Multi-disease Predictive Analytics: A Clinical Knowledge-aware Approach

LIN QIU, Department of Information Systems and Analytics, National University of Singapore
SRUTHI GORANTLA, Department of Computer Science, National University of Singapore
VAIBHAV RAJAN and BERNARD C. Y. TAN, Department of Information Systems and Analytics, National University of Singapore

Multi-Disease Predictive Analytics (MDPA) models simultaneously predict the risks of multiple diseases in patients and are valuable in early diagnoses. Patients tend to have multiple diseases simultaneously or develop multiple complications over time, and MDPA models can learn and effectively utilize such correlations between diseases. Data from large-scale Electronic Health Records (EHR) can be used through Multi-Label Learning (MLL) methods to develop MDPA models. However, data-driven approaches for MDPA face the challenge of data imbalance, because rare diseases tend to have much less data than common diseases. Insufficient data for rare diseases makes it difficult to leverage correlations with other diseases. These correlations are studied and recorded in biomedical literature but are rarely utilized in predictive analytics. This article presents a novel method called Knowledge-Aware Approach (KAA) that learns clinical correlations from the rapidly growing body of clinical knowledge. KAA can be combined with any data-driven MLL model for MDPA to refine the predictions of the model. Our extensive experiments, on real EHR data, show that the use of KAA improves the predictive performance of commonly used MDPA models, particularly for rare diseases. KAA is also found to be superior to existing general approaches of combining clinical knowledge with data-driven models. Further, a counterfactual analysis shows the efficacy of KAA in improving physicians' ability to prescribe preventive treatments.

CCS Concepts: • **Applied computing → Health informatics**;

Additional Key Words and Phrases: Electronic health records, diagnosis prediction, rare diseases, multi-label learning, knowledge graph, biomedical literature

ACM Reference format:
Lin Qiu, Sruthi Gorantla, Vaibhav Rajan, and Bernard C. Y. Tan. 2021. Multi-disease Predictive Analytics: A Clinical Knowledge-aware Approach. *ACM Trans. Manage. Inf. Syst.* 12, 3, Article 19 (May 2021), 34 pages.
https://doi.org/10.1145/3447942

This article was supported by Singapore Ministry of Education Academic Research Fund Tier 1 [R-253-000-138-133].
Authors' addresses: L. Qiu (corresponding author), Department of Information Systems and Analytics, National University of Singapore, Computing Drive 15, 117418, Singapore; email: lin_qiu@u.nus.edu; S. Gorantla, Department of Computer Science and Automation, Indian Institute of Science, Mirinji Marg, Mathikere, Bengaluru, Karnataka 560012, Bangalore, India; email: gorantlas@iisc.ac.in; V. Rajan, Department of Information Systems and Analytics, National University of Singapore, Computing Drive 15, 117418, Singapore; email: vaibhav.rajan@nus.edu.sg; B. C. Y. Tan, Department of Information Systems and Analytics, National University of Singapore, Computing Drive 15, 117418, Singapore; email: btan@comp.nus.edu.sg.

2158-656X/2021/05-ART19 $15.00
https://doi.org/10.1145/3447942

1 INTRODUCTION

The ability to identify high-risk individuals prior to the onset of diseases facilitates personalized preventive care and reduces overall rates of morbidity and mortality. Compared with predicting diseases individually, it is more useful and realistic to predict multiple diseases, because patients tend to have multiple diseases simultaneously and patients with a single disease may develop other complications over time. For example, patients with diabetes are likely to have stroke, heart disease, and renal failure [8]. Rare diseases, which affect fewer than 200,000 patients in the U.S., are especially challenging to diagnose and lead to substantial clinical and economic burden [47].

Traditionally, physicians play the role of identifying potential diseases that patients may have, based on patients' health information and physicians' experience and knowledge. However, this process is challenging, because all physicians may not be equally knowledgeable or experienced to identify all possible diseases [2]. Physicians are usually trained in a specialized narrow domain and so, many potential diseases may be out of their scope of expertise, particularly rare diseases. Also, biomedical knowledge is being produced at a rapid rate—more than 0.8M new articles are added annually into the biomedical database MEDLINE [38]—and physicians, busy in their practice, may not always have up-to-date knowledge. Hence, reliable automated decision support tools that can aid physicians in accurate diagnoses is crucial. Diagnostic decision support systems, which use predictive models for diagnoses, have helped physicians play this role since the 1970s [36].

The availability of large-scale **Electronic Health Records (EHR)** has spurred the development of Machine Learning methods for a variety of clinical predictive tasks, including diagnosis prediction [43, 51]. Multi-label and multi-task learning models have been developed for **multi-disease predictive analytics (MDPA)**, i.e., the simultaneous prediction of risks of multiple diseases. Such models can learn correlations across multiple diseases, from historical EHR data, to improve overall prediction. However, there are considerably more patient samples for common diseases (e.g., diabetes) in EHR, for a predictive model to learn from, than for rare diseases (e.g., Hodgkin's disease). Due to lack of sufficient samples, correlations between rare diseases and other diseases can be very difficult to capture.

A common approach to address the problem of insufficient samples is to use auxiliary information from Biomedical **Knowledge Graphs (KG)** to develop combined data-driven knowledge-based models. Knowledge graphs are large, heterogeneous information networks with multiple node types representing clinical concepts (e.g., diseases, drugs) and multiple edge types (e.g., "treats," "predisposes") representing associations between pairs of clinical concepts. These KGs are being actively developed by a combination of manual curation and automatic text mining from biomedical literature. Such combined data-driven knowledge-based models have been found to improve the accuracy and interpretability of diagnosis prediction [12, 34, 52]. However, previous studies have two limitations. Most of the models have been designed to use only a relatively small part of the knowledge graph containing hierarchical information, e.g., ICD disease classification hierarchy [44]. KGs contain considerably more information on, for example, adverse drug events, risk factors, and co-morbidities. Recognizing the need for modeling additional information, some recent works, such as Reference [52], have utilized complete KGs for (single) diagnosis prediction. But, they have specialized architectures that cannot be generalized easily for MDPA. Further, to our knowledge, no previous study has evaluated the efficacy of combined data-driven knowledge-based models for rare disease prediction.

In this article, we address these gaps through two contributions. First, we develop a general approach, called **Knowledge Aware Approach (KAA)**, of using information from knowledge graphs in **Multi-Label Learning (MLL)** models. Unlike previous approaches where KGs are used at training time, our approach uses them to refine the predictions from a trained MLL model. By

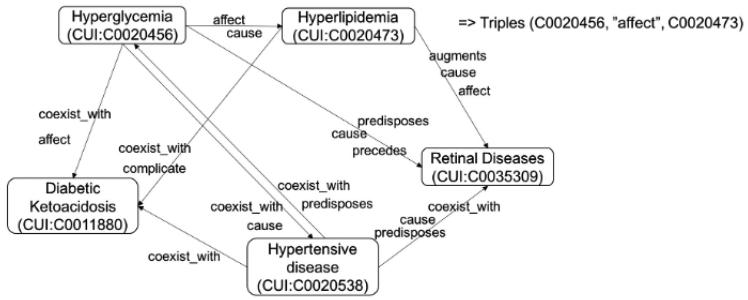


Fig. 1. Portion of Knowledge Graph constructed from SemMedDB triples.

thus decoupling the use of KGs from model training, KAA offers a unique and highly flexible approach that can be used with any underlying data-driven MDPA model. Second, we extensively evaluate the performance of several MLL models, with the use of KAA, for predicting multiple common and rare diseases on a large real EHR dataset. In addition to experiments to compare predictive accuracy, we also conduct a counterfactual analysis to evaluate the utility of predictive models to augment the physician's decision-making ability. Our experiments demonstrate the efficacy of KAA in improving the predictive accuracy of all the MLL models evaluated, particularly for rare diseases. Our KAA model can be found in https://bitbucket.org/q_lin/kaa.git/src.

2 RESEARCH BACKGROUND

2.1 Literature-derived Knowledge Graphs

Biomedical literature is the primary source of clinical knowledge, obtained from research (e.g., clinical trials) as well as experience (e.g., case studies). This literature can be freely accessed from databases such as PubMed, which contains more than 28M articles from MEDLINE [38]. There are ongoing efforts to create standardized vocabularies and ontologies to encode this vast body of biomedical literature. For instance, Metathesaurus from the Unified Medical Language System created by the U.S. National Library of Medicine is the largest thesaurus in the biomedical domain [1] that comprises over 1M biomedical concepts and 5M concept names. The Metathesaurus is organized by concepts represented by **Controlled Unique Identifier (CUI)** for biomedical vocabulary as well as relationships between the concepts such as "causes" or "co-exists with."

Manual creation of ontologies for the enormous amount of biomedical literature is infeasible. Thus, automated and scalable Natural Language Processing systems have been designed to encode knowledge from the primary literature directly [20]. SemMedDB is one such system that extracts "semantic predications" automatically from the titles and abstracts of all PubMed articles [28]. Each semantic prediction consists of a triple (subject, predicate, object). The subject and object are clinical concepts from the Metathesaurus. The predicate indicates a relationship between the subject and the object, such as "treats" or "causes." These triples can be viewed as directed edges and nodes in a knowledge graph, where each node corresponds to a subject or an object, labelled by their CUI, and each directed edge corresponds to a predicate, labelled by the relationship. Figure 1 shows part of a KG from SemMedDB.

SemMedDB includes 30 types of relationships (listed in Appendix A). The corresponding KG is dense, with about 94M predications (edges) [28]. Two clinical concepts may be connected by multiple paths (sequence of one or more edges) representing various direct and indirect relationships. To enable reasoning on such large, complex graphs and to automate feature engineering, knowledge graph embeddings have been developed that yield state-of-the-art results in many graph mining applications. These embeddings encode the global structural properties of the KG into vectorial

representations of its vertices [26, 48]. In our experiments, we use one such method, TransE [4], to obtain vectorial representations of clinical concepts from SemMedDB. We provide more details on TransE in Appendix B. With such representations, similarity between diseases can be computed algebraically using vectorial measures of similarity, e.g., cosine similarity. A high similarity between diseases indicates that they have similar properties in the graph, e.g., similar neighborhood, that in turn would indicate shared clinical relations such as causes, treatments, or risk factors.

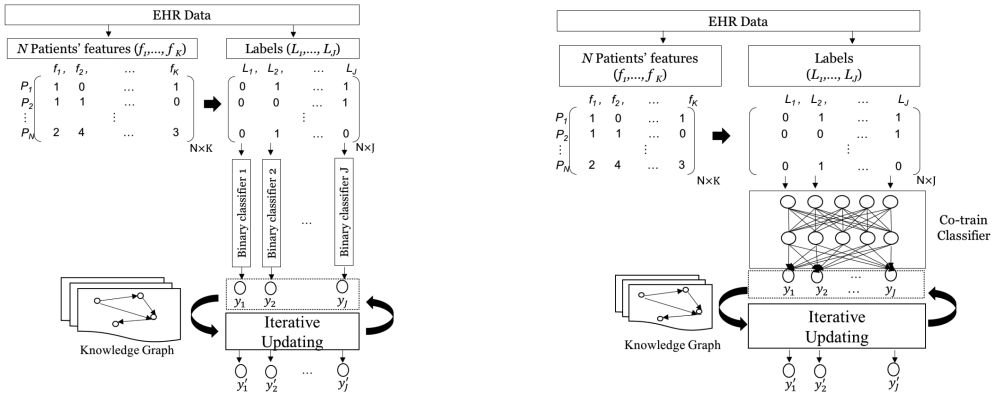
2.2 Multi-disease Predictive Analytics

Wide adoption of EHR in recent years has led to the development of many machine learning models for diagnosis prediction using historical EHR data comprising clinical investigations, drug prescriptions, previous diagnoses, and demographic information, e.g., References [11, 24, 31, 33]. Majority of the previous works on diagnosis prediction has focused on the prediction of specific single diseases, e.g., References [3, 13, 14, 55]. Data-driven models for MDPA can be formulated through **multi-task learning (MTL)** or **multi-label learning (MLL)**. MTL views each patient as an example with multiple tasks (one disease per task) and trains the multiple tasks jointly with shared computational structure to improve learning, e.g., Reference [30]. MLL associates each patient with multiple labels (one label per disease). MLL is applicable for large label spaces while MTL is not, which makes MLL more suitable for health risk modelling, since the number of possible diseases can be large. Furthermore, MTL assumes that every example (patient) is associated with all tasks (diseases), while MLL allows each example (patient) to be associated with a subset of labels (diseases) [54], making MLL a more general model. Hence, most previous works on MDPA have used MLL models [39, 42, 49, 57].

There are numerous ways of building MLL models; a detailed review can be found in Reference [54]. We highlight two categories of methods that are most common in diagnosis prediction and later show how our proposed KAA method improves predictive accuracy in both cases. In the first category a model is trained for each disease, e.g., Binary Relevance [5]. This is equivalent to modeling each disease independently and we denote it by **MDS (model diseases separately)**. In the second category of MLL methods, which we call **MDJ (model diseases jointly)**, the training process is shared, i.e., the likelihood of each disease is learned jointly. Since all the disease labels are used together in the training process, correlations between the diseases are learned well. In both cases, all patient features are utilized during training and the final predictions are given for all diseases simultaneously. See Figure 2.

Using Biomedical Knowledge in Predictive Models. Clinical knowledge is essential to select important features for use in predictive models. To obviate the need for such manual feature engineering, which is cumbersome and time-consuming, most recent approaches have developed "end-to-end" models with minimal or no feature selection. Biomedical knowledge has been utilized in these models during training to improve the predictive accuracy and address the problem of insufficient samples for less common diseases.

The work of Reference [10] was an early deep neural network–based framework that can utilize knowledge-based priors to regularize parameters in their networks. They evaluated their model using only the ICD hierarchy and did not test its predictive accuracy for MDPA. Subsequent work developed more specialized neural architectures to effectively utilize the hierarchical structure of disease ontologies to obtain clinical concept embeddings. For instance, GRAM uses an attention-based mechanism that gives more weight to an ancestor of a clinical concept that has less frequency in the data. This allows it to learn from related diseases guided by the parent-child relations in the disease hierarchy [12]. In GRAM, the ICD hierarchy is only used while learning disease representations, while KAME [34] uses it additionally in generating hospital visit embeddings. Both these



(a) For MDS (b) For MDJ

Fig. 2. Our KAA applied to MLL predictions.

works evaluate next-visit diagnoses prediction, but they do not specifically evaluate accuracy of rare disease prediction. Further, both these models can use only hierarchical information from disease ontologies and not the entire KG. Recently, DG-RNN was proposed by Reference [52], which uses KG embeddings to obtain attention weights for each clinical concept used for prediction. They initialize the embeddings with TransE embeddings from the entire KG and then fine-tune the embeddings using neighbors (from the KG) of each clinical concept found in the data. For this crucial step of neighborhood selection, only disease nodes and two kinds of edges that are most relevant to heart failure prediction are used from the KG. Without such task-specific selection, each clinical event will have a very large number of potentially less useful neighbors in the dense KG, leading to difficulties in training and loss of interpretability. Further, their DG-RNN has been designed for risk prediction of a single disease (heart failure) and not evaluated for MLL, which may require additional model development, as discussed in Reference [31].

Among these approaches the most general approach, with respect to use of KGs, is that of Reference [10]. Their model can use auxiliary information of label similarity from any source, including knowledge graph embeddings. We denote this method by MDJ-Regu and provide more details in Appendix C. All the previous approaches are closely tied to the underlying data-driven model architectures. MDJ-Regu can only be used with neural network based MDPA models. GRAM, KAME and DG-RNN have even more specialized neural architectures, as described above.

Another general approach of using auxiliary knowledge in predictive models is to use pretrained embeddings as features. This has been used in many applications, not just in MDPA. For example, knowledge graph embeddings of drugs and diseases have been used to predict adverse drug events [35]. Such an approach can be used with both MDS and MDJ, and we call them MDS-Embed and MDJ-Embed, respectively, where the embeddings are obtained using TransE.

In contrast to all these models, where KGs are used during model training, our proposed KAA is a general framework that uses KGs to refine the predictions from an MLL model. Hence, it can be applied on the predictions obtained from any MLL model.

3 MODEL DEVELOPMENT

We now describe our KAA, which takes disease prediction probabilities from any MLL method as input and refines them using a knowledge graph to produce new prediction probabilities.

3.1 Notation and Problem Setting

Let N be the number of patients in the training data, K be the dimensionality of the feature set, and J be the total number of labels. Each patient is associated with a K-dimensional feature vector that encodes patients' features, such as medications and laboratory investigation results, extracted from historical EHR data. Each patient is diagnosed with a set of diseases and the J-dimensional binary label vector encodes this information: 1 indicates presence of a disease and 0 indicates absence. We denote the label for the jth disease by y_j $(1 \leq j \leq J)$.

After an MLL method is trained on historical data of N patients, it can predict the probabilities of the J labels given any K-dimensional feature vector x of a patient. In the MDS category, independent classifiers g_j $(1 \leq j \leq J)$ are learned separately, e.g., through Binary Relevance. After training, g_j can produce the probability P_j of the jth disease (i.e., the probability of label $y_j = 1$). In the MDJ category, a shared function C is learned, e.g., through a neural network. The final predictions are obtained through $\delta_j(C(x))$, where δ_j is the function that takes the outcome of $C(x)$ and outputs the prediction probability of the jth disease as well. In this manner, a jointly trained MLL model can predict the probabilities P_j (for $1 \leq j \leq J$) for all the J diseases simultaneously. See Figure 2.

3.2 KAA: Leveraging Biomedical Literature during Multi-label Prediction

The key idea of KAA is that the predictions for two similar diseases can be fine-tuned based on their similarity obtained from the knowledge graph to reduce the divergence between their predictions. If two diseases are not similar, then we rely primarily on data-driven probabilities. KAA iteratively applies this to all pairs of diseases. Similar iterative procedures have been used in other contexts, e.g., Reference [17].

Let $S_{d,d'}$ be the similarity between the knowledge graph embeddings of diseases d and $d^{'}$.

To impose the condition that two diseases with high similarity tend to have similar occurrence probabilities, we define the following loss function where the probabilities $P_j : j \in J$ represent purely data-driven predictions from any existing algorithm for MDPA and the probabilities $P_j^{'}, P_k^{'}$: $j \in J, k \in J$ represent our proposed knowledge-refined predictions.

$$f(P_j^{'}) = (1-\epsilon) \sum_{j=1}^{J} \sum_{k=1}^{J} S_{j,k}(P_j^{'} - P_k^{'})^2 + \epsilon \sum_{j=1}^{J} \|S_{j,*}\|_1 (P_j^{'} - P_j)^2 \tag{1}$$

Equation (1) consists of two terms. The first term captures the constraint on disease similarity. For a pair of diseases, j and k, with large similarity value $S_{j,k}$, minimizing $f(P_j^{'})$ will lead to smaller difference between the knowledge-refined probabilities $P_j^{'}$ and $P_k^{'}$. However, a small similarity value $S_{j,k}$ will not enforce such a constraint on the difference between the knowledge-refined probabilities. The second term is used to ensure that knowledge-corrected probabilities do not deviate too much from the data-driven predictions. This can be viewed as a form of regularization, as we do not want this optimization to completely overrule the model learned during training. Hence, the squared difference $(P_j^{'} - P_j)^2$ is minimized with a coefficient $\|S_{j,*}\|_1 = \sum_{k=1}^{J} |S_{j,k}|$ to ensure that both terms have comparable coefficients. This coefficient prevents the loss function from attaching more importance to the first term over summations involving $P_j^{'} : j \in J$ when $\|S_{j,*}\|_1$ is large. The tradeoff between the first and second terms in Equation (1) is controlled by a hyperparameter $\epsilon \in (0, 1)$. To minimize Equation (1), we find its critical point:

$$\frac{\partial(f(P_j^{'}))}{\partial(P_j^{'})} \propto (1-\epsilon) \sum_{j=1}^{J} \sum_{k=1}^{J} S_{j,k}(P_j^{'} - P_k^{'}) + \epsilon \sum_{j=1}^{J} \|S_{j,*}\|_1 (P_j^{'} - P_j). \qquad \forall j \in J \tag{2}$$

Setting the above to zero, we obtain the refined prediction for disease j, which consists of the original prediction P_j and a "weighted-average" of optimal $P_k^{'}$, $(\forall k \in J)$:

$$P_j^{'} = (1-\epsilon)\frac{\sum_{k=1}^{J} S_{j,k}, P_k^{'}}{\|S_{j,*}\|_1} + \epsilon P_j. \tag{3}$$

We note that the second-order partial derivative, shown in Equation (4), is greater than zero for positive similarity values ensuring that the critical point is a minimum.

$$\frac{\partial^2(f(P_j^{'}))}{\partial^2(P_j^{'})} = (1-\epsilon)\sum_{j=1}^{J}\sum_{k=1}^{J} S_{j,k} + \epsilon \sum_{j=1}^{J} \|S_{j,*}\|_1 = \sum_{j=1}^{J}\sum_{k=1}^{J} S_{j,k} \tag{4}$$

We then obtain an iterative process that allows us to input a set of knowledge-refined predictions to obtain another set of updated predictions, as shown in Equation (5). We initialize $P_j^{'(t=0)}$ (t is the iteration index) with the data-driven predictions P_j (i.e., $P_j^{'(0)} = P_j$). We repeat this step for all J diseases to obtain $P_j^{'(t=1)}$, $1 \le j \le J$ and iteratively continue until convergence.

$$P_j^{'(t)} = (1-\epsilon)\frac{\sum_{k=1}^{J} S_{j,k}, P_k^{'(t-1)}}{\|S_{j,*}\|_1} + \epsilon P_j \tag{5}$$

Convergence is determined by no decrease in loss in successive iterations, which typically occurs in 10–15 iterations in our experiments. The time complexity of obtaining the refined predictions for all J labels, for each patient, is $O(J^2T)$ where T is the total number of iterations.

To calculate disease similarities, we use knowledge graph embeddings from TransE [4] in our experiments. We provide a brief overview of TransE in Appendix B. TransE outputs k-dimensional vector representations (where k is a hyperparameter, set to 200 in our experiments) of all the clinical entities in the graph. Thus, we obtain a 200-dimensional vector representation for each disease. The similarity between two disease vectors d and $d^{'}$ is computed using the Gaussian kernel, $S_{d,d'} = \exp(-\gamma\|d - d^{'}\|)^2$, where exp denotes the exponential function and γ is a hyperparameter. This similarity value is always positive, thereby ensuring a minima in Equation (4).

Figure 2 shows a schematic of how our KAA approach can be combined with MLL methods. More details on the classifiers MDS and MDJ used in our experiments are in Appendix D. Appendix E has a summary of all the steps in KAA described above, including how each step is applied in our experiments.

4 EXPERIMENT SET UP

In this study, experiments are conducted using de-identified data from the MIMIC-III dataset [27], a publicly available EHR dataset of 46,520 patients, collected from intensive care units of the Beth Israel Deaconess Medical Center in Boston, USA, between 2001 to 2012. To evaluate our method in not only common diseases but also rare diseases, we choose the 20 most common diseases and the 10 rarest diseases from our EHR dataset. The occurrence distribution of the selected diseases in our dataset is given in Table 1. The **imbalance ratio (IR)** that reflects the imbalance among diseases (i.e., the rarity of diseases in this dataset) is computed as below [32], where $|D_j^+|$ and $|D_j^-|$ are the number of positive and negative samples, respectively, for disease j.

$$IR_j = \frac{max(|D_j^+|, |D_j^-|)}{min(|D_j^+|, |D_j^-|)}$$

Additional details of the selected diseases are in Appendix F.

Table 1. Summary of Occurrences of the 30 Selected Diseases (Appendix F Has More Details)

Disease No.	Short Name	Number (Percent) of Occurrences	Imbalance ratio	Disease No.	Short Name	Number (Percent) of Occurrences	Imbalance ratio
1	SEP	6,925 (14.9%)	5.7	16	NSD	6,924 (14.9%)	5.7
2	ARF	9,549 (20.5%)	3.9	17	HVD	6,627 (14.2%)	6.0
3	CHF	10,432 (22.4%)	3.5	18	ED	6,374 (13.7%)	6.3
4	PNE	6,558 (14.1%)	6.1	19	HD*	128 (0.3%)	362.4
5	FED	12,731 (27.4%)	2.7	20	RF	9,884 (21.2%)	3.7
6	LD	4,156 (8.9%)	10.2	21	MM*	154 (0.3%)	301.1
7	HCSH	4,889 (10.5%)	8.5	22	CKD	5,039 (10.8%)	8.2
8	DA	9,540 (20.5%)	3.9	23	SLE*	311 (0.7%)	148.6
9	MD	4,405 (9.5%)	9.6	24	NECS*	513 (1.1%)	89.7
10	EH	17,924 (38.5%)	1.6	25	CIBD*	421 (0.9%)	109.5
11	CAH	13,115 (28.2%)	2.5	26	MOS*	406 (0.9%)	113.6
12	DWC	8,064 (17.3%)	4.8	27	MS*	222 (0.5%)	208.5
13	DLM	12,250 (26.3%)	2.8	28	CNS*	269 (0.6%)	171.9
14	DMC	3,414 (7.3%)	12.6	29	LYM*	214 (0.5%)	216.4
15	CD	13,705 (29.5%)	2.4	30	CRA*	185 (0.4%)	250.5

*indicates rare disease.

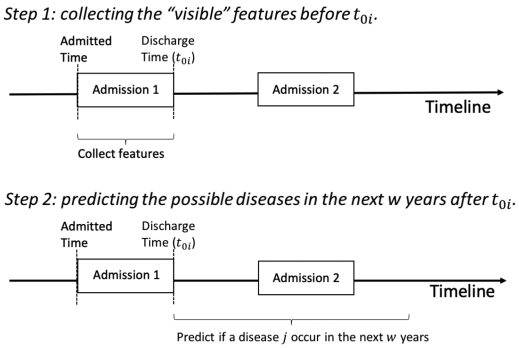


Fig. 3. Our experiment design.

Similar to previous works on MDPA [12, 34], we use clinical data from previous admission episodes to predict diagnoses in future episodes, since the granularity of timestamps for many clinical investigations in MIMIC-III is more reliable at an episode-level than at a daily or hourly level. Hence, patients with single episodes are excluded and we use data of 7,537 patients, who have two or more episodes.

Figure 3 illustrates our experiment design. For each patient i $(i = 1, \ldots, N)$, we set the discharge time of the first admission as t_{0i} and use the features collected during the first admission episode as predictors, all of which are the "visible" information for the time after the first admission. We

Table 2. Disease Occurrences across Time Windows

Short Name	Before t_{oi}	During (cumulative)				
		Year=w1	Year=w2	Year=w3	Year=w4	Year=w5
SEP	1,014	1,120	1,418	1,599	1,724	1,825
ARF	1,625	1,371	1,760	2,006	2,187	2,327
CHF	2,175	1,631	2,048	2,337	2,529	2,657
PNE	1,115	969	1,236	1,391	1,513	1,599
FED	2,132	1,778	2,252	2,551	2,763	2,917
LD	782	606	766	864	932	982
HCSH	983	816	1,077	1,263	1,374	1,479
DA	1,720	1,433	1,824	2,068	2,228	2,354
MD	727	559	712	809	881	928
EH	3,025	1,832	2,274	2,546	2,781	2,916
CAH	2,310	1,407	1,800	2,058	2,253	2,384
DWC	1,466	1,026	1,300	1,462	1,579	1,652
DLM	1,861	1,258	1,612	1,811	1,998	2,110
DMC	817	543	699	805	886	939
CD	2,348	1,740	2,148	2,407	2,586	2,730
NSD	1,053	844	1,091	1,246	1,371	1,471
HVD	1,182	770	979	1,113	1,211	1,283
ED	1,031	801	1,041	1,194	1,309	1,390
HD	30	26	27	29	30	30
RF	1,585	1,378	1,727	1,937	2,071	2,184
MM	27	23	28	34	35	39
CKD	958	842	1,125	1,309	1,440	1,559
SLE	88	63	72	81	82	84
NECS	75	64	81	95	103	109
CIBD	94	73	93	101	104	111
MOS	74	56	76	80	82	85
MS	50	30	38	45	51	52
CNS	48	37	41	44	47	50
LYM	39	24	37	38	40	47
CRA	36	25	30	34	37	40

predict the status of y_i^j $(j = 1, \ldots, J)$, which is defined by whether patient i is diagnosed with the jth disease from the label set of $J = 30$ diseases in the next w years after t_{0i}. We vary w from one to five years to examine if different time windows affect our predictive performance.

Table 2 shows the distribution of diseases for different time windows. The first column ("before t_{0i}") has the distribution of diseases in the first admission episode, t_{0i}. The second column shows the distribution of diseases that occur within $w = 1$ year after t_{0i}. Subsequent columns for w years $(w = 2, 3, 4, 5)$ after t_{0i} shows the distribution of diseases in a cumulative manner.

A total of 5,062 features are extracted from patient demographics, lab tests, procedures, medications, and known diseases. These features are from the hospital record system, which contains measurements for the entire hospital stay, including those in non-ICU care, of these patients [27]. Patient demographics, such as gender, race, marital status, and age are encoded as binary variables (age is binarized into intervals of 10 years, $(0 - 10, \ldots, 70 - 80$, and $\geq 80)$.

Table 3. Summary of Features Used

Feature Category	No. of Columns	Examples
Demographics	76	Female, married, white
Lab tests	266	Hematocrit, Eosinophils
Procedures	1,181	ICD-9-CM procedure codes (e.g., 3727 for cardiac mapping)
Medications	3,271	"Warfarin, 5mg Tablet"
Diagnoses	268	ICD-9-CM diagnosis codes (e.g., 5,848 for acute kidney failure)

Each medication is encoded as a binary variable indicating whether or not that medication was given. Each lab-test is numerically encoded as two variables (the number of times the test was taken and the number of times the test result was abnormal). Each procedure is encoded as a variable (the number of times the procedure was conducted). The disease diagnosis variables that are known at t_{0i} are also included as binary predictors. The features used are summarized in Table 3.

We conduct two sets of experiments. In the first set, we compare the improvement in predictive accuracy when KAA is used with multiple MLL models. The selected MLL models include:

(1) Neural network based models for MDS and MDJ (described in Section 2.2).
(2) Classifiers CC-J48 [40] and ML-KNN [53] that were reported to be the best-performing MLL methods on this dataset [57].

The aim of the first experiment is to evaluate whether the use of KAA improves the predictive performance of these MLL models. In addition, we also evaluate two other methods of incorporating general clinical knowledge in MDPA models:

(1) MDJ-Regu [10]: the closest previous approach that uses disease similarities as a regularizer in training MDJ models.
(2) MDS-Embed and MDJ-Embed: Following the common approach of using pretrained embeddings as features, we use KG embeddings of diseases, diagnosed at the first admission as additional inputs in MDS and MDJ.

Since neural networks may overfit such high-dimensional data, we use lasso regularization for MDS and MDJ in all cases, i.e., both with and without knowledge infusion through KAA, knowledge-based regularization (-Regu), and feature embeddings (-Embed). For a fair comparison, we have used exactly the same neural network structure in MDJ, MDJ-Regu, MDJ-Embed, and MDJ-KAA. Similarly, the network structure of MDS, MDS-Embed, and MDS-KAA are identical. We also provide the same disease similarity values, obtained from TransE, to MDJ-Regu, MDJ-Embed, and MDJ-KAA. In MDJ-Embed, multiple diseases, if present in the input diagnoses, are represented by the element-wise sum of the individual disease embeddings. More details of MDS and MDJ networks are in Appendix D.

All the methods are compared using 10-fold **cross-validation (CV)** to estimate the predictive performance [22]. Since there are multiple labels, we use the micro-AUC, which computes the **Area Under the ROC Curve (AUC)** by considering the predictions for all the labels together and is preferred when there is class imbalance [50].

In the second set of experiments, we demonstrate the practical impact of KAA by counterfactually analyzing how healthcare predictive models can augment clinicians' capability in identifying high-risk patients and providing preventive treatments to reduce their risks.

Table 4. Mean Micro-AUC, with Standard Deviations: Prediction of All Diseases across 5 Time Windows

Time Window	w1	w2	w3	w4	w5
MDS	0.6263 ± 0.0094	0.6293 ± 0.0034	0.6408 ± 0.0062	0.6483 ± 0.0038	0.6496 ± 0.0054
MDS-Embed[1]	0.6259 ± 0.0016	0.6419 ± 0.0033**	0.6529 ± 0.0078**	0.6542 ± 0.0027**	0.6521 ± 0.0045
MDS-KAA[1]	**0.6642 ± 0.0118****	**0.6744 ± 0.0039****	**0.6850 ± 0.0041****	**0.6889 ± 0.0066****	**0.6919 ± 0.0079****
MDJ	0.6885 ± 0.0192	0.6930 ± 0.0099	0.7075 ± 0.0055	0.7157 ± 0.0116	0.7148 ± 0.0082
MDJ-Embed[2]	0.6763 ± 0.0139	0.6827 ± 0.0136	0.6951 ± 0.0089	0.6922 ± 0.0103	0.7051 ± 0.0087
MDJ-Regu[2]	0.6704 ± 0.0173	0.6947 ± 0.0099	0.7051 ± 0.0074	0.7175 ± 0.0090	0.7189 ± 0.0091
MDJ-KAA[2]	**0.7142 ± 0.0138****	**0.7261 ± 0.0094****	**0.7295 ± 0.0100****	**0.7274 ± 0.0099****	**0.7348 ± 0.0063****
CC-J48	0.7451 ± 0.0157	0.6333 ± 0.0061	0.6782 ± 0.0231	0.6944 ± 0.0048	0.7041 ± 0.0025
CC-J48-KAA[3]	**0.7746 ± 0.0165****	**0.6870 ± 0.0110****	**0.7260 ± 0.0229****	**0.7349 ± 0.0049****	**0.7427 ± 0.0025****
ML-KNN	0.7248 ± 0.0037	0.7286 ± 0.0021	0.7326 ± 0.0013	0.7365 ± 0.0016	0.7370 ± 0.0040
ML-KNN-KAA[4]	**0.7244 ± 0.0040**	**0.7281 ± 0.0017**	**0.7320 ± 0.0013**	**0.7362 ± 0.0017**	**0.7368 ± 0.0041**

**$p < 0.05$. Baseline used for comparison: [1]MDS, [2]MDJ, [3]CC-J48, [4]ML-KNN.

Table 5. Mean Micro-AUC, with Standard Deviations: Prediction of Rare Diseases across 5 Time Windows

Time Window	w1	w2	w3	w4	w5
MDS	0.5031 ± 0.0101	0.5088 ± 0.0044	0.5041 ± 0.0221	0.5072 ± 0.0019	0.5095 ± 0.0016
MDS-Embed[1]	0.5243 ±0.0194**	0.5179 ± 0.0169	0.5195 ± 0.0121*	0.5122 ± 0.0092	0.5161 ± 0.0099
MDS-KAA[1]	**0.5385 ± 0.0205****	**0.5252 ± 0.0078****	**0.5253 ± 0.0178****	**0.5173 ± 0.0058****	**0.5290 ± 0.0139****
MDJ	0.6521 ± 0.0299	0.6464 ± 0.0169	0.6559 ± 0.0184	0.6496 ± 0.0297	0.6607 ± 0.0149
MDJ-Embed[2]	0.6594 ± 0.0321	0.6511 ± 0.0235	0.6567 ± 0.0222	0.6549 ± 0.0219	0.6697 ± 0.0292
MDJ-Regu[2]	0.6560 ± 0.0438	0.6475 ± 0.0398	0.6525 ± 0.0179	0.6550 ± 0.0355	0.6504 ± 0.0151
MDJ-KAA[2]	**0.6627 ± 0.0275****	**0.6638 ± 0.0178****	**0.6704 ± 0.0197****	**0.6630 ± 0.0296****	**0.6770 ± 0.0193****
CC-J48	0.8249 ± 0.0379	0.5723 ± 0.0366	0.6257 ± 0.0267	0.6361 ± 0.0393	0.6281 ± 0.0275
CC-J48-KAA[3]	**0.8392 ± 0.0349****	**0.5897 ± 0.0282****	**0.6367 ± 0.0269***	**0.6542 ± 0.0499***	**0.6440 ± 0.0317****
ML-KNN	0.5576 ± 0.0233	0.5543 ± 0.0218	0.5564 ± 0.0350	0.5608 ± 0.0462	0.5498 ± 0.0488
ML-KNN-KAA[4]	**0.5891 ± 0.0338****	**0.5888 ± 0.0305****	**0.5942 ± 0.0370****	**0.5890 ± 0.0602****	**0.5868 ± 0.0491****

*$p < 0.1$, ** $p < 0.05$. Baseline used for comparison: [1]MDS, [2]MDJ, [3]CC-J48, [4]ML-KNN.

5 EXPERIMENTAL RESULTS

5.1 Improvement in Predictive Accuracy

Tables 4 and 5 show the mean and standard deviation (over 10-fold CV) of the micro-AUC obtained by all the methods for all the diseases and the rare diseases, respectively. They are broken down by the length of the time window, w. Results at the level of disease labels (i.e., for each of the 30 disease labels) at $w1$ are presented in Appendix G.

Both MDS-KAA and MDJ-KAA models significantly outperform MDS and MDJ models, respectively, for all five time windows. This shows that the use of KG through KAA indeed improves the predictive performance for all five time windows irrespective of whether multiple diseases are modeled separately or jointly. In addition, we notice that the performance increase from MDS to MDS-KAA model is larger than that from MDJ to MDJ-KAA model. This is understandable, since MDS models do not learn the correlations between diseases during model training, while MDJ models do learn these data-driven correlations due to co-training. Thus, the incremental improvement due to KAA is more for MDS models. More interestingly, our KAA approach improves the performance of MDJ as well where data-driven correlations have been learned by the model during

training. This clearly illustrates the benefits of additional information from clinical knowledge that is added through KAA for both overall predictive performance and, in particular, for rare diseases.

We observe that the performance values of MDJ and MDJ-Regu model are similar: There is no significant difference in the micro-AUC values in all cases. These results demonstrate that the regularization method employed in MDJ-Regu is not effective in improving predictive performance. Similar findings, that there is limited increase in predictive performance, was reported by Reference [10] when the ICD-9 hierarchy was used to obtain disease similarities in their experiments. Incorporating the knowledge embeddings as inputs, through MDS-Embed, improves the prediction performance of MDS for all diseases as well as rare diseases. However, the improvement in performance from MDS-KAA is higher. In the case of MDJ, using embeddings as inputs through MDJ-Embed does not outperform MDJ-Regu when all diseases are considered. In the case of rare disease, the performance of MDJ-Embed is comparable or marginally better than MDJ-Regu. In both cases, MDJ-KAA outperforms MDJ-Embed. Note that in our experiments, MDJ, MDJ-Embed, MDJ-Regu, and MDJ-KAA use the same co-training model (neural network) and use the same disease similarities obtained from the KG. These results demonstrate that the performance improvement observed is not just due to the use of similarities from KG embeddings but also the way it is used in KAA.

The results for the other classifiers show that KAA improves the performance of CC-J48 for all diseases as well as rare diseases across all five time windows. For ML-KNN, KAA is able to improve the predictive performance in predicting rare diseases, and for all diseases, the performance is comparable to the case without KAA. A paired sample t-test [56] at significance level 0.05, considering all the time windows over all 10 folds, was conducted to evaluate the significance of performance improvement. We find that performance improvements in rare disease prediction due to the use of KAA are statistically significant for all evaluated MLL models. In the case of all diseases, the use of KAA results in significant improvements over MDS, MDJ, MDJ-Embed, MDJ-Regu, and CC-J48. These results illustrate the generality of our KAA method and performance gain especially for rare diseases.

5.2 Counterfactual Analysis

For clinical predictive models, it is vital that "prediction can lead to actions that reduce risk beyond what would occur without the prediction rule" [19]. The practical value of a predictive model is reflected by its ability to facilitate preventive treatments for high-risk patients who otherwise would not get such interventions. Ideally, the impact of a predictive model should be evaluated through a randomized clinical trial on two comparable case and control groups. However, clinical trials are extremely time-consuming and expensive and so, very few clinical predictive models can be evaluated through such trials.

An alternative, proposed by Reference [30], to quantify the utility of a predictive model in improving decision-making, is through a counterfactual analysis. This is done by constructing a counterfactual table, as shown in Figure 4, for each predicted disease (or a group of diseases). We consider all patients who were actually diagnosed with the disease in the period between t_{i1} and $t_{i1} + 5$ and divide them into four groups. The variables a, b, c, and d in the table denote the number of patients in each of the four groups that satisfy the conditions described below.

Such a counterfactual table can offer several insights. The values a and d represent the events that are correctly detected by only the physicians (physician utility) and only our predictive model (model utility), respectively. Values b and c indicate consistency between the physicians' decision and our model's prediction—b: both are correct (positive consistency); c: both are incorrect (negative consistency). Thus, we view a and c as the model's errors (these are the cases where the model predicts "low risk" but the patients are diagnosed with the disease later) and c and d as the

		Predicted Risk (from a model)	
		Low	High
Preventive treatment prescribed at/before t_{io}	Yes	a	b
	No	c	d

a. Physician utility: the number of events captured by physician, not by model
b. Positive consistency: the number of events captured by both physician and model
c. Negative consistency: the number of events captured by neither physician nor model
d. Model utility: the number of events captured by model, not by physician

Fig. 4. Counterfactual table.

physician's limitations (no preventive treatment given to patients who were diagnosed later with the disease). In particular, c should be low, considering that this is the blind area in which both physicians and predictive models cannot recognize the potential risks.

In our case, the most important value is d; a higher value indicates greater model utility in enhancing physician's capability. Smaller values of a and c indicate fewer prediction errors and are desirable. To compare two models, a model can be considered more valuable than another if it yields a smaller value for both a and c and a larger value for d. Constructing this table requires (1) identifying the cases where physicians recognized future risks (and possibly prescribed preventive interventions) and (2) determining a threshold for distinguishing high and low predicted risks from the considered model. We describe below how they both were determined in our experiments.

EHR data does not have information about whether or not the physician actually recognized the risk of a disease and advised preventive interventions. Moreover, for these 30 diseases, which includes rare diseases, it is difficult for physicians to recognize future risks and prescribe preventive interventions. For a rigorous evaluation of our model, we overestimate the physician's ability by assuming that if there is a known relation in the biomedical literature (as seen in the knowledge graph extracted from the literature), then the physician recognizes it and prescribes the appropriate preventive intervention. No assumptions are made regarding what interventions are provided. The exact nature of the intervention is not required for our evaluation, since we are only evaluating the binary decision of whether or not an intervention was prescribed. The relations considered from the biomedical literature are listed in Appendix H and include 14 relations such as "predisposes," "precedes," "coexists with," and "causes." So, if a patient has disease A and disease B is known to be related to disease A by any of these 14 relations, then we assume that the physician would prescribe relevant interventions to the patient to prevent disease B. E.g., diabetes (A) is a known to induce hypertension (B), then we assume that preventive intervention (e.g., lifestyle modifications) for disease B is advised.

We consider multiple thresholds (10%, 8%, and 6%) for distinguishing high and low predicted risks from the predictive model. From previous medical literature, we find that it is common to use 20% risk over 10 years as a cutoff between high- and low-risk patients (e.g., Reference [29]). Previous works, e.g., References [16, 30], proportionally use 10% risk cut-off for 5-year span. Hence, in our case considering the 4-year span, we can proportionally use 8% risk as the cut-off. Once again, we make the evaluation more stringent by using a 10% threshold. With a 10% threshold, a patient is considered to be at high-risk of a disease if the predicted probability of the disease is higher than 0.1. Using a lower risk threshold would result in the model predicting more patients at high risk. This would increase the values of b and d in the counterfactual table, thereby showing higher improvement over physician ability. We illustrate this at two other risk thresholds: 6% and 8%.

Figure 5 presents the aggregated counterfactual results for all the 30 diseases and separately for the 10 rare diseases using risk threshold of 10% for the predictive models. We compare the results from MDS (MDJ) with the results from MDS-KAA (MDJ-KAA) to illustrate the practical impact

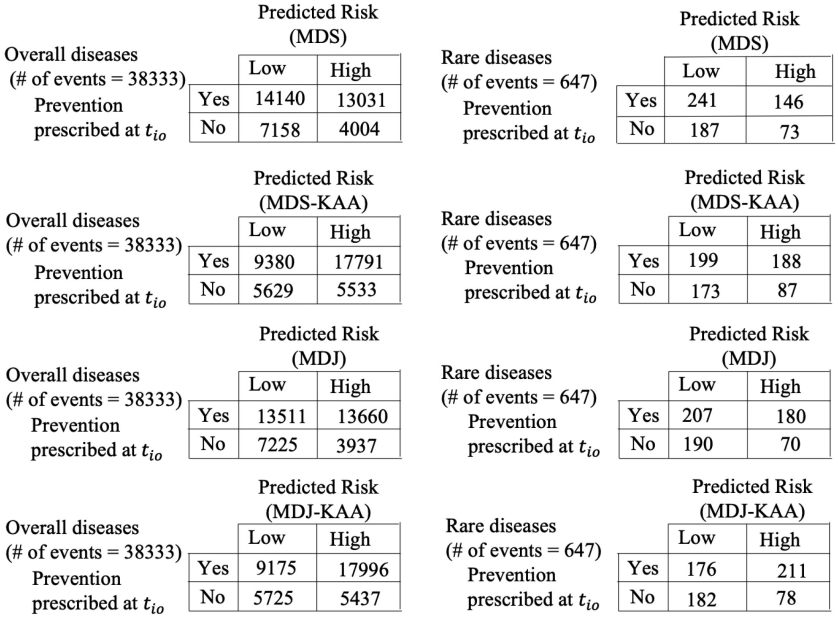


Fig. 5. Results of counterfactual analysis (10% cut-off).

of incorporating clinical knowledge. The complete results for each of the 30 diseases are shown in Appendix H. Despite the assumed overestimate of physician capability, all values for d are nonzero, indicating that both MDS and MDJ frameworks provide useful insights for augmenting physicians' decision-making (regardless of diseases, time windows, and models). Notably, both MDS-KAA and MDJ-KAA models that incorporate clinical knowledge outperform the vanilla MDS and MDJ models without knowledge by producing a much smaller value for both a and c (causing fewer errors) and a much larger value for d (augmenting physician's capability). These results demonstrate the efficacy of using our KAA method to leverage clinical knowledge for both common and rare diseases.

Note that our assumptions impose a stringent standard and the results are a conservative underestimate of the model abilities. Since we have overestimated physician ability, the values of a and b in our counterfactual tables are expected to be much lower in practice than what we have assumed. Lower values of a and b would lead to higher values of c and d, indicating better model performance compared to the values of d shown now. Moreover, by choosing higher risk thresholds for the predictive models, we underestimate the model capability. The results in Appendix H for other risk thresholds (8% and 6%) show the expected increase in model capability. In particular, for rare diseases, high-risk thresholds should be much lower than 10%. If we reduce the thresholds to below 6%, then we expect the model performance to improve even further. We observe improvements even in this rigorous evaluation regime and expect better performance in less demanding conditions.

6 DISCUSSION AND IMPLICATIONS

6.1 Contributions to the Literature

This study contributes to existing knowledge base in several ways. First, we design a novel approach to augment MDPA models that can significantly enhance physicians' capabilities for decision-making. MDPA models combined with KAA can capture and utilize correlations between

diseases from not only data-driven models but also KGs derived from biomedical literature. We validate our KAA approach using real EHR data. Empirical results reveal that the predictive performance of several MDPA models significantly improve with the use of KAA. Overall improvement in predictive performance is seen for all the diseases tested, including rare diseases that are harder to diagnose.

Second, our optimization procedure is flexible and generalizable. It can be combined with any existing MLL approach and can be used with any knowledge graph. Unlike previous approaches that are applied during model training and require modifications in the training process, KAA is applied on the initial predictions obtained from data-driven predictive models. Thus, it can be used with any MDPA model and can easily incorporate the latest knowledge from the biomedical literature without re-training of the MDPA model. In addition, our KAA model does not use a limited subset of biomedical literature (e.g., ICD-9 hierarchy) but associations derived from the entire literature. Moreover, since these associations are automatically mined, the KAA approach can scale well with the fast-growing body of biomedical literature.

Finally, we establish several design principles (i.e., corresponding to the steps in our KAA approach) for MDPA. These design principles can be applied when building an MLL framework to refine the predictions obtained from purely data-driven predictive models through the optimization process using similarities generated from a KG. Such prescriptive knowledge advanced in this study, as "nascent design theory" [21], is generalizable to other contexts of integrating human knowledge with predictive models.

6.2 Managerial and Practical Implications

Healthcare resources in hospitals are always limited and have to be utilized effectively for maximum impact on patients' welfare. To do so, hospitals need to not only design the best treatment plans for current conditions of patients but also provide preventive interventions to alleviate potential future problems of patients. Over-investing in the provision of preventive interventions leads to wastage of hospital resources, while under-investing in the provision of preventive interventions potentially increases morbidity rates. Accurate MDPA models allow hospitals to prescribe appropriate levels of preventive interventions for each patient, thereby alleviating the problem of over-investing or under-investing in the provision of preventive interventions. From a patient's perspective, if they are made aware of their risks of various diseases in advance, then they can take actions to reduce the chances of occurrence of these diseases or reduce the severity of their effects, such as through self-management of their daily lives.

Early diagnosis can also lead to significant financial savings. A study revealed that having one additional chronic condition, on average, can cause a patient to incur an additional $2,000 in healthcare spending per year [18]. Common diseases (e.g., Diabetes mellitus and Hypertension) have very large patient size and broad societal impact. For instance, as of 2015, 415M people are estimated to have diabetes worldwide [25] and diabetic patients in the United States alone incur a medical cost of $322B per year [8]. Rare diseases are more difficult to identify and may be life-threatening. It can take over seven years, on average, to accurately detect rare diseases [46], leading to prolonged clinical and economic burden.

Although physicians are highly trained professionals in their specific care domain, rare diseases may be overlooked or may be out of the scope of their expertise. Biomedical literature is being created at a rapid rate, and physicians may not have enough time beyond their daily practice to review new knowledge. By leveraging such biomedical literature in a scalable manner, our KAA framework significantly strengthens the capability of predictive systems to offer physicians valuable clinical decision support during the care-giving process.

6.3 Limitations and Future Research

KAA has been evaluated using predictions that were refined using disease similarities using KG embeddings, where other clinical relationships were indirectly modeled. Future studies can consider other relationships (e.g., drug-disease relationships) and extend KAA to directly incorporate these relationships. Our empirical results are obtained based on analyses performed using one EHR dataset. Future work can validate KAA using additional data. Finally, KAA incorporates clinical knowledge after the process of modeling multiple diseases. While there are benefits of such an approach as described earlier, it may be possible to improve the predictive accuracy further by integrating clinical knowledge both during modeling and prediction, which future studies can explore.

7 CONCLUSION

This study proposes a novel KAA to enhance data-driven models for MDPA with knowledge from biomedical literature during risk prediction. Our empirical results demonstrate that KAA indeed increases the predictive performance of purely data-driven models for both common and rare diseases. Compared to extant approaches of combined knowledge-based data-driven modelling, KAA offers a more flexible approach, as it can be used in any underlying data-driven MDPA model. This study provides the impetus for further research to leverage the combination of knowledge graphs, which are automatically mined from rapidly growing biomedical literature, and machine learning models, trained on increasingly available EHR data, in creative ways to address complex healthcare challenges.

APPENDICES

A SEMMEDDB RELATIONSHIPS

Tables 6 and 7 lists the relationships found in SemMedDB [28].

Table 6. Summary of Relationships in SemMedDB

No.	Relationship Name: Explanation	Examples and Remarks
1	ADMINISTERED_TO: Given to an entity, when no assertion is made that the substance or procedure is being given as treatment.	Patients with single brain lesion received an extra 3 Gy x 5 radiotherapy .... *C0034618: Radiation therapy (Therapeutic or Preventive Procedure) - ADMINISTERED_TO - C0030705: Patients (Human).*
2	AFFECTS: Produces a direct effect on. Implied here is the altering or influencing of an existing condition, state, situation, or entity.	BAP31 and its caspase cleavage product regulate cell surface expression of tetraspanins and integrin-mediated cell survival. *C1424489: BCAP31 gene (Gene or Genome) - AFFECTS - C0007620: Cell Survival (Cell Function)*
3	ASSOCIATED_WITH: Has a relationship to (gene-disease relation).	EP2 plays a critical role in tumorigenesis in mouse mammary gland ... *C1419062: PTGER2 gene (Gene or Genome) - ASSO- CIATED_WITH - C1326912: Tumorigenesis (Neoplastic Process)*
4	AUGMENTS: Expands or stimulates a process.	Nicotine induces conditioned place preferences over a large range of doses in rats. *C0028040: Nicotine (Organic Chemical) - AUG- MENTS - C0815102: place preference learning (Mental Process)*
5	CAUSES: Brings about a condition or an effect. Implied here is that an agent, such as for example, a pharmacologic substance or an organism, has brought about the effect. This includes induces, effects, evokes, and etiology.	Neurocysticercosis (NCC) is one of the major causes of neurological disease ... *C0338437:Neurocysticercosis (Disease or Syndrome) - CAUSES - C0027765: nervous system disorder (Disease or Syndrome)*
6	COEXISTS_WITH: Occurs together with, or jointly.	Food intolerance-related constipation is characterized by proctitis. *C0009806: Constipation (Sign or Symptom) - COEX ISTS_WITH - C0033246: Proctitis (Disease or Syndrome)*
7	CONVERTS_TO: Changes from one form to another (both substances).	... plasma nitrite is readily oxidized to nitrate within plasma ... *C0028137: Nitrites (Chemical Viewed Structurally) - CONVERTS_TO - C0699857: Nitrate (Inorganic Chemical)*
8	COMPLICATES: Causes to become more severe or complex, or results in adverse effects.	Infections can trigger GBS and exacerbate CIDP. *C0021311: Infection (Disease or Syndrome) - COMPLI CATES - C0393819: Polyradiculoneuropathy, Chronic Inflammatory Demyelinating (Disease or Syndrome)*
9	DIAGNOSES: Distinguishes or identifies the nature or characteristics of.	Manometry showed a higher anal sphincter resting pressure ... *C0024751: Manometry (Laboratory Procedure) - DIAG NOSES - C0429217: Anal sphincter pressure (Finding)*
10	DISRUPTS: Alters or influences an already existing condition, state, or situation. Produces a negative effect on.	Overexpression of NF-kappaB inhibits tumor cell apoptosis. *C0079904: NF-kappaB (Amino Acid, Peptide, or Protein) - DISRUPTS - C0162638: Apoptosis (Cell Function)*
11	INHIBITS: Decreases, limits, or blocks the action or function of (substance interaction).	In recent studies, the BDNF expression was reduced by typical neuroleptics. *C0040615: Antipsychotic Agents (Pharmacologic Sub stance) - INHIBITS - C0107103: Brain-Derived Neurotrophic Factor (Biologically Active Substance)*
12	INTERACTS_WITH: Substance interaction.	Here, we show that chymases, which are chymotryptic peptidases secreted by mast cells, hydrolyze HGF ... *C0055673: Chymase (Enzyme) - INTERACTS_WITH- C0062534: Hepatocyte Growth Factor (Amino Acid, Peptide, or Protein)*
13	ISA: The basic hierarchical link in the UMLS Semantic Network. If one item is another item, then the first item is more specific in meaning than the second item.	The sympathetic neurotransmitter norepinephrine has been found ... *C0028351: Norepinephrine (Neuroreactive Substance or Biogenic Amine) - ISA - C0027908: Neurotransmitters (Neuroreactive Substance or Biogenic Amine)*
14	LOCATION_OF: The position, site, or region of an entity or the site of a process.	We report a case of primary cardiac epithelioid hemangioendothelioma arising from the right atrium of a 2-month-old infant. *C1269890: Entire right atrium (Body Part, Organ, or Organ Component) - LOCATION_OF - C0206732: Hemangioendothelioma, Epithelioid (Neoplastic Process)*
15	MANIFESTATION_OF: That part of a phenomenon that is directly observable or concretely or visibly expressed, or that gives evidence to the underlying process. This includes expression of, display of, and exhibition of.	Recurrence of glomerulopathy underlying ESRD was frequent for IgAN, FSG... *C1261469: End stage renal failure (Disease or Syn drome) - MANIFESTATION_OF - C1398810: glomerulopathy (Disease or Syndrome)*

Table 7. Summary of Relationships in SemMedDB

No.	Relationship Name: Explanation	Examples and Remarks
16	METHOD_OF: The manner and sequence of events in performing an act or procedure.	... because of the use of SSCP as a screening method and sequencing only a part of TSHR exon 10. *C0243031: Single-Stranded Conformational Poly morphism (Laboratory or Test Result) - METHOD_OF- C0220908:Screening procedure (Health Care Activity)*
17	OCCURS_IN: Has incidence in a group or population.	Older populations are more prone to bone loss with weight loss ... *C0599877: loss; bone (Pathologic Function) - OCCUR S_IN - C1518563: Older Population (Human)*
18	PART_OF: Composes, with one or more other physical units, some larger whole. This includes component of, division of, portion of, fragment of, section of, and layer of.	The probasal bodies and microtubules within the blepharoplast cavities... *C0026046: Microtubules (Cell Component) - PAR T_OF - C0230744: Basal body of cilium or flagellum, not bacterial (Cell Component)*
19	PRECEDES: Occurs earlier in time. This includes antedates, comes before, is in advance of, predates, and is prior to.	... the risk of tissue plasminogen activator-induced hemorrhagic transformation following ischemic stroke in mice ... *C0948008: Ischemic stroke (Disease or Syndrome) - PRECEDES - C1096400: Haemorrhagic transformation stroke (Disease or Syndrome)*
20	PREDISPOSES: To be a risk to a disorder, pathology, or condition.	... high ghrelin levels contribute to obesity in Prader-Willi syndrome (PWS) ... *C0911014: ghrelin (Amino Acid, Peptide, or Protein) - PREDISPOSES - C0028754: Obesity (Disease or Syndrome)*
21	PREVENTS: Stops, hinders or eliminates an action or condition.	Ipsapirone and ketanserin protects against circulatory shock, intracranial hypertension, and cerebral ischemia during heatstroke. *C0123905: ipsapirone (Pharmacologic Substance) - PRE VENTS - C0151740: Intracranial Hypertension (Disease or Syndrome)*
22	PROCESS_OF: Disorder occurs in (higher) organism.	... no information is available in CAD patients with normal glomerular filtration rate (GFR). *C0010054: Coronary Arteriosclerosis (Disease or Syndrome) - PROCESS_OF - C0030705: Patients (Human)*
23	PRODUCES: Brings forth, generates or creates. This includes yields, secretes, emits, biosynthesizes, generates, releases, discharges, and creates.	Human EPCs express functional PAR-1... *C0038250: Stem cells (Cell) - PRODUCES - C0668084: Receptor, PAR-1 (Amino Acid, Peptide, or Protein)*
24	STIMULATES: Increases or facilitates the action or function of (substance interaction).	Candesartan therapy significantly reduced inflammation and increased adiponectin levels ... *C0717550: candesartan (Pharmacologic Substance) - STIMULATES - C0389071: Adiponectin (Amino Acid, Peptide, or Protein)*
25	TREATS: Applies a remedy with the object of effecting a cure or managing a condition.	This study initially surveyed 163 patients with clinical stage Ib or IIa cervical adenocarcinoma treated with radical hysterectomy and pelvic lymphadenectomy. *C0677962: Radical hysterectomy (Therapeutic or Pre ventive Procedure) - TREATS - C0279672: Cervical Adenocarcinoma (Neoplastic Process)*
26	USES: Employs in the carrying out of some activity. This includes applies, utilizes, employs, and avails.	Pre-emptive therapy with oral valganciclovir for CMV infections... *C0087111: Therapeutic procedure (Therapeutic or Preventive Procedure) - USES - C0909381: valganciclovir (Pharmacologic Substance)*
27	COMPARED_WITH: Comparative predicate.	Compared with placebo, candesartan therapy significantly lowered plasma hsCRP levels... *C0032042: Placebos (Medical Device) - COMPARED_- WITH - C0717550: candesartan (Pharmacologic Substance)*
28	HIGHER_THAN: Comparative predicate.	Losartan was more effective than atenolol in reducing cardiovascular morbidity... *C0126174: Losartan (Organic Chemical) - HIGH ER_THAN - C0004147: Atenolol (Organic Chemical)*
29	LOWER_THAN: Comparative predicate.	Amoxicillin - Clavulanate was not as effective as ciprofloxacin for treating uncomplicated bladder infection in women. *C0054066: Amoxicillin-Potassium Clavulanate Combi nation (Antibiotics) - LOWER_THAN - C0008809: Ciprofloxacin (Pharmacologic Substance)*
30	SAME_AS: Comparative predicate.	Candesartan is as effective as lisinopril once daily in reducing blood pressure. *C0717550: candesartan (Organic Chemical) - SAME_AS - C0065374: Lisinopril (Amino Acid, Peptide, or Protein)*

B KNOWLEDGE GRAPH EMBEDDING USING TRANSE

The TransE model is a technical tool used in this study to obtain graph embeddings (i.e., these vector representations containing information on relationships between diseases and other clinical concepts (e.g., drugs)). So, diseases with similar clinical relationships in the knowledge graph will have similar vector representations. These vector representations are further used to obtain similarity between diseases in the augmentation procedure of our KAA approach. The graph in this study is the SemMedDB knowledge graph organized in the form of triples (CUI of disease1, relationship, CUI of disease2), as described above. The embedding obtained through TransE aims to encode global structural knowledge of the graph. We briefly describe the intuition behind TransE and refer the reader to Reference [4] for more details. Given a training set S of triplets (h, l, t) composed of two entities head disease h, tail disease t, and $h, t \in E$ (the set of clinical concepts), and a relationship $l \in L$ (the set of relationships; head disease has the relationship l to tail disease), TransE model outputs k-dimensional vector representations of the entities and the relationships (where k is a hyperparameter, set to 200 in our experiments). The basic idea behind TransE model is that the relationship induced by the l-labeled edges corresponds to a translation of the vector representations. That is, we want that in vector-space $h + l \approx t$ when (h, l, t) holds in the graphs (i.e., head disease h adding the relationship l should be close to tail disease t). If the triplet (h, l, t) is not present, then $h + l$ should be far away from t in vector space. Following an energy-based framework, the energy of a triplet is equal to $d(h + l, t)$, where d is a dissimilarity measurement function (e.g., $L1$ or the $L2$ norm). To learn the vector representations, a margin-based ranking criterion is minimized over the training set:

$$L = \sum_{(h,l,t)\in S} \sum_{(h',l',t')\in S'_{(h,l,t)}} [\gamma + d(h + l, t) - d(h' + l, t')]_+, \tag{6}$$

where $[x]_+$ denotes the positive part of $x, \gamma > 0$ denotes a positive margin hyperparameter, and

$$S'_{(h,l,t)} = (h', l, t)|h' \notin E \cup (h, l, t')|t' \notin E. \tag{7}$$

The set of corrupted triplets S' in Equation (7), consists of training triplets containing either the head diseases or tail diseases replaced by a random entity (but not both at the same time, denoted by h' or t'). The loss function (6) favors lower values of the energy for training triplets than for corrupted triplets and thus implement the intended criterion. For a given entity, its vector representation is the same as either the head or the tail of a triplet.

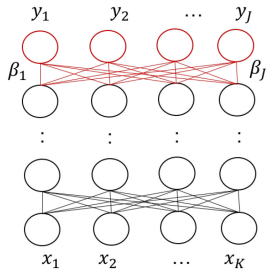


Fig. 6. Regularization of network weights.

C NEURAL NETWORK REGULARIZATION USING DISEASE SIMILARITIES (MDJ-REGU)

As shown in Figure 6, Reference [10] used a deep feed-forward neural network, with L hidden layers and an output prediction layer for multi-label classification, given K features of a set of samples $X = x_1, x_2, \ldots, x_N$ and the set of labels $Y = y_1, y_2, \ldots, y_J$, where $x_i \in R^K, y_i \in (0,1)^j, (1 \leq i \leq N)$. We use $\theta = (\theta_{hid}, B)$ to denote the model parameters and $\theta_{hid} = (W^{(l)}, b^{(l)})_{(l=1)}^{L}$ denotes the weights for the hidden layers (each with $D^{(l)}$ units) where $D^{(0)} = K$. The J columns of $B = [\beta_1, \beta_2, \ldots, \beta_J]$ from the last hidden layer are the prediction parameters (where $\beta_j \in R^{D^{(L)}}$). The neural network consists of fully connected layers, linear activation ($W^{(l)}h^{(l-1)} + b^{(l)}$) (where $h^{(0)} = x_i$) and sigmoid nonlinearities ($\delta(z) = 1/(1 + exp(-z))$. The conditional likelihood of y_i given x_i and model parameters θ can be written as:

$$\log P(y_i|x_i, \theta) = \sum_{j=1}^{J} [\epsilon_j \log \delta_j(\beta_j^{\intercal} h_j) + (1 - \epsilon_j) \log(1 - \delta_j)(\beta_j^{\intercal} h_j)], \tag{8}$$

$$tr(B^{\intercal} B) = \frac{1}{2} \sum_{1 \leq j, j' \leq J} A_{j,j'} |\beta_j - \beta_j^{'}|_2^2, \tag{9}$$

where $tr(.)$ represents the trace operator, the graph Laplacian regularizer enforces the parameters β_j and $\beta_{j'}$ to be similar, proportional to $A_{j,j'}$. The regularized loss function is:

$$L = -\sum_{i=1}^{N} \log p(y_i|x_i, \theta) + \frac{\rho}{2} tr(B^{\intercal} B), \tag{10}$$

where $rho > 0$ are the Laplacian hyperparameters. Thus, the similarity value in $A_{j,j'}$ affects the parameters β_j and $\beta_{j'}$, brings them closer, which in turn makes the predictions for diseases j, $j^{'}$ similar.

D TRAINING MDS AND MDJ MODELS

During training, a classifier has to be chosen in both MDS and MDJ frameworks. Any statistical or machine learning method to obtain J binary classifiers g_j (for MDS) and co-trained structure $C(x)$ and δ_j, $(1 \leq j \leq J)$ (for MDJ) can be adopted. We use a **deep neural network (DNN)** model with two hidden layers. Hyperparameter tuning can be done on a randomly chosen validation set (that is separated out from the training data). In our experiments, the learning rate, batch size, and two hidden layer sizes for binary classifier g_j are set to 0.008, 50, 100, 100, respectively. The same set of hyperparameters for the co-trained neural network C are set to 0.008, 75, 2000, 100, and a sigmoid activation function is used for δ_j. Number of epochs was set to 100 for MDS and 180 for MDJ. If a binary label is required from the prediction probabilities ($P_j^{'}$), then a threshold probability value has to be chosen: The label is set to 1 above this threshold and to 0 below the threshold. We set the threshold value to a constant a across all labels. This value is chosen empirically to maximize the overall F-measure on the training data [45].

E KAA: ALGORITHM DESCRIPTION

Table 8 below lists all the steps for using KAA with an MDPA model and a Knowledge Graph.

Hyperparameter ϵ for KAA and γ for the RBF kernel may be selected by grid search by evaluating the performance on a randomly chosen validation set (that is separated out from the training data). In our experiments, we use $\epsilon = 0.9, \gamma = 0.01$. λ can be set to a low value, e.g., 0.0001; in

Table 8. Algorithm for KAA

INPUT: Disease Probabilities P_j for J diseases from any MDPA model, Knowledge Graph KG

1: Obtain disease embeddings d_j for all $1 \leq j \leq J$ by using TransE on KG

2: Compute pairwise similarities of diseases, $S_{j,k} = \exp(-\gamma \|d_j - d_k\|)^2$ where $1 \leq j, k \leq J$

3: Initialize $P_j^{'(t=0)} = P_j$ for $1 \leq j \leq J$

4: Do:

5: For each j (disease):

6: Update $P_j^{'(t)} = (1-\epsilon)\frac{\sum_{k=1}^{J} S_{j,k}, P_k^{'(t-1)}}{\|S_{j,*}\|_1} + \epsilon P_j$ (using Equation (5))

7: $t = t + 1$

7: Until $f(P_j^{'(t+1)}) - f(P_j^{'(t)}) < \lambda$

OUTPUT: Knowledge-refined probabilities $P_j^{'}$ for J input diseases

Table 9. Summary of the 30 Selected Diseases

No.	Disease Name	Abbreviation	Disease Examples (ICD-9-CM codes)
1	Septicemia (except in labor)	SEP	Septicemia (038)
2	Acute and unspecified renal failure	ARF	Kidney Failure, Acute (5849)
3	Congestive heart failure; nonhypertensive	CHF	Congestive Heart Failure (4280)
4	Pneumonia (except that caused by tuberculosis or sexually transmitted disease)	PNE	Pneumonia due to other specified bacteria (4828)
5	Fluid and electrolyte disorders	FED	Electrolyte and fluid disorders not elsewhere classified (2769)
6	Other liver diseases	LD	Non-alcoholic Fatty Liver (5718)
7	Hypertension with complications and secondary hypertension	HCSH	Cardiovascular Diseases (3062); Coronary Heart Diseases (402)
8	Deficiency and other anemia	DA	folate-deficiency anemia (2812)
9	Other nutritional; endocrine; and metabolic disorders	MD	Obesity (2780)
10	Essential hypertension	EH	Essential Hypertension (401)
11	Coronary atherosclerosis and other heart disease	CAH	Coronary Arteriosclerosis (4140)
12	Diabetes mellitus without complication	DWC	Glucose Intolerance (2713)
13	Disorders of lipid metabolism	DLM	Hypercholesterolemia (2720, 2722)
14	Diabetes mellitus with complications	DMC	Diabetic Retinopathy (36201-36207)
15	Cardiac dysrhythmias	CD	Cardiac dysrhythmias (427)
16	Other nervous system disorders	NSD	Unspecified disorders of nervous system (3499)
17	Heart valve disorders	HVD	Aortic Valve Stenosis (3960)
18	Esophageal disorders	ED	Esophageal hemorrhage (53082)
19	Hodgkin‘s disease	HD	Hodgkin’s disease (201)
20	Respiratory failure; insufficiency; arrest (adult)	RF	Acute respiratory failure (51881)
21	Multiple myeloma	MM	Multiple myeloma (2030)
22	Chronic kidney disease	CKD	Chronic kidney disease (585)

(Continued)

Table 9. Continued

No.	Disease Name	Abbreviation	Disease Examples (ICD-9-CM codes)
23	Systemic lupus erythematosus and connective tissue disorders	SLE	Systemic lupus erythematosus (7100)
24	Other non-epithelial cancer of skin	NECS	Skin of trunk, except scrotum (1735)
25	Cancer of liver and intrahepatic bile duct	CIBD	Intrahepatic bile ducts (1551)
26	Melanomas of skin	MOS	Malignant melanoma of skin (172)
27	Multiple sclerosis	MS	Multiple sclerosis (340)
28	Other CNS infection and poliomyelitis	CNS	Acute poliomyelitis (045)
29	Lymphadenitis	LYM	Acute lymphadenitis (683)
30	Cancer of rectum and anus	CRA	Malignant neoplasm of rectum, rectosigmoid junction, and anus (154)

our experiments, the difference in loss between successive iterations reaches zero within 10–15 iterations.

In our experiments, detailed in Section 4, we have considered 30 diseases (shown in Table 9) from the MIMIC-III clinical database. The Knowledge Graph used is SemMedDB [28]. Table 11 in Appendix G lists the risk probabilities obtained for these 30 diseases at time window 1 from the neural network models MDS and MDJ in columns (1) and (3), respectively. The KAA procedure is applied on these probabilities to obtain knowledge-refined probabilities from MDS-KAA and MDJ-KAA in columns (2) and (4), respectively. Similarly, KAA is applied on other Multi-label Learning methods and the overall summary of results for all time windows are shown in Tables 4 and 5.

F DETAILS OF 30 SELECTED DISEASES

We simultaneously model and predict patients' risks of having 30 diseases (listed in Table 9). These diseases are identified based on the diagnostic definitions from the **Health Cost and Utilization Clinical Classification Software (HCUP CCS)** [23]. HCUP CCS definitions can cluster ICD-9 diagnostic codes found in EHR data into mutually exclusive, broadly homogeneous disease categories to reduce the noise, ambiguity, and redundancy in the original ICD-9 diagnostic codes. As a result, there are a total of 268 diseases identified from the MIMIC-III dataset.

Table 10. The Summary of the Rarity of Rare Diseases

Rare diseases name	Frequency	Imbalance ratio
Hodgkin's disease	574,000	362.4
Multiple myeloma	488,200	301.1
Systemic lupus erythematosus and connective tissue disorders	20–70 per 100,000	148.6
Other non-epithelial cancer of skin	392,000	89.7
Cancer of liver and intrahepatic bile duct	1–2 per 100,000	109.5
Melanomas of skin	3,100,000	113.6
Multiple sclerosis	2,000,000	208.5
Other CNS infection and poliomyelitis	-	171.9
Lymphadenitis	-	216.4
Cancer of rectum and anus	3,000,000	250.5

Rare Diseases

To validate the the rarity of the selected rare diseases, we present their frequency (i.e., the proportion of world population affected by these diseases in 2015) [6, 15, 41] and their imbalance ratios calculated from our dataset, as shown in Table 10. Any disease, disorder, illness, or condition affecting fewer than 200,000 people in the United States is regarded rare by the **National Institutes of Health (NIH)** and the U.S. **Food and Drug Administration (FDA)** [9]. To generalize this definition to the rest of the world, we compute the prevalence proportion as 6.23×10^{-4} using the estimated population of the U.S. as 321M [7]. Using this proportion and the estimated world population of 7.358B [37], we get the definition of rare disease to be less than either 4,580,000, or 63 per 100,000 in the world. As shown in Table 3, we find that the diseases selected from our dataset are rare by this definition. The frequency of two diseases—"Other CNS infection and poliomyelitis" and "Lymphadenitis"—could not be found, but both of them should be rare, too, according to their high imbalance ratios.

G RESULTS OF INDIVIDUAL DISEASES

Table 11 shows the results for individual diseases.

Table 11. Results of Individual Disease at w1

NO.	Disease name	MDS (1)	MDS-KAA (2)	MDJ (3)	MDJ-KAA (4)
1	SEP	0.5144	0.5494	0.5357	0.5498
2	ARF	0.5503	0.5804	0.5766	0.5920
3	CHF	0.5815	0.6158	0.6398	0.6472
4	PNE	0.5053	0.5309	0.5179	0.5464
5	FED	0.5161	0.5161	0.5704	0.5704
6	LD	0.5757	0.5861	0.6247	0.6365
7	HCSH	0.5782	0.5782	0.6654	0.6654
8	DA	0.5236	0.5427	0.5452	0.5603
9	MD	0.5350	0.5358	0.5613	0.5887
10	EH	0.5491	0.5491	0.5773	0.5783
11	CAH	0.5817	0.5770	0.6633	0.6681
12	DWC	0.5471	0.5545	0.6476	0.6509
13	DLM	0.5772	0.5772	0.6039	0.6039
14	DMC	0.5769	0.5771	0.6825	0.6758
15	CD	0.5729	0.5953	0.6078	0.6204
16	NSD	0.5396	0.5291	0.5899	0.5922
17	HVD	0.5661	0.5867	0.6187	0.6264
18	ED	0.5610	0.5610	0.5920	0.5920
19	HD	0.4876	0.4876	0.7684	0.7684
20	RF	0.5293	0.5546	0.5696	0.5875
21	MM	0.4896	0.7449	0.5570	0.7066
22	CKD	0.6265	0.6570	0.6982	0.7064
23	SLE	0.5271	0.6267	0.6717	0.6897
24	NECS	0.5003	0.5003	0.5701	0.5701
25	CIBD	0.5256	0.5256	0.6734	0.6831
26	MOS	0.5127	0.5127	0.6648	0.6648

(Continued)

Table 11. Continued

NO.	Disease name	MDS (1)	MDS-KAA (2)	MDJ (3)	MDJ-KAA (4)
27	MS	0.5209	0.5209	0.8261	0.8419
28	CNS	0.4882	0.4882	0.6263	0.6263
29	LYM	0.4882	0.4882	0.5146	0.5146
30	CRA	0.4909	0.4909	0.5713	0.5713

H COUNTERFACTUAL ANALYSIS

Table 12 lists the positive associations in SemMedDB considered as triggers for diseases.

Table 12. Summary of Positive Relationships between Diseases

Relationship	Explanation
AFFECTS	Produces a direct effect on
AUGMENTS	Expands or stimulates a process
CAUSES	Brings about a condition or an effect
COEXISTS_WITH	Occurs together with, or jointly
CONVERTS_TO	Changes from one form to another
COMPLICATES	Causes to become more severe or complex, or results in adverse effects
ISA	The first item is more specific in meaning than the second item
MANIFESTATION_OF	That part of a phenomenon that is directly observable
OCCURS_IN	Has incidence in a group
PRECEDES	Occurs earlier in time
PREDISPOSES	To be a risk to a disorder, pathology, or condition
PRODUCES	Brings forth, generates or creates
STIMULATES	Increases or facilitates the action or function of
SAME_AS	Comparative predicate

Figures 7 and 8 show the results at 8% and 6% cut-off values to distinguish between high-risk and low-risk based on model predictions.

Overall diseases (# of events = 38333)

Prevention prescribed at t_{io}

	Predicted Risk (MDS)	
	Low	High
Yes	14140	13031
No	7158	4004

Rare diseases (# of events = 647)

Prevention prescribed at t_{io}

	Predicted Risk (MDS)	
	Low	High
Yes	241	146
No	187	73

Overall diseases (# of events = 38333)

Prevention prescribed at t_{io}

	Predicted Risk (MDS-KAA)	
	Low	High
Yes	8645	18526
No	5392	5770

Rare diseases (# of events = 647)

Prevention prescribed at t_{io}

	Predicted Risk (MDS-KAA)	
	Low	High
Yes	190	197
No	172	88

Overall diseases (# of events = 38333)

Prevention prescribed at t_{io}

	Predicted Risk (MDJ)	
	Low	High
Yes	13506	13665
No	7225	3937

Rare diseases (# of events = 647)

Prevention prescribed at t_{io}

	Predicted Risk (MDJ)	
	Low	High
Yes	207	180
No	190	70

Overall diseases (# of events = 38333)

Prevention prescribed at t_{io}

	Predicted Risk (MDJ-KAA)	
	Low	High
Yes	8584	18587
No	5478	5684

Rare diseases (# of events = 647)

Prevention prescribed at t_{io}

	Predicted Risk (MDJ-KAA)	
	Low	High
Yes	173	214
No	180	80

Fig. 7. Results of counterfactual analysis (8% cut-off).

Overall diseases (# of events = 38333)

Prevention prescribed at t_{io}

	Predicted Risk (MDS)	
	Low	High
Yes	14138	13033
No	7158	4004

Rare diseases (# of events = 647)

Prevention prescribed at t_{io}

	Predicted Risk (MDS)	
	Low	High
Yes	241	146
No	187	73

Overall diseases (# of events = 38333)

Prevention prescribed at t_{io}

	Predicted Risk (MDS-KAA)	
	Low	High
Yes	7858	19313
No	5027	6135

Rare diseases (# of events = 647)

Prevention prescribed at t_{io}

	Predicted Risk (MDS-KAA)	
	Low	High
Yes	179	208
No	167	93

Overall diseases (# of events = 38333)

Prevention prescribed at t_{io}

	Predicted Risk (MDJ)	
	Low	High
Yes	13511	13660
No	7225	3937

Rare diseases (# of events = 647)

Prevention prescribed at t_{io}

	Predicted Risk (MDJ)	
	Low	High
Yes	207	180
No	190	70

Overall diseases (# of events = 38333)

Prevention prescribed at t_{io}

	Predicted Risk (MDJ-KAA)	
	Low	High
Yes	9175	17996
No	5725	5437

Rare diseases (# of events = 647)

Prevention prescribed at t_{io}

	Predicted Risk (MDJ-KAA)	
	Low	High
Yes	164	223
No	178	82

Fig. 8. Results of counterfactual analysis (6% cut-off).

The results for individual diseases with MDS model are shown in Tables 13, 14, and 15 for cut-off values 10%, 8%, and 6%, respectively.

Table 13. Results of Counterfactual Analyses for MDS (10% Cut-off)

Disease name	MDS-not-KAA				MDS-KAA			
	a	b	c	d	a	b	c	d
SEP	932	726	105	62	492	1,166	69	98
ARF	1,217	980	85	45	719	1,478	61	69
CHF	1,268	1,285	67	37	768	1,785	41	63
PNE	767	567	177	88	416	918	116	149
FED	516	588	1,072	741	516	588	1,072	741
LD	468	369	100	45	255	582	57	88
HCSH	388	335	524	232	388	335	524	232
DA	1,101	962	178	113	630	1,433	125	166
MD	208	190	348	182	116	282	210	320
EH	1,229	1,298	217	172	994	1,533	169	220
CAH	840	889	416	239	523	1,206	253	402
DWC	865	671	67	49	485	1,051	39	77
DLM	588	619	574	329	588	619	574	329
DMC	415	305	157	62	415	305	157	62
CD	994	1,137	354	245	568	1,563	232	367
NSD	574	421	349	127	323	672	220	256
HVD	466	375	306	136	257	584	171	271
ED	310	311	505	264	310	311	505	264
HD	15	9	4	2	15	9	4	2
RF	387	457	793	547	208	636	473	867
MM	16	8	12	3	11	13	9	6
CKD	366	400	577	216	210	556	388	405
SLE	43	27	8	6	16	54	3	11
NECS	19	12	55	23	19	12	55	23
CIBD	42	30	27	12	34	38	21	18
MOS	39	21	18	7	39	21	18	7
MS	32	12	6	2	30	14	6	2
CNS	21	13	9	7	21	13	9	7
LYM	2	3	34	8	2	3	34	8
CRA	12	11	14	3	12	11	14	3
Overall Diseases	14,140	13,031	7,158	4,004	9,380	17,791	5,629	5,533
Rare Diseases	241	146	187	73	199	188	173	87

Table 14. Results of Counterfactual Analyses for MDS (8% Cut-off)

Disease	MDS-not-KAA				MDS-KAA			
name	a	b	c	d	a	b	c	d
SEP	932	726	105	62	424	1,234	63	104
ARF	1,217	980	85	45	622	1,575	57	73
CHF	1,268	1,285	67	37	675	1,878	39	65
PNE	767	567	177	88	355	979	104	161
FED	516	588	1,072	741	516	588	1,072	741
LD	468	369	100	45	229	608	53	92
HCSH	388	335	524	232	388	335	524	232
DA	1,101	962	178	113	544	1,519	104	187
MD	208	190	348	182	96	302	177	353
EH	1,229	1,298	217	172	944	1,583	164	225
CAH	840	889	416	239	500	1,229	229	426
DWC	865	671	67	49	424	1,112	36	80
DLM	588	619	574	329	588	619	574	329
DMC	415	305	157	62	415	305	157	62
CD	994	1,137	354	245	538	1,593	224	375
NSD	574	421	349	127	277	718	191	285
HVD	466	375	306	136	236	605	163	279
ED	310	311	505	264	310	311	505	264
HD	15	9	4	2	15	9	4	2
RF	387	457	793	547	185	659	427	913
MM	16	8	12	3	8	16	9	6
CKD	366	400	577	216	189	577	357	436
SLE	43	27	8	6	15	55	3	11
NECS	19	12	55	23	19	12	55	23
CIBD	42	30	27	12	32	40	20	19
MOS	39	21	18	7	39	21	18	7
MS	32	12	6	2	27	17	6	2
CNS	21	13	9	7	21	13	9	7
LYM	2	3	34	8	2	3	34	8
CRA	12	11	14	3	12	11	14	3
Overall Diseases	14,140	13,031	7,158	4,004	8,645	18,526	5,392	5,770
Rare Diseases	241	146	187	73	190	197	172	88

Table 15. Results of Counterfactual Analyses for MDS (6% Cut-off)

Disease	MDS-not-KAA				MDS-KAA			
name	a	b	c	d	a	b	c	d
SEP	931	727	105	62	382	1,276	54	113
ARF	1,217	980	85	45	573	1,624	53	77
CHF	1,267	1,286	67	37	625	1,928	36	68
PNE	767	567	177	88	315	1,019	96	169
FED	516	588	1,072	741	516	588	1,072	741
LD	468	369	100	45	171	666	46	99
HCSH	388	335	524	232	388	335	524	232
DA	1,101	962	178	113	491	1,572	96	195
MD	208	190	348	182	82	316	141	389
EH	1,229	1,298	217	172	843	1,684	149	240
CAH	840	889	416	239	459	1,270	208	447
DWC	865	671	67	49	378	1,158	33	83
DLM	588	619	574	329	588	619	574	329
DMC	415	305	157	62	415	305	157	62
CD	994	1,137	354	245	432	1,699	176	423
NSD	574	421	349	127	209	786	154	322
HVD	466	375	306	136	199	642	137	305
ED	310	311	505	264	310	311	505	264
HD	15	9	4	2	15	9	4	2
RF	387	457	793	547	139	705	346	994
MM	16	8	12	3	4	20	8	7
CKD	366	400	577	216	164	602	303	490
SLE	43	27	8	6	14	56	2	12
NECS	19	12	55	23	19	12	55	23
CIBD	42	30	27	12	31	41	19	20
MOS	39	21	18	7	39	21	18	7
MS	32	12	6	2	22	22	4	4
CNS	21	13	9	7	21	13	9	7
LYM	2	3	34	8	2	3	34	8
CRA	12	11	14	3	12	11	14	3
Overall Diseases	14,138	13,033	7,158	4,004	7,858	19,313	5,027	6,135
Rare Diseases	241	146	187	73	179	208	167	93

The results for individual diseases with MDJ model are shown in Tables 16, 17, and 18 for cut-off values 10%, 8%, and 6%, respectively.

Table 16. Results of Counterfactual Analyses for MDJ (10% Cut-off)

Disease	MDJ-not-KAA				MDJ-KAA			
name	a	b	c	d	a	b	c	d
SEP	936	722	105	62	521	1,137	75	92
ARF	1,232	965	95	35	720	1,477	63	67
CHF	1,239	1,314	65	39	807	1,746	43	61
PNE	789	545	176	89	435	899	116	149
FED	504	600	1,057	756	504	600	1,057	756
LD	408	429	108	37	233	604	72	73
HCSH	305	418	512	244	305	418	512	244
DA	1,070	993	182	109	638	1,425	119	172
MD	196	202	360	170	112	286	228	302
EH	1,241	1,286	225	164	992	1,535	182	207
CAH	803	926	440	215	549	1,180	258	397
DWC	788	748	73	43	448	1,088	45	71
DLM	533	674	586	317	533	674	586	317
DMC	382	338	159	60	382	338	159	60
CD	954	1,177	385	214	599	1,532	254	345
NSD	558	437	341	135	325	670	233	243
HVD	406	435	303	139	235	606	172	270
ED	266	355	488	281	266	355	488	281
HD	15	9	4	2	15	9	4	2
RF	368	476	811	529	205	639	501	839
MM	15	9	11	4	8	16	8	7
CKD	326	440	564	229	190	576	380	413
SLE	36	34	8	6	19	51	6	8
NECS	18	13	58	20	18	13	58	20
CIBD	28	44	29	10	23	49	26	13
MOS	34	26	18	7	34	26	18	7
MS	29	15	6	2	27	17	6	2
CNS	19	15	8	8	19	15	8	8
LYM	2	3	34	8	2	3	34	8
CRA	11	12	14	3	11	12	14	3
Overall Diseases	13,511	13,660	7,225	3,937	9,175	17,996	5,725	5,437
Rare Diseases	207	180	190	70	176	211	182	78

Table 17. Results of Counterfactual Analyses for MDJ (8% Cut-off)

Disease	MDJ-not-KAA				MDJ-KAA			
name	a	b	c	d	a	b	c	d
SEP	936	722	105	62	468	1,190	63	104
ARF	1,231	966	95	35	644	1,553	58	72
CHF	1,239	1,314	65	39	735	1,818	38	66
PNE	789	545	176	89	384	950	105	160
FED	504	600	1,057	756	504	600	1,057	756
LD	407	430	108	37	220	617	68	77
HCSH	305	418	512	244	305	418	512	244
DA	1,070	993	182	109	558	1,505	105	186
MD	196	202	360	170	94	304	197	333
EH	1,240	1,287	225	164	951	1,576	177	212
CAH	803	926	440	215	522	1,207	245	410
DWC	788	748	73	43	419	1,117	40	76
DLM	533	674	586	317	533	674	586	317
DMC	382	338	159	60	382	338	159	60
CD	952	1,179	385	214	556	1,575	237	362
NSD	558	437	341	135	288	707	210	266
HVD	406	435	303	139	223	618	159	283
ED	266	355	488	281	266	355	488	281
HD	15	9	4	2	15	9	4	2
RF	368	476	811	529	184	660	448	892
MM	15	9	11	4	8	16	8	7
CKD	326	440	564	229	175	591	346	447
SLE	36	34	8	6	17	53	5	9
NECS	18	13	58	20	18	13	58	20
CIBD	28	44	29	10	22	50	25	14
MOS	34	26	18	7	34	26	18	7
MS	29	15	6	2	27	17	6	2
CNS	19	15	8	8	19	15	8	8
LYM	2	3	34	8	2	3	34	8
CRA	11	12	14	3	11	12	14	3
Overall Diseases	13,506	13,665	7,225	3,937	8,584	18,587	5,478	5,684
Rare Diseases	207	180	190	70	173	214	180	80

Table 18. Results of Counterfactual Analyses for MDJ (6% Cut-off)

Disease	MDJ-not-KAA				MDJ-KAA			
name	a	b	c	d	a	b	c	d
SEP	935	723	105	62	425	1,233	57	110
ARF	1,228	969	95	35	605	1,592	56	74
CHF	1,238	1,315	65	39	691	1,862	36	68
PNE	788	546	176	89	355	979	101	164
FED	504	600	1,056	757	504	600	1,056	757
LD	407	430	108	37	168	669	53	92
HCSH	305	418	512	244	305	418	512	244
DA	1,070	993	182	109	512	1,551	100	191
MD	196	202	359	171	80	318	167	363
EH	1,239	1,288	225	164	866	1,661	162	227
CAH	802	927	440	215	481	1,248	232	423
DWC	787	749	73	43	371	1,165	34	82
DLM	531	676	586	317	531	676	586	317
DMC	382	338	159	60	382	338	159	60
CD	951	1,180	385	214	475	1,656	187	412
NSD	558	437	341	135	225	770	174	302
HVD	406	435	303	139	204	637	133	309
ED	266	355	488	281	266	355	488	281
HD	15	9	4	2	15	9	4	2
RF	368	476	811	529	143	701	360	980
MM	15	9	11	4	6	18	7	8
CKD	326	440	564	229	157	609	309	484
SLE	36	34	8	6	15	55	5	9
NECS	18	13	58	20	18	13	58	20
CIBD	28	44	29	10	20	52	25	14
MOS	34	26	18	7	34	26	18	7
MS	29	15	6	2	24	20	5	3
CNS	19	15	8	8	19	15	8	8
LYM	2	3	34	8	2	3	34	8
CRA	11	12	14	3	11	12	14	3
Overall Diseases	13,494	13,677	7,223	3,939	7,910	19,261	5,140	6,022
Rare Diseases	207	180	190	70	164	223	178	82

REFERENCES

[1] Alan R. Aronson. 2001. Effective mapping of biomedical text to the UMLS metathesaurus: The metamap program. In *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 17.

[2] Eta S. Berner. 2007. *Clinical Decision Support Systems*. Vol. 233. Springer.

[3] Sakyajit Bhattacharya, Vijay Huddar, Vaibhav Rajan, and Chandan K. Reddy. 2018. A dual boundary classifier for predicting acute hypotensive episodes in critical care. *PloS One* 13, 2 (2018).

[4] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Proceedings of the International Conference on Advances in Neural Information Processing Systems*. 2787–2795.

[5] Matthew R. Boutell, Jiebo Luo, Xipeng Shen, and Christopher M. Brown. 2004. Learning multi-label scene classification. *Pattern Recog.* 37, 9 (2004), 1757–1771.

[6] John A. Bridgewater, Karyn A. Goodman, Aparna Kalyan, and Mary F. Mulcahy. 2016. Biliary tract cancer: Epidemiology, radiotherapy, and molecular profiling. *Amer. Soc. Clin. Oncol. Educ. Book* 36 (2016), e194–e203.

[7] U.S. Census Bureau. 2015. Annual Estimates of the Resident Population for Selected Age Groups by Sex for the United States, States, Counties and Puerto Rico Commonwealth and Municipios: April 1, 2010 to July 1, 2015. Retrieved from https://www.census.gov/data/tables/time-series/demo/popest/2010s-national-detail.html.

[8] CDC. 2014. National Diabetes Statistics Report, Center for Disease Control and Prevention. Retrieved from http://www.cdc.gov/diabetes/pdfs/data/2014-report-estimates-of-diabetes-and-its-burden-in-the-united-states.pdf.

[9] CDN. 2014. Rare Disease Facts, National Organization for Rare Disorders. Retrieved from https://www.nidcd.nih.gov/directory/national-organization-rare-disorders-nord#:~:text=Description%3A,affects%20fewer%20than%20200%2C000%20Americans.

[10] Zhengping Che, David Kale, Wenzhe Li, Mohammad Taha Bahadori, and Yan Liu. 2015. Deep computational phenotyping. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 507–516.

[11] Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F. Stewart, and Jimeng Sun. 2016. Doctor AI: Predicting clinical events via recurrent neural networks. In *Proceedings of the Machine Learning for Healthcare Conference*. 301–318.

[12] Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, and Jimeng Sun. 2017. GRAM: Graph-based attention model for healthcare representation learning. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 787–795.

[13] Edward Choi, Mohammad Taha Bahadori, Jimeng Sun, Joshua Kulas, Andy Schuetz, and Walter Stewart. 2016. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. In *Proceedings of the International Conference on Advances in Neural Information Processing Systems*. 3504–3512.

[14] Edward Choi, Cao Xiao, Walter Stewart, and Jimeng Sun. 2018. MiME: Multilevel medical embedding of electronic health records for predictive healthcare. In *Proceedings of the International Conference on Advances in Neural Information Processing Systems*. 4547–4557.

[15] N. Danchenko, J. A. Satia, and M. S. Anthony. 2006. Epidemiology of systemic lupus erythematosus: A comparison of worldwide disease burden. *Lupus* 15, 5 (2006), 308–318.

[16] Mandip S. Dhamoon and Mitchell S. V. Elkind. 2010. Inclusion of stroke as an outcome and risk equivalent in risk scores for primary and secondary prevention of vascular disease. *Circulation* 121, 18 (2010), 2071–2078.

[17] Yuan Fang, Kingsley Kuan, Jie Lin, Cheston Tan, and Vijay Chandrasekhar. 2017. Object detection meets knowledge graphs. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17)*.

[18] Jessie Gerteis, David Izrael, Deborah Deitz, Lisa LeRoy, Richard Ricciardi, Therese Miller, and Jayasree Basu. 2014. *Multiple Chronic Conditions Chartbook*. Agency for Healthcare Research and Quality, Rockville, MD..

[19] Deborah Grady and Seth A. Berkowitz. 2011. Why is a good clinical prediction rule so hard to find? *Arch. Internal Medic.* 171, 19 (2011), 1701–1702.

[20] Rebecca S. Graves. 2002. Users' guides to the medical literature: A manual for evidence-based clinical practice. *J. Medic. Libr. Assoc.* 90, 4 (2002), 483.

[21] Shirley Gregor and Alan R. Hevner. 2013. Positioning and presenting design science research for maximum impact. *MIS Quart.* 37, 2 (2013), 337–355.

[22] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.

[23] HCUP. 2020. Clinical Classications Software (CCS) for ICD-9-CM Fact Sheet, Agency for Healthcare Research and Quality. Retrieved from https://www.hcup-us.ahrq.gov/ toolssoftware/ccs/ccsfactsheet.jsp.

[24] Vijay Huddar, Bapu Koundinya Desiraju, Vaibhav Rajan, Sakyajit Bhattacharya, Shourya Roy, and Chandan K. Reddy. 2016. Predicting complications in critical care using heterogeneous clinical data. *IEEE Access* 4 (2016), 7988–8001.

[25] IDF. 2015. What Is Diabetes? Retrieved from https://www.idf.org/about-diabetes/what-is-diabetes.html.

[26] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. 2020. A survey on knowledge graphs: Representation, acquisition and applications. *arXiv preprint arXiv:2002.00388* (2020).
[27] Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, H. Lehman Li-wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. *Sci. Data* 3 (2016), 160035.
[28] Halil Kilicoglu, Dongwook Shin, Marcelo Fiszman, Graciela Rosemblat, and Thomas C. Rindflesch. 2012. SemMedDB: A PubMed-scale repository of biomedical semantic predications. *Bioinformatics* 28, 23 (2012), 3158–3160.
[29] M. S. V. Elkind, R. D'Agostino, M. S. Dhamoon, D. T. Lackland, D. C. Goff, R. T. Higashida, L. A. McClure, P. H. Mitchell, R. L. Sacco, C. A. Sila, S. C. Smith, D. Tanne, D. L. Tirschwell, E. Touzé, and L. R. Wechsler. 2012. Inclusion of stroke in cardiovascular risk prediction instruments a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 43, 7 (2012), 1998–2027.
[30] Yu-Kai Lin, Hsinchun Chen, Randall A. Brown, Shu-Hsing Li, and Hung-Jen Yang. 2017. Healthcare predictive analytics for risk profiling in chronic care: A Bayesian multitask learning approach. *MIS Quart.* 41, 2 (2017).
[31] Zachary C. Lipton, David C. Kale, Charles Elkan, and Randall Wetzel. 2016. Learning to diagnose with LSTM recurrent neural networks. In *Proceedings of the International Conference on Learning Representations*.
[32] Xu-Ying Liu, Jianxin Wu, and Zhi-Hua Zhou. 2008. Exploratory undersampling for class-imbalance learning. *IEEE Trans. Syst. Man Cyber. Part B (Cyber.)* 39, 2 (2008), 539–550.
[33] Fenglong Ma, Radha Chitta, Jing Zhou, Quanzeng You, Tong Sun, and Jing Gao. 2017. Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1903–1911.
[34] Fenglong Ma, Quanzeng You, Houping Xiao, Radha Chitta, Jing Zhou, and Jing Gao. 2018. KAME: Knowledge-based attention model for diagnosis prediction in healthcare. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 743–752.
[35] Justin Mower, Devika Subramanian, and Trevor Cohen. 2018. Learning predictive models of drug side-effect relationships from distributed representations of literature-derived semantic predications. *J. Amer. Medic. Inform. Assoc.* 25, 10, 1339–1350.
[36] Mark A. Musen, Blackford Middleton, and Robert A. Greenes. 2014. Clinical decision-support systems. In *Biomedical Informatics*. Springer, 643–674.
[37] United Nations. 2015. The World Population Prospects: 2015 Revision. Retrieved from https://www.un.org/en/development/desa/publications/world-population-prospects-2015-revision.html.
[38] NLM. 2019. MEDLINE, U.S. National Library of Medicine. Retrieved from https://www.nlm.nih.gov/bsd/medline.html.
[39] Narges Razavian, Jake Marcus, and David Sontag. 2016. Multi-task prediction of disease onsets from longitudinal laboratory tests. In *Proceedings of the Machine Learning for Healthcare Conference*. 73–100.
[40] Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. 2011. Classifier chains for multi-label classification. *Mach. Learn.* 85, 3 (2011), 333.
[41] K. Srinath Reddy. 2016. Global burden of disease study 2015 provides GPS for global health 2030. *Lancet* 388, 10053 (2016), 1448–1449.
[42] Junyuan Shang, Shenda Hong, Yuxi Zhou, Meng Wu, and Hongyan Li. 2018. Knowledge guided multi-instance multi-label learning via neural networks in medicines prediction. In *Proceedings of the Asian Conference on Machine Learning*. 831–846.
[43] Benjamin Shickel, Patrick James Tighe, Azra Bihorac, and Parisa Rashidi. 2017. Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J. Biomed. Health Inform.* 22, 5 (2017), 1589–1604.
[44] Vergil N. Slee. 1978. The International Classification of Diseases: Ninth Revision (ICD-9). Retrieved from https://www.cdc.gov/nchs/icd/icd9cm.htm.
[45] Michael Steinbach, George Karypis, Vipin Kumar et al. 2000. A Comparison of Document Clustering Techniques KDD Workshop on Text Mining. Retrieved from https://hdl.handle.net/11299/215421.
[46] Shire Human Genetic Therapies. 2013. *Rare Disease Impact Report: Insights from Patients and the Medical Community.* Technical Report. *Shire Human Genetic Therapies.*
[47] Philip J. Vickers. 2013. Challenges and opportunities in the treatment of rare diseases. *Drug Discov. World* 14 (2013), 9–16.
[48] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.* 29, 12 (2017), 2724–2743.
[49] Agnieszka Wosiak, Kinga Glinka, and Danuta Zakrzewska. 2018. Multi-label classification methods for improving comorbidities identification. *Comput. Biol. Med.* 100 (2018), 279–288.
[50] Xi-Zhu Wu and Zhi-Hua Zhou. 2017. A unified view of multi-label performance measures. In *Proceedings of the 34th International Conference on Machine Learning, Vol. 70*. JMLR.org, 3780–3788.

[51] Pranjul Yadav, Michael Steinbach, Vipin Kumar, and Gyorgy Simon. 2018. Mining electronic health records (EHRs) a survey. *ACM Comput. Surv.* 50, 6 (2018), 1–40.

[52] Changchang Yin, Rongjian Zhao, Buyue Qian, Xin Lv, and Ping Zhang. 2019. Domain knowledge guided deep learning with electronic health records. In *Proceedings of the IEEE International Conference on Data Mining (ICDM'19)*. IEEE, 738–747.

[53] Min-Ling Zhang and Zhi-Hua Zhou. 2007. ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recog.* 40, 7 (2007), 2038–2048.

[54] Min-Ling Zhang and Zhi-Hua Zhou. 2013. A review on multi-label learning algorithms. *IEEE Trans. Knowl. Data Eng.* 26, 8 (2013), 1819–1837.

[55] Jiayu Zhou, Jimeng Sun, Yashu Liu, Jianying Hu, and Jieping Ye. 2013. Patient risk prediction model via top-k stability selection. In *Proceedings of the SIAM International Conference on Data Mining*. SIAM, 55–63.

[56] Donald W. Zimmerman. 1997. Teacher's corner: A note on interpretation of the paired-samples T test. *J. Educ. Behav. Statist.* 22, 3 (1997), 349–360.

[57] Damien Zufferey, Thomas Hofer, Jean Hennebert, Michael Schumacher, Rolf Ingold, and Stefano Bromuri. 2015. Performance comparison of multi-label learning algorithms on clinical data for chronic diseases. *Comput. Biol. Med.* 65 (2015), 34–43.

Received June 2020; revised January 2021; accepted January 2021