corMatrix_i \times diagnal\ matrix\ (sd_i)$ allows to generate the Covariance Matrix using correlation matrix. This makes it possible to model MTL for multiple complications of chronic diseases.
Standard Deviation: sd_i	The standard deviation of ith predictor $predictor_{p,i}$ for the complication $comp$ among m complications is sd_i^{comp}. According to the recommendation of Gelman et al. [58], sd_i^{comp} follows the the truncated Lorentz distribution ($x_0 = 0,\ \gamma = 2.5$). sd_i is defined as a vector of the standard deviations of $predictor_{p,i}$ for all m complications i.e. $sd_i = [sd_i^1, sd_i^2, ..., sd_i^m]$.	In order to predict each complication, for each predictor, one standard deviation is defined.

complications.

To evaluate different dimensions of the present study, two comparisons were conducted.

- *IPMC versus CPMC*: First, the performances of IPMC and CPMC were compared through the implementation of different predictive models; namely Bayesian logistic regression (BLR), DT, SVM, and ANN. These comparisons were a head-to-head comparison; for example, BLR implementations of IPMC and CPMC were compared with each other and similarly for LR, SVM, DT and ANN.
- *Implementation of CPMC using predictive models BLR, DT, SVM and ANN:* Second, different implementations of CPMC were compared with each other to see which predictive model generated the best performance for CPMC. The first and the second evaluations were independent of the period of prediction; meaning that the comparisons were repeated for 1-year, 3-year and 8-year predictions to determine if the period of prediction would impact the results of the above-mentioned comparisons.

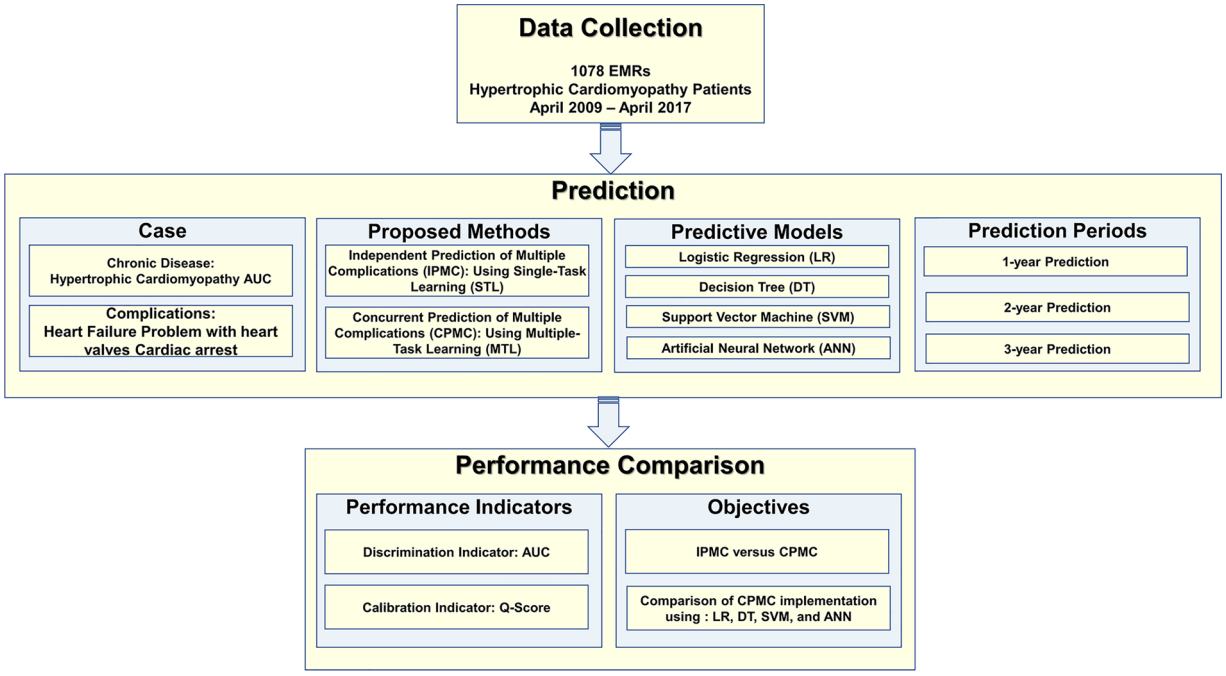

Fig. 1. Proposed methodology.

3.1. Evaluation site and data collection

In 2015, a system was developed to continuously analyze daily-generated EMRs for improving practices in chronic cardiovascular diseases for patients across twenty-four different Cardiology Clinics in Australia. Within these twenty-four clinics, an integration infrastructure was implemented that linked the daily generated EMRs of the cardiac patients. The dataset used in this study is the un-linked EMRs from these twenty-four practices. The data has been de-identified by the system while linking EMRs.

In order to clean the data, we implemented the following five steps. First, we removed all the intra- and post- factors because our study aims to predict the development of complications of hypertrophic cardiomyopathy. We adopted outlier detection algorithms in STATISTICA II to eliminate entry errors and duplicated data. In the third step, we removed all the variables that cannot be used in the predication such as the system-generated ID for patients. In the fourth step, we eliminated variables that, according to our preliminary analysis, showed no contribution towards prediction. These variables are Occupation, Education, City of birth (not Ethnicity), City of Address, and Marital Status. Finally, we removed records with missing values. There is a wide range of methods dealing with missing values in electronic medical records. A review of these methods are presented in Liu et al. [69]. The imputation methods, where missing values are replaced with substitute values, have shown improvement in the population-based model prediction. For example, García-Laencina et al. [70] use imputation methods on the entire dataset to handle missing values. Shukla et al. [71] Propose an approach that can analyze survivability of breast cancer patients in presence of missing data. Imputation methods are not suitable for the aim of this study. However, when an analysis offers a method for prediction of specific conditions in individual patients rather than a population of patients, the imputation methods can mislead the prediction for individual patients using assumed values. However, performance of this algorithm depends on the number of predictors under the study [72]. As in this study we are testing 107 EMR variables, this approach would need an extensive computational power. In such situations, removal of missing records to avoid the bias of replacing them with substitute values are recommended [73]. Because our study aims at individual patients rather than population-based models, we opted to not include records with missing values in our analysis. Our initial results showed that removing records with missing values do not influence the accuracy of our predictions.

We define a base time of $t = 0$ on the year each patient became a hypertrophic cardiomyopathy patient from April 2009 to April 2017. We only collected patient record $EMR_{t=0}$ for that year, and then collect complications test results for 8 years after the base year. For example, if a patient becomes a hypertrophic cardiomyopathy patient in year 2009, then 2009 would be $t = 0$ for that specific patient. So we collect the patient's EMR_0 for year 2009 (2009-EMR) and then also collect all the records of complications test results for up to 8 years inclusive, which for this example means from 2009 ($t = 0$) up to 2017 ($t = 8$) inclusive; see Fig. 2. Let's assume a patient became hypertrophic cardiomyopathy patient at 2010 and we are doing a 3-year prediction. If there is no record of all three complications tests (positive or negative) in 2011, 2012 or 2013, then the patient's EMR would be censored from the analysis. In addition, for example in 3-year prediction, we censor the patient's EMR, if the patient has been diagnosed to one of the complications between 2009–2012 inclusive. That is, our prediction focuses on 1, 3, and 8 years after a patient becomes a hypertrophic cardiomyopathy patient and gets diagnosed to a complication.

Table 3 presents the number of records after cleaning. The study incorporates the de-identified EMRs for 1078 hypertrophic cardiomyopathy patients. The EMRs under the evaluations had the patients' information from April 2009-April 2017, inclusive.

We adopted SNOMED CT-AU Ontology (SCAO) [74] for chronic diseases because SCAO was the terminology used by the EMR software providers in the clinics. The predictors chosen have values for more than 18 % of EMRs. This is an acceptable cut-off value proposed by Sariyar et al. [75] for studies involving EMRs. 107 EMR variables were then examined in order to predict the three most commonly experienced complications associated with hypertrophic cardiomyopathy: heart failure, problems with heart valves, and cardiac arrest [40]. The list of 107 EMR variables under this study is given in Appendix 1.

3.2. Predictions

Fig. 2 presents the prediction process designed to compare IPMC and CPMC. The above-mentioned data was randomly divided into train

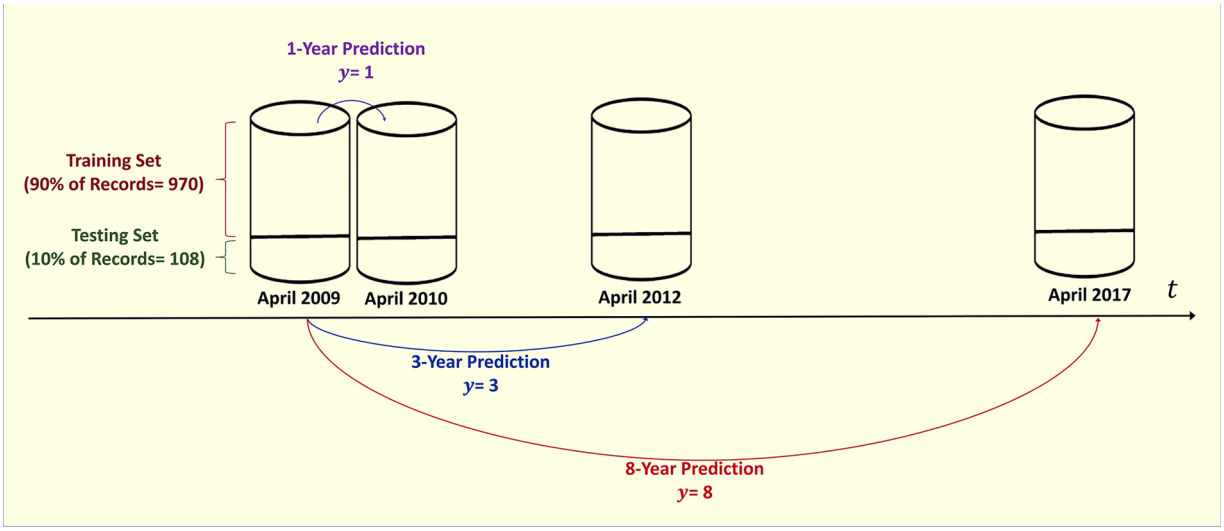

Fig. 2. 1-year, 3-year and 8-year Predictions.

and test datasets with the ratio of 9:1. The IPMC and CPM were trained using 90 % of the records (970 records) and tested on 10 % of the records (108 records). The predicted variables were heart failure, problems with heart valves, and cardiac arrest. The predicted values (i.e. 0: the complication has not been experienced, 1: the complication has been experienced) were compared to actual value in the testing dataset to measure the performance. The performance indicators are discussed in Sections 3.3.1 and 3.3.2.

To also evaluate the role of prediction periods (i.e. y) in the performance of prediction with the methods, this study ran IPMC and CPMC for a 1-year prediction (i.e. $y = 1$), a 3-year prediction (i.e. $y = 3$) and an 8-year prediction (i.e. $y = 8$). For example, in a 3-year prediction, the predictors $predictor_p = [predictor_{p,1}, predictor_{p,2}, ..., predictor_{p,a}]^\tau$ were taken from 90 % of the EMRs on April 2009 and predicted variables (i.e. heart failure, problems with heart valves, cardiac arrest) were taken from the same EMRs on April 2012. To measure the performance of IPMC and CPMC, we inputted the trained models by predictors extracted from the rest of the EMRs on April 2009. The predicted values were compared to the 10 % remaining EMRs in 2012. The performance was measured using performance indicators. Similarly, 1-year and 8-year predictions were conducted.

3.3. Comparison

The performance of IPMC and CPMC in this study was measured based on their accuracy. In order to understand the accuracy of the models, two concerns were taken into account: discrimination and calibration. Accurate prediction of each complication differs between hypertrophic cardiomyopathy patients that are at risk of developing that complication, versus those that are not. This refers to discrimination, which demonstrates how well a prediction can classify hypertrophic cardiomyopathy patients for developing a particular complication. Calibration is a probabilistic understanding of error which measures deviation of the predicted probability from the observed probability [76], or in other words, how close a prediction probability is to an observed probability.

3.3.1. Measuring discrimination

Discrimination was measured by utilizing a commonly-used method in clinical medicine called the receiver operating characteristics (ROC) curve [77]. The ROC curve has two dimensions: sensitivity on the Y-axis and specificity on the X-axis. *Sensitivity* measures the proportion of positives that are correctly identified; for instance, in our study, it presents a portion of between 0–1 of patients diagnosed with heart failure, problems with heart valves, and/or cardiac arrest who are correctly identified by the method. *Specificity* measures the proportion of negatives that are correctly identified; for instance, in our study, specificity presents a portion of 0–1 of patients that are not diagnosed with heart failure, problems with heart valves, and/or cardiac arrest, and they are correctly identified by the method as not developing these complications. The performance indicator used in this study is area under the curve (*AUC*) of ROC. AUC is a scalar indicator that a test with no better performance than chance has an AUC of 0.5, but a perfect test has an AUC of 1. AUC is often utilized in clinical medicine to assess the tradeoff between sensitivity and specificity [78]. Table 4 presents the results of discrimination.

3.3.2. Measuring calibration

The calibration of the predictions was measured using an indicator called *Q-Score* proposed by Ghil et al. [79]. This indicator has been adopted from the Mean Squared Error, which demonstrates how much the predicted probability deviates from the observed probability and penalizes strong deviations [76]. In order to build a comparative measure, Ghil et al. [79] divides the mean square errors of the prediction under evaluation with unskilled prediction that predicts

Table 3
Number of electronic medical records in each clinic*.

General Practice Office	Number of Electronic Medical Records	General Practice Office	Number of Electronic Medical Records	General Practice Office	Number of Electronic Medical Records	General Practice Office	Number of Electronic Medical Records
1	126	4	119	7	47	10	64
2	74	5	144	8	209	11	71
3	103	6	97	9	86	12	112

* Some of the EMRs were shared between clinics.

Table 4
Discrimination analysis results.

Prediction Period	Predictive Model	Heart failure		Problems with heart valves		Cardiac arrest	
		IPMC	CPMC	IPMC	CPMC	IPMC	CPMC
1-Year Prediction	BLR	0.862	0.893***	0.810	0.848***	0.791	0.830***
	DT	0.769	0.799***	0.731	0.768***	0.729	0.771***
	SVM	0.774	0.797**	0.751	0.788***	0.724	0.761***
	ANN	0.791	0.821***	0.762	0.799***	0.746	0.781***
3-Year Prediction	BLR	0.892	0.941***	0.851	0.907***	0.843	0.863**
	DT	0.779	0.812***	0.740	0.802***	0.776	0.799**
	SVM	0.811	0.840***	0.792	0.822***	0.762	0.785**
	ANN	0.747	0.781***	0.769	0.814***	0.781	0.803***
8-Year Prediction	BLR	0.881	0.893*	0.821	0.848**	0.813	0.830**
	DT	0.788	0.811***	0.764	0.780***	0.726	0.783***
	SVM	0.761	0.814***	0.786	0.788**	0.741	0.761**
	ANN	0.771	0.813***	0.779	0.781***	0.735	0.794***

Comparison of CMPS versus IMPC: *, ** and *** compare prediction of a particular complication using the same predictive model.
*** Comparing CPMC with IPMC, the AUC is significantly higher at.
**Comparing CPMC with IPMC, the AUC is significantly higher at.
*Comparing CPMC with IPMC, the AUC is significantly higher at.
Example: For predicting Heart Failure, CPMC demonstrates a significantly better AUC than IPMC at.
Comparison of BLR with other predictive models to implement CMPC: Underlines compare CPMC prediction of a particular complication when implementing using BLR versus other predictive models.
: Comparing the predictive model with BLR to implement CPMC, the AUC is significantly lower at.
=: Comparing the predictive model with BLR to implement CPMC, the AUC is significantly lower at.
—: Comparing the predictive model with BLR to implement CPMC, the AUC is significantly lower at.
Example: For predicting heart failure, CPMC_LR demonstrates a significantly lower AUC than CPMC_BLR at $\rho < 0.05$. In other words, implementation of CPMC using BLR rather than LR significantly improves AUC at $\rho < 0.05$ when predicting heart failure.

constant historic average count. Specifically, we define this with Eq. (4).

$$Q - Score = \frac{MSE_{prediction}}{MSE_{unskilled\ prediction}} \quad (4)$$

The Q-Score may take positive values. The *Q-Score* = 1 if the prediction under the evaluation generates similar results to the unskilled prediction producing a constant average. A desired time series analysis produces a *Q-Score* < 1. Therefore, the aim is to minimize the *Q-Score*.

Table 5 presents the results of the calibration analysis.

4. Analysis of evaluation

4.1. CPMC performs better than IPMC

As previously discussed, the first objective of evaluation was to compare IPMC and CPMC to examine the performance of STL and MTL. IPMC and CPMC have been trained to predict three complications of

Table 5
Calibration analysis results.

Prediction Period	Predictive Model	Heart failure		Problems with heart valves		Cardiac arrest	
		IPMC	CPMC	IPMC	CPMC	IPMC	CPMC
1-Year Prediction	BLR	0.283	0.192***	0.391	0.374*	0.265	0.182***
	DT	0.501	0.497	1.038	0.475**	0.412	0.391*
	SVM	0.331	0.319*	0.487	0.409***	0.403	0.204***
	ANN	0.352	0.207***	0.539	0.411***	0.394	0.299***
3-Year Prediction	BLR	0.246	0.211**	0.165	0.084***	0.298	0.201***
	DT	1.030	0.498***	0.743	0.472***	0.601	0.443***
	SVM	0.414	0.399*	0.345	0.101***	0.307	0.297
	ANN	0.419	0.240***	0.139	0.123*	0.365	0.223***
8-Year Prediction	BLR	0.221	0.107***	0.256	0.176***	0.556	0.297***
	DT	0.678	0.317***	0.561	0.387***	1.088	0.485***
	SVM	0.457	0.294***	0.401	0.184***	0.781	0.325***
	ANN	0.349	0.128***	0.310	0.293*	0.502	0.329***

Comparison of CMPS versus IMPC: *, ** and *** compare prediction of a particular complication using the same predictive model.
*** Comparing CPMC with IPMC, the Q-Score is significantly lower at.
** Comparing CPMC with IPMC, the Q-Score is significantly lower at.
* Comparing CPMC with IPMC, the Q-Score is significantly lower at.
Example: For predicting Heart Failure, CPMC demonstrates a significantly better Q-Score than IPMC at.
Comparison of BLR with other predictive models to implement CMPC: Underlines compare CPMC prediction of a particular complication when implementing by BLR versus other predictive models.
:Comparing the predictive model with BLR to implement CPMC, the Q-Score is significantly higher at.
=:Comparing the predictive model with BLR to implement CPMC, the Q-Score is significantly higher at.
—:Comparing the predictive model with BLR to implement CPMC, the Q-Score is significantly higher at.
Example: For predicting heart failure, CPMC_LR demonstrates a significantly higher Q-Score than CPMC_BLR at $\rho < 0.05$. In other words, implementation of CPMC using BLR rather than LR significantly improves the Q-Score at $\rho < 0.05$ when predicting heart failure.

hypertrophic cardiomyopathy patients using four different predictive models: LR, DT, SVM and ANN. However, the evaluations have been repeated for 1-year, 3-year, and 8-year predictions. Table 4 and Table 5 show that CPMC outperforms the alternative IPMC predictive models. While a better discrimination can be demonstrated by an AUC closer to 1, the calibration can be improved by a *Q-Score* closer to zero. A *Q-Score* of more than 1 is unacceptable calibration [79]. The non-parametric significance test of the AUC and *Q-Score* demonstrated that in most of the cases CPMC had significantly better discrimination and calibration than IPMC. This reveals the better performance of MTL in comparison with STL.

To compare the discrimination capabilities of IPMC and CPMC, we examined the AUCs of different implementations for IPMC and CPMS across three prediction periods and for three complications. Generally, both CPMC and IPMC show acceptable AUC measures of more than 0.500. Table 4 reveals that the AUC measurements in different predictions were significantly higher by CMPC than IPMC. On average, the difference between the AUCs of CPMC and IPMC was 0.029. The discrimination differences of CPMC and IPMC were similar to each other and were also close to the overall average for predictions of heart failure (i.e. 0.028), problems with heart valves (i.e. 0.030) and cardiac arrest (i.e. 0.029).

Like discrimination comparisons, to compare the calibration of IPMC and CPMC, we examined the *Q-Scores* of different implementations of IPMC and CPMC across three prediction periods and for three complications. Although on some occasions IPMC revealed *Q-Scores* of more than 1 (i.e., unacceptable predictions), CPMC consistently stayed under 0.498, demonstrating a promising performance. Table 5 reveals that even with a couple of incorrect predictions (i.e. 1-year and 3-year predictions of heart failure by DT), when CPMC was compared with IPMC, CPMC significantly outperformed in terms of calibration. On average, the difference between the *Q-Scores* of IPMC and CPMC was 0.148. The discrimination differences of IPMC and CPMC were similar to each other and were also close to the overall average for predictions of heart failure (i.e. 0.147), problems with heart valves (i.e. 0.147) and cardiac arrest (i.e. 0.151).

In summary, taking both the AUCs and *Q-Scores* into consideration, CPMC shows better performance when compared with IPMC; therefore, when compared with STL, MTL outperforms in terms of discrimination and calibration. Although we have seen some exceptions in which the performance of CPMC was not significantly different than IPMC, there was no case prediction reported in Table 4 and Table 5 that demonstrate improving performance in STL when compared with MTL.

4.2. CPMC with logistic regression performs better than CPMC with other predictive models

The second objective of evaluation was to compare implementations of CPMS using different predictive models. MTL using LR, DT, SVM and ANN were trained to predict three complications of hypertrophic cardiomyopathy patients. However, the evaluations have been repeated for 1-year, 3-year, and 8-year predictions. Table 4 and Table 5 show that CPMC using Logistic Regression outperforms when compared to the implementation of CPMC using alternative predictive models. While a better discrimination can be demonstrated by an AUC closer to 1, the calibration can be improved by a *Q-Score* closer to zero. A *Q-Score* of more than 1 is unacceptable calibration [79]. The non-parametric significance test of the AUC and *Q-Score* demonstrated that in most cases, BLR had significantly better discrimination and calibration than other predictive models. This reveals that the performance of MTL to predict complications of hypertrophic cardiomyopathy patients improves by LR in comparison with DT, SVM and ANN.

To compare discrimination capabilities of implementations of CPMC using different predictive models, we examined AUCs of CPMC across three prediction periods and for three complications using different predictive models. Table 4 reveals that the AUC measures are significantly higher in LR compared to DT, SVM and ANN. In average, the difference between AUCs of CPMC implemented by LR to other predictive models is 0.115. The discrimination differences of predictive models implementing CPMC were similar to each other and were also close to the overall average for predictions of heart failure (i.e. 0.096), problems with heart valves (i.e. 0.116) and cardiac arrest (i.e. 0.136).

Like discrimination, in order to compare the calibration capabilities of implementations of CPMC using different predictive models, we examined the *Q-Scores* of CPMC across three prediction periods and for three complications using different predictive models. Table 5 reveals that except in few predictions, LR compared with DT, SVM and ANN significantly improves the calibration of CPMC. Although we have some exceptions in significantly better calibration of BR, all CPMC predictions using LR show smaller *Q-Scores* (though not significantly smaller). On average, the difference between the *Q-Scores* of IPMC and CPMC was 0.091. The calibration differences of predictive models implementing CPMC were similar to others and were also close to the overall average for predictions of heart failure (i.e. 0.096), problems with heart valves (i.e. 0.088) and cardiac arrest (i.e. 0.089).

In summary, taking both AUCs and *Q-Scores* into consideration, CPMC shows better performance when using LR compared with DT, SVM and ANN. Therefore, MTL outperforms in terms of discrimination and calibration by LR. Although we have seen some exceptions in which the performance of LR was not significantly different than other predictive models, there was no case prediction alternative models when compared with MTL. reported in Table 4 and Table 5 that demonstrate improving performance in the

5. Discussion and conclusion

Policy makers and healthcare professionals have sought to reconstruct daily generated EMRs and to integrate them into large clinical data warehouses for use in auditing, continuous quality improvement, health service planning, epidemiological studies, and evaluation research [80–83]. Although managing the increasing amount of daily generated EMRs is essential [84], the potential benefits of EMRs are not limited to storage and retrieval of patients' records. EMRs stored over time can be used as an analytics resource for developing and advancing clinical decision support.

This paper focused on predicting multiple complications of chronic diseases using daily generated EMRs. There is a body of literature that conducts analytics techniques in order to predict various medical outcomes [1–3,7]. Following that stand of literature, the current study proposed concurrent prediction of multiple complications (CPMC) that use MTL, as opposed to independent prediction of multiple complications (IPMC) that uses STL in order to predict multiple complications caused by chronic diseases. This paper also suggests the utilization of logistic regression in order to implement MTL (i.e. CPMC). While the chronic care management plan is an important part of primary care that leads to better outcomes and higher patient satisfaction, the differences of chronic patients in their risk to develop medical complications has made generating a comprehensive plan a difficult task. The present research developed a method that can be implemented in clinical decision support systems to assist medical professionals to identify patients at risk of progressing complications caused by chronic diseases.

In order to evaluate the CPMC-LR proposed by this paper, as an exemplar, we predicted three main complications of hypertrophic cardiomyopathy patients: heart failure, problems with heart valves, and cardiac arrest. Our experiment showed that CPMC has consistently demonstrated better performance when compared to IPMC. The performance was compared in terms of discrimination and calibration measured by the AUC and *Q-Score*, respectively. Furthermore, in our evaluation, we showed that MTL improves the prediction of independent STL of different chronic disease complications. Our findings also showed that Bayesian Logistic Regression improves the performance of predicting multiple complications of chronic diseases. Even

though there were cases that did not demonstrate the significant improvement in both above-mentioned comparisons, all experiments demonstrated higher AUCs and lower *Q-Scores*.

Evaluating both discriminations and calibrations asured us that the CPMC-LR implementing MTL using logistic regression improves: 1) the classification of chronic patients based on whether they are at risk of developing complications, and 2) the identification of the risk probability of developing complications of chronic disease. While the first outcome can be used to identify individual chronic patients at risk of developing complications, the second outcome can be utilized in population healthcare programs for chronic patients.

5.1. Academic implications

This study made two important contributions. First, the present paper contributed to the healthcare analytics literature [27–29,31]; in particular, analytics techniques used in the context of predicting complications of chronic diseases [8,37–39]. Following the work of Zolbanin et al [36] for predicting survivability when multiple chronic diseases are presented, an awareness among the healthcare analytics community was raised about the inter-correlation of chronic conditions. For example, in order to predict diabetic vision loss, Piri et al. [8] considered the co-existence of other diabetes-related complications as predictors and implemented single-task learning. However, complications may develop concurrently. The contribution of the present work is represented by our focus on concurrent prediction of interrelated complications, which contrasts with the existing literature in this area [8,45]. Despite the proposed method, the prior literature (in order to capture the occurrence of each particular complication) separately predicts the complications of chronic diseases independently or by setting other complications as predictors, rather than outcomes. However, multiple complications of chronic diseases are often related and can be developed concurrently [11]. Following this line of literature, the study proposes the use of MTL to predict multiple complications of chronic diseases to address the interrelationships among multiple complications. Second, the current study established the outperforming application of logistic regression for the prediction of multiple complications for chronic patients. The paper concluded that in comparison to DT, SVM and ANN, logistic regression shows better performance in terms of both discrimination and calibration for concurrent prediction of multiple complications of chronic diseases using MTL. This is in response to the call from a survey of the literature of predictive models for MTL [85]. This review recommends researchers to compare predictive models to implement MTL for predictions in different application contexts.

The unique feature of this study was the comprehensive comparison of different implementations of single- and MTL using LR, DT, SVM and ANN and in different prediction periods. To the best of our knowledge, this is the first time in the literature that such a comprehensive comparison was performed.

5.2. Practical implications

We know from previous studies that intervention for chronic diseases, such as making changes to a patient's diet and exercise, can postpone and in some cases reverse the progression of complications [86,87]. As such, many institutes and organizations recommend periodic screening. However, many people do not receive appropriate tests early enough to avoid irreversible complications of their chronic diseases. It is reasonable to assume that in many cases, the complexity of guidance regarding screening coupled with challenges in public education have inhibited efficiency in population-level chronic disease management. With this paper, we highlighted a method to inform public education and rapidly screen population databases in order to prioritize patients. In such an environment, the CPMC offers an automatable technique that can be computerized and used to identify chronic patients who are at risk of developing complications based on the data in their EMRs (Goldstein et al., 2017). However, the choice of predictive techniques and advances in approaches such as MTL are important.

This study suggests CPMC as a MTL-based approach. The paper highlights that the outperforming capability of logistic regression in implementation of this method makes it capable of providing guidelines for chief information and quality assurance officers. CPMC allows corporate healthcare providers to utilize proper methods for right conditions when predicting multiple complications of chronic diseases. This would be implemented by an automated system that can prompt patients for further screening.

5.3. Comparison with related studies

There is a body of literature in the use of MTL for predicting patients at risk of developing complications caused by chronic diseases. For example, Lin et al. [88] and Liu et al. [89] utilize MTL to predict the complications of diabetes and compare MTL performance with STL. The results show that MTL outperforms. Although the present study agrees with the findings of these two papers, it has four differences with the work of Lin et al. [88] and Liu et al. [89]. First, while the previous literature focuses only on Bayesian logistic regression, the current work also generalizes the comparison when different predictive models such as decision tree, support vector model and artificial neural networks are used. According to Table 4 and Table 5, MTL outperforms STL, in all above-mentioned predictive models. Second, the present paper enhances Lin et al. [88] and Liu et al. [89] by changing the predicting period from 1 to 3 and 8 years. The results acknowledge that MTL outperforms STL, regardless of the prediction period. Third, The present study advances the evaluation measures used by Lin et al. [88] and Liu et al. [89]. In addition to AUC as a measure of discrimination, Q-Score is used to compare the STL and MTL based on their calibration. Fourth, the evaluation conducted by Lin et al. [88] and Liu et al. [89] occurred in the context of diabetes, however the current work compared these two approaches to predict the complications of hypertrophic cardiomyopathy.

5.4. Comparison with MTL baseline methods

MTL is a well-studied problem and generally improves single task peers. In this section, we have included two MTL baseline methods and show the method proposed in this study outperforms these two baseline methods. The two baseline methods chosen for this purpose are: (A) Argyriou et al. [90] as a feature-based MTL method that aims to learn common predictors s; and (B) Zhang and Yeung [91] as a task correlation based MTL that seeks to improve the performance of a learning task with the help of some other related tasks.

Table 6 presents the results of comparing our measure of discrimination, AUC, for Argyriou et al. [90]Zhang and Yeung [91], and the CPMC method. The results demonstrate the CPMC method has a higher AUC compared to the other two methods since CPMC has a significantly higher AUC in most predictions.

In order to compare calibration, Table 7 compares Q-scores for Argyriou et al. [90] Zhang and Yeung [91], and the CPMC method. The results demonstrate that the CPMC method has a lower Q-Score compared to the other two methods. In all predictions, CPMC has a significantly better Q-Score.

The results shown in Table 6 and Table 7 confirm that the CPMC method outperforms both baseline MTL methods previously proposed in literature [90,91].

5.5. Limitations

In this section, we indicate the limitations that have the potential to open new avenues for future research.

Table 6
Comparison with MTL Baseline Methods: Discrimination analysis results.

Prediction Period	Heart failure			Problems with heart valves			Cardiac arrest		
	Argyriou et al. [90]	Zhang and Yeung [91]	CPMC	Argyriou et al. [90]	Zhang and Yeung [91]	CPMC	Argyriou et al. [90]	Zhang and Yeung [91]	CPMC
1-Year Prediction	0.809**	0.817*	0.821	0.764***	0.72***	0.799	0.752***	0.764***	0.781
3-Year Prediction	0.754***	0.761**	0.781	0.773***	0.788***	0.814	0.787***	0.789***	0.803
8-Year Prediction	0.779***	0.784***	0.813	0.778*	0.779	0.781	0.749***	0.759***	0.794

***the AUC is significantly lower at $\rho < 0.01$.
**the AUC is significantly lower at $\rho < 0.05$.
*the AUC is significantly lower at $\rho < 0.1$.
Example: For predicting Heart Failure in 8 years, CPMC demonstrates a significantly better AUC than Argyriou et al. [90] and Zhang and Yeung [91] at $\rho < 0.01$.

Table 7
Comparison with MTL Baseline Methods: Calibration analysis results.

Prediction Period	**Heart failure**			**Problems with heart valves**			**Cardiac arrest**		
	Argyriou et al. [90]	Zhang and Yeung [91]	**CPMC**	Argyriou et al. [90]	Zhang and Yeung [91]	**CPMC**	Argyriou et al. [90]	Zhang and Yeung [91]	**CPMC**
1-Year Prediction	0.331***	0.298***	0.207	0.499***	0.454***	0.411	0.344***	0.315*	0.299
3-Year Prediction	0.385***	0.281***	0.240	0.131*	0.133*	0.123	0.351***	0.289***	0.233
8-Year Prediction	0.276***	0.209***	0.128	0.281**	0.285*	0.293	0.482***	0.411***	0.329

*** the Q-Score is significantly higher at.
** the Q-Score is significantly higher at.
* the Q-Score is significantly higher at.
Example: For predicting Heart Failure in 8 years, CPMC demonstrates a significantly better Q-Score than Argyriou et al. [90] and Zhang and Yeung [91] at $\rho < 0.01$.

- This study is limited to the dataset that it has used. Using a baseline comparison could have improved the evaluation presented in this study. However, since we compare STL and MTL to predict the complications of hypertrophic cardiomyopathy in 1, 3 and 8 years, there is a limited data available. Therefore, this study collected its own data and compared the performance of STL and MTL. Section 5.3 compares the findings of this comparison with related literature [88,89]. Therefore, we recommend that researchers reproduce the current study and compare the additional results.
- Although this paper has focused on hypertrophic cardiomyopathy patients, it is recommended that researchers utilize the CPMC proposed in this article to predict multiple complications of other chronic diseases.
- This study assumes the same predictors for all complications. However, this is an idealization and predictions of complications might be dependent on different sets of predictors. It would be rewarding in future to pay attention to framing EMR-based classification algorithms as an all-encompassing strategy to inform interventions, particularly in an area where effective interventions may have been informed by analysis of different predictors using data that might not be routinely available via patient records, such as data relating to actual lifestyle choices. The lack of key predictor data (i.e. diet and exercise) in the study dataset has provided some limitations to the results of this study, in particular for long term prediction. It is common knowledge that long-term predictions of diabetes rely on modifiable life-style choices for which there is no data in the reported study. Therefore, the lack of life-style data in our dataset has committed this study to ignore life-style changes for diabetes management, which is clearly a limitation for the present work.
- Further studies are recommended to model the clinical cost-benefit analysis of any predictive data effort seeking to identify people at risk of developing multiple complications of chronic diseases. Individual healthcare providers would argue that they are already implementing the findings of this study, but on a patient-by-patient basis. The key of this endeavor is to highlight opportunities for population-level education and patient-initiated screening and analysis of entire patient panels at scale for prioritized screening and intervention.

Acknowledgment

The authors would like to thank the anonymous reviewers and the editor for their insightful comments and suggestions. Dr. Madjid Tavana is grateful for the partial support he received from the Czech Science Foundation (GAˇCR19-13946S) for this research.

Appendix 1. EMR predictors under study

Ability for face-to-face interactions	Delivery of a baby weighing > 9 lb	History of chlamydia	Parental history of blood pressure
Acanthosis Nigerians	Dementia	History of corticosteroids intake	Parental history of cancer
Age	Diastolic Blood Pressure	History of depression	Parental history of diabetes
Allergy to eggs	Egg substance	History of Gynecological Surgery	Peripheral Arterial Disease
Allergy to influenza vaccine	Ethnicity	History of hemodialysis	Physical Activity
Antipsychotic therapy for schizophrenia and/or severe bipolar disease	Excessive thirst	Pneumococcal Vaccine	Asthma
Arrhythmia	History of high-risk medication intake	Polycystic ovary syndrome	Atherosclerosis Disease

Fasting plasma glucose	History of hospice care	Ratio of polyunsaturated–saturated fatty acids	Behavioral/Neuropsych Assessment
Flexible Sigmoidoscopy	History of influenza vaccine	Rhabdomyolysis	Blurry vision
Fracture - Lower Body	History of kidney transplant	Sex, male	Body Mass Index
Frequent urination at night	History of leucovorin intake	Sexually Active	Bradycardia
Genital Herpes	History of liver disease	Shortness of Breath	Cardiovascular disease
Glomerulonephritis and Nephrotic Syndrome	History of mood instability	Sleep disorders in the presence of glucose intolerance	Carotid artery disease
Gonococcal Infections	History of pregnancy	Smoking status	CD4+ count
Height	Hemoglobin A1c, %	History of pregnancy Dx	Statin Allergen
CD4+ percentage	Hemorrhagic Stroke	History of spinal infection	Syphilis
Chronic glucocorticoid exposure	Hepatitis A	History of suicide	Systolic blood pressure
Cognitive assessment	Hepatitis B	HIV 1	Systolic Blood Pressure, mm Hg
Colonoscopy	Hepatitis C	HIV 2	Trauma
Corticosteroids	High density lipoprotein cholesterol, mg/dL	Injection of leuprolide acetate	Triglycerides, per 10 mg/dL
CT Colonography	History of Beta Blocker Therapy	Ischemic Stroke	Upper Respiratory Infection
CT Scan of Lower Spine	History of bipolar disorder	Isotretinoin	Venereal Diseases
Daily Alcohol Intake	History of breastfeeding	Low-density lipoprotein	Waist Circumference, per 10 cm
Daily Cereal fiber intake	History of CABG surgeries	Major Organ Transplant Other Than Kidney	Waist: hip ratio, per 0.1 unit
Daily consumption of vegetables	History of cancer	Malignant Neoplasm of Colon	X-Ray of lower spine
Daily Trans-fat intake	History of Cerebrovascular disease, brain stroke, or TIA	Neurologic impairment	
Dapsone and pyrimethamine	History of Cesarean Birth	Nonalcoholic fatty liver	

References

[1] Amarasingham R, Patzer RE, Huesch M, Nguyen NQ, Xie B. Implementing electronic health care predictive analytics: considerations and challenges. Health Aff. (Millwood) 2014;33:1148–54.

[2] Bardhan I, Oh J, Zheng Z, Kirksey K. Predictive analytics for readmission of patients with congestive heart failure. Inf. Syst. Res. 2014;26:19–39.

[3] Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff. (Millwood) 2014;33:1123–31.

[4] Chen L, Li X, Yang Y, Kurniawati H, Sheng QZ, Hu H-Y, et al. Personal health indexing based on medical examinations: a data mining approach. Decis Support Syst 2016;81:54–65. https://doi.org/10.1016/j.dss.2015.10.008.

[5] Dag A, Oztekin A, Yucel A, Bulur S, Megahed FM. Predicting heart transplantation outcomes through data analytics. Decis Support Syst 2017;94:42–52.

[6] Delen D, Oztekin A, Tomak L. An analytic approach to better understanding and management of coronary surgeries. Decis Support Syst 2012;52:698–705. https://doi.org/10.1016/j.dss.2011.11.004.

[7] Meyer G, Adomavicius G, Johnson PE, Elidrisi M, Rush WA, Sperl-Hillen JAM, O'Connor PJ. A machine learning approach to improving dynamic decision making. Inf. Syst. Res. 2014;25:239–63. https://doi.org/10.1287/isre.2014.0513.

[8] Piri S, Delen D, Liu T, Zolbanin HM. A data analytics approach to building a clinical decision support system for diabetic retinopathy: developing and deploying a model ensemble. Decis Support Syst 2017;101:12–27.

[9] Tseng C-J, Lu C-J, Chang C-C, Chen G-D, Cheewakriangkrai C. Integration of data mining classification techniques and ensemble learning to identify risk factors and diagnose ovarian cancer recurrence. Artif Intell Med 2017;78:47–54. https://doi.org/10.1016/j.artmed.2017.06.003.

[10] Wulff A, Haarbrandt B, Tute E, Marschollek M, Beerbaum P, Jack T. An interoperable clinical decision-support system for early detection of SIRS in pediatric intensive care using openEHR. Artif Intell Med 2018. https://doi.org/10.1016/j.artmed.2018.04.012.

[11] Brown CL, Hammill BG, Qualls LG, Curtis LH, Muir AJ. Significant morbidity and mortality among hospitalized end-stage liver disease patients in Medicare. J Pain Symptom Manage 2016;52:412–9. e1.

[12] Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Vis Sci 2016;57:5200–6.

[13] Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. J Am Med Inform Assoc 2016;112. https://doi.org/10.1093/jamia/ocw112.

[14] Dagliati A, Marini S, Sacchi L, Cogni G, Teliti M, Tibollo V, et al. Machine learning methods to predict diabetes complications. J Diabetes Sci Technol 2017:1932296817706375.

[15] Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012;13:395.

[16] Kohli R, Tan SS-L. Electronic Health Records: How Can IS Researchers Contribute to Transforming Healthcare? MIS Q 2016;40:553–73.

[17] López B, Martin C, Viñas PH. Special section on artificial intelligence for diabetes. Artif Intell Med 2018;85:26–7. https://doi.org/10.1016/j.artmed.2017.09.008.

[18] Park K, Ali A, Kim D, An Y, Kim M, Shin H. Robust predictive model for evaluating breast cancer survivability. Eng Appl Artif Intell 2013;26:2194–205.

[19] Stoean R, Stoean C, Lupsor M, Stefanescu H, Badea R. Evolutionary-driven support vector machines for determining the degree of liver fibrosis in chronic hepatitis C. Artif Intell Med 2011;51:53–65. https://doi.org/10.1016/j.artmed.2010.06.002.

[20] Tabak YP, Sun X, Nunez CM, Johannes RS. Using electronic health record data to develop inpatient mortality predictive model: acute Laboratory Risk of Mortality Score (ALaRMS). J Am Med Inform Assoc 2013;21:455–63.

[21] Yeh J-Y, Wu T-H, Tsao C-W. Using data mining techniques to predict hospitalization of hemodialysis patients. Decis Support Syst 2011;50:439–48. https://doi.org/10.1016/j.dss.2010.11.001.

[22] Sangi M, Win KT, Shirvani F, Namazi-Rad M-R, Shukla N. Applying a novel combination of techniques to develop a predictive model for diabetes complications. PLoS One 2015;10:e0121569https://doi.org/10.1371/journal.pone.0121569.

[23] Maron BJ. Hypertrophic cardiomyopathy: a systematic review. JAMA 2002;287:1308–20. https://doi.org/10.1001/jama.287.10.1308.

[24] Kotsiantis SB, Zaharakis ID, Pintelas PE. Machine learning: a review of classification and combining techniques. Artif Intell Rev 2006;26:159–90.

[25] Bardou D, Zhang K, Ahmad SM. Lung sounds classification using convolutional neural networks. Artif Intell Med 2018;88:58–69. https://doi.org/10.1016/j.artmed.2018.04.008.

[26] Kang S. Personalized prediction of drug efficacy for diabetes treatment via patient-level sequential modeling with neural networks. Artif Intell Med 2018;85:1–6. https://doi.org/10.1016/j.artmed.2018.02.004.

[27] Liang Z, Zhang G, Huang JX, Hu QV. Deep Learning for Healthcare Decision Making With EMRs, in: Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference On. 2014. p. 556–9.

[28] Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief. Bioinform 2017.

[29] Ravì D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep learning for health informatics. IEEE J Biomed Health Inform 2017;21:4–21.

[30] Schetinin V, Jakaite L, Krzanowski W. Bayesian averaging over Decision Tree models for trauma severity scoring. Artif Intell Med 2018;84:139–45. https://doi.org/10.1016/j.artmed.2017.12.003.

[31] Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE J Biomed Health Inform 2017.

[32] Suner A, Çelikoğlu CC, Dicle O, Sökmen S. Sequential decision tree using the analytic hierarchy process for decision support in rectal cancer. Artif Intell Med 2012;56:59–68. https://doi.org/10.1016/j.artmed.2012.05.003.

[33] Nguyen P, Tran T, Wickramasinghe N, Venkatesh S. Deepr: a convolutional net for medical records. IEEE J Biomed Health Inform 2017;21:22–30.

[34] Zheng T, Xie W, Xu L, He X, Zhang Y, You M, et al. A machine learning-based framework to identify type 2 diabetes through electronic health records. Int J Media Inf Lit 2017;97:120–7. https://doi.org/10.1016/j.ijmedinf.2016.09.014.

[35] Walczak S, Velanovich V. An evaluation of artificial neural networks in predicting pancreatic Cancer survival. J Gastrointest Surg 2017;21:1606–12.

[36] Zolbanin HM, Delen D, Hassan Zadeh A. Predicting overall survivability in comorbidity of cancers: a data mining approach. Decis Support Syst 2015;74:150–61. https://doi.org/10.1016/j.dss.2015.04.003.

[37] Ting DSW, Cheung CY-L, Lim G, Tan GSW, Quang ND, Gan A, Hamzah H, Garcia-Franco R, San Yeo IY, Lee SY. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. Jama 2017;318:2211–23.

[38] Kothari V, Stevens RJ, Adler AI, Stratton IM, Manley SE, Neil HA, Holman RR. UKPDS 60: risk of stroke in type 2 diabetes estimated by the UK Prospective Diabetes Study risk engine. Stroke 2002;33:1776–81.

[39] Pahl E, Sleeper LA, Canter CE, Hsu DT, Lu M, Webber SA, et al. Incidence of and risk factors for sudden cardiac death in children with dilated cardiomyopathy: a report from the pediatric cardiomyopathy registry. J Am Coll Cardiol 2012;59:607–15. https://doi.org/10.1016/j.jacc.2011.10.878.

[40] American Heart Association. Hypertrophic cardiomyopathy [WWW document] URL http://www.heart.org/HEARTORG/Conditions/More/Cardiomyopathy/Hypertrophic-Cardiomyopathy_UCM_444317_Article.jsp#.WoczfKjwbD4 (accessed 2.16.18). 2016.

[41] Caruana R. Multitask learning, in: learning to learn. Boston, MA: Springer; 1998. p. 95–133. https://doi.org/10.1007/978-1-4615-5529-2_5.

[42] Tan M. Prediction of anti-cancer drug response by kernelized multi-task learning. Artif Intell Med 2016;73:70–7. https://doi.org/10.1016/j.artmed.2016.09.004.

[43] Zhou D, Miao L, He Y. Position-aware deep multi-task learning for drug–drug interaction extraction. Artif Intell Med 2018;87:1–8. https://doi.org/10.1016/j.artmed.2018.03.001.

[44] Tangri N, Stevens LA, Griffith J, Tighiouart H, Djurdjev O, Naimark D, Levin A, Levey AS. A predictive model for progression of chronic kidney disease to kidney failure. Jama 2011;305:1553–9.

[45] Zhang D, Shen D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease. NeuroImage 2012;59:895–907. https://doi.org/10.1016/j.neuroimage.2011.09.069.

[46] Razavian N, Marcus J, Sontag D. Multi-task prediction of disease onsets from longitudinal laboratory tests, in: machine learning for healthcare Conference. Presented at the Machine Learning for Healthcare Conference 2016:73–100.

[47] Chuang HH-C. Mathematical modeling and Bayesian estimation for error-prone retail shelf audits. Decis Support Syst 2015;80:72–82. https://doi.org/10.1016/j.dss.2015.10.003.

[48] Heinrich B, Klier M, Schiller A, Wagner G. Assessing data quality – a probability-based metric for semantic consistency. Decis Support Syst 2018;110:95–106. https://doi.org/10.1016/j.dss.2018.03.011.

[49] Liu C, Talaei-Khoei A, Zowghi D, Daniel J. Data completeness in healthcare: a literature survey. Pac. Asia J. Assoc. Inf. Syst. 2017;9.

[50] Ando RK, Zhang T. A framework for learning predictive structures from multiple tasks and unlabeled data. J Mach Learn Res 2005;6:1817–53.

[51] Bakker B, Heskes T. Task clustering and gating for bayesian multitask learning. J Mach Learn Res 2003;4:83–99.

[52] Baxter J. A model of inductive bias learning. J Artif Intell Res 2000;12:149–98.

[53] Namburete AI, Xie W, Yaqub M, Zisserman A, Noble JA. Fully-automated alignment of 3D fetal brain ultrasound to a canonical reference space using multi-task learning. Med Image Anal 2018.

[54] Ranjan R, Patel VM, Chellappa R. Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Trans Pattern Anal Mach Intell 2017.

[55] Yu J, Zhang B, Kuang Z, Lin D, Fan J. iPrivacy: image privacy protection by identifying sensitive objects via deep multi-task learning. IEEE Trans. Inf. Forensics Secur. 2017;12:1005–16.

[56] Baxter J. A model of inductive bias learning. J Artif Intell ResJAIR 2000;12:3.

[57] Liu Y, Jiang C, Zhao H. Using contextual features and multi-view ensemble learning in product defect identification from online discussion forums. Decis Support Syst 2018;105:1–12. https://doi.org/10.1016/j.dss.2017.10.009.

[58] Gelman A, Jakulin A, Pittau MG, Su Y-S. A weakly informative default prior distribution for logistic and other regression models. Ann Appl Stat 2008;2:1360–83.

[59] Jammalamadaka SR, Qiu J, Ning N. Multivariate bayesian structural time series model. ArXiv Prepr 2018. ArXiv180103222.

[60] Melie-Garcia L, Draganski B, Ashburner J, Kherif F. Multiple linear regression: bayesian inference for distributed and big data in the medical informatics platform of the human brain project. bioRxiv 2018:242883.

[61] Gribling S, de Laat D, Laurent M. Matrices with high completely positive semi-definite rank. Linear Algebra Its Appl. 2017;513:122–48.

[62] Follett L, Yu C. Achieving parsimony in bayesian VARs with the horseshoe prior. ArXiv Prepr 2017. ArXiv170907524.

[63] Lewandowski D, Kurowicka D, Joe H. Generating random correlation matrices based on vines and extended onion method. J Multivar Anal 2009;100:1989–2001.

[64] Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JDF. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med 2004;23:2567–86.

[65] De Mol C, Giannone D, Reichlin L. Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? J Econom 2008;146:318–28.

[66] Lee SK. On generalized multivariate decision tree by using GEE. Comput Stat Data Anal 2005;49:1105–19.

[67] Shih P, Liu C. Face detection using discriminating feature analysis and support vector machine. Pattern Recognit 2006;39:260–76.

[68] Ahmadzadeh F. Change point detection with multivariate control charts by artificial neural network. Int. J. Adv. Manuf. Technol. 2009:1–12.

[69] Liu T, Tao D, Song M, Maybank SJ. Algorithm-dependent generalization bounds for multi-task learning. IEEE Trans Pattern Anal Mach Intell 2017;39:227–41.

[70] García-Laencina PJ, Abreu PH, Abreu MH, Afonoso N. Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values. Comput Biol Med 2015;59:125–33. https://doi.org/10.1016/j.compbiomed.2015.02.006.

[71] Shukla N, Hagenbuchner M, Win KT, Yang J. Breast cancer data analysis for survivability studies and prediction. Comput Methods Programs Biomed 2018;155:199–208. https://doi.org/10.1016/j.cmpb.2017.12.011.

[72] Cunningham P, Delany SJ. k-Nearest neighbour classifiers. Mult. Classif. Syst. 2007;34:1–17.

[73] Kusiak A, Dixon B, Shah S. Predicting survival time for kidney dialysis patients: a data mining approach. Comput Biol Med 2005;35:311–27.

[74] SCAO. National Clinical Terminology Service (NCTS) website [WWW document] URL https://www.healthterminologies.gov.au/ (accessed 12.29.17). 2016.

[75] Sariyar M, Borg A, Pommerening K. Missing values in deduplication of electronic patient data. J Am Med Inform Assoc 2011;19:e76–82.

[76] Ferri C, Hernández-Orallo J, Modroiu R. An experimental comparison of performance measures for classification. Pattern Recognit Lett 2009;30:27–38. https://doi.org/10.1016/j.patrec.2008.08.010.

[77] Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 1993;39:561–77.

[78] Dag A, Topuz K, Oztekin A, Bulur S, Megahed FM. A probabilistic data-driven framework for scoring the preoperative recipient-donor heart transplant survival. Decis Support Syst 2016;86:1–12.

[79] Ghil M, Yiou P, Hallegatte S, Malamud BD, Naveau P, Soloviev A, et al. Extreme events: dynamics, statistics and prediction. Nonlinear Process Geophys 2011;18:295–350.

[80] Demir E. A decision support tool for predicting patients at risk of readmission: a comparison of classification trees, logistic regression, generalized additive models, and multivariate adaptive regression splines. Decis. Sci. 2014;45:849–80.

[81] Ivanović M, Budimac Z. An overview of ontologies and data resources in medical domains. Expert Syst Appl 2014;41:5158–66. https://doi.org/10.1016/j.eswa.2014.02.045.

[82] McGuire MF. Pancreatic Cancer: insights from counterterrorism theories. Decis Anal 2014;11:265–76. https://doi.org/10.1287/deca.2014.0301.

[83] Zandi F. A bi-level interactive decision support framework to identify data mining-oriented electronic health record architectures. Appl Soft Comput 2014;18:136–45. https://doi.org/10.1016/j.asoc.2014.01.001.

[84] Liaw S-T, Rahimi A, Ray P, Taggart J, Dennis S, de Lusignan S, et al. Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. Int J Media Inf Lit 2013;82:10–24.

[85] Zhang M-L, Zhou Z-H. A review on multi-label learning algorithms. IEEE Trans Knowl Data Eng 2014;26:1819–37.

[86] Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus—present and future perspectives. Nat Rev Endocrinol 2012;8:228–36.

[87] White MG, Shaw JAM, Taylor R. Type 2 diabetes: the pathologic basis of reversible β-Cell dysfunction. Diabetes Care 2016;39:2080–8. https://doi.org/10.2337/dc16-0619.

[88] Lin Y-K, Chen H, Brown RA, Li S-H, Yang H-J. Healthcare predictive analytics for risk profiling in chronic care: a Bayesian multitask learning approach. MIS Q 2017:41.

[89] Liu B, Li Y, Ghosh S, Sun Z, Ng K, Hu J. Complication risk profiling in diabetes care: a bayesian multi-task and feature relationship learning approach. IEEE Trans Knowl Data Eng 2019.

[90] Argyriou A, Evgeniou T, Pontil M. Convex multi-task feature learning. Mach Learn 2008;73:243–72.

[91] Zhang Y, Yeung D-Y. A convex formulation for learning task relationships in multi-task learning. ArXiv Prepr 2012:ArXiv12033536.