Predictive Modeling of Hospital Readmission: Challenges and Solutions

Shuwen Wang, Xingquan Zhu, *Senior Member, IEEE*

Abstract—Hospital readmission prediction is a study to learn models from historical medical data to predict probability of a patient returning to hospital in a certain period, *e.g.* 30 or 90 days, after the discharge. The motivation is to help health providers deliver better treatment and post-discharge strategies, lower the hospital readmission rate, and eventually reduce the medical costs. Due to inherent complexity of diseases and healthcare ecosystems, modeling hospital readmission is facing many challenges. By now, a variety of methods have been developed, but existing literature fails to deliver a complete picture to answer some fundamental questions, such as what are the main challenges and solutions in modeling hospital readmission; what are typical features/models used for readmission prediction; how to achieve meaningful and transparent predictions for decision making; and what are possible conflicts when deploying predictive approaches for real-world usages. In this paper, we systematically review computational models for hospital readmission prediction, and propose a taxonomy of challenges featuring four main categories: (1) data variety and complexity; (2) data imbalance, locality and privacy; (3) model interpretability; and (4) model implementation. The review summarizes methods in each category, and highlights technical solutions proposed to address the challenges. In addition, a review of datasets and resources available for hospital readmission modeling also provides firsthand materials to support researchers and practitioners to design new approaches for effective and efficient hospital readmission prediction.

Keywords—hospital readmission, predictive modeling, classification, clustering, electronic health records

✦

1 INTRODUCTION

Hospital readmission is defined as a hospital visit of a discharged patient being admitted again to the same or a different medical institution within a specific period of time, such as 30 days or 90 days, after the previous visit. A revisit usually implies an incomplete or unsuccessful treatment from the previous in-patient visit, therefore it is a defined metrics of the US healthcare systems [1]. Reasons behind hospital readmissions are complicated and various, among which many are avoidable especially those related to doctors, nurses, and healthcare system. For example, patients discharged ahead of the schedule are more likely to be readmitted [2]. Medication errors, like no proper prescriptions for necessary medicines when a patient was discharged, are also responsible for preventable readmissions [3], [4]. In addition, socioeconomic factors, including patients of both lower and higher socioeconomic status, are tied to the readmission risk [5], and a disparity in hospital readmission between ethnicity has also been observed [6], in which the possibility that Medicare patients, being readmitted to hospital after major surgeries, among Black patients is 19% higher than White patients.

A succession of hospital revisits, in a short period, are costly to patients and healthcare system [1]. Avoidable hospital readmission not only incurs $41.3 billion annual cost [7], it also places signature pressure to the medical resources as well as the high working intensity of medical staff.

In order to mitigate the severity of high readmission rates, the US federal government introduced a series of

plans [111]. Combining payment and readmissions through Hospital Readmissions Reduction Program (HRRP) is part of the initiatives to reduce readmission rates [112].

Following the HRRP initiative, many researches have been conducted to tackle the problem. One commonly used approach is to apply predictive models, such as logistic regression, to learn from historical data, and then predict the readmission possibility of a patient after being discharged from hospital [29]–[31]. Other advanced machine learning methods, such as decision trees [72] and deep learning [100], are also used for readmission prediction. Table 1 summarizes commonly used predictive models in the field. In general, eleven popular predictive model types are considered such as clinical rule based method, case-based reasoning, regression based method and deep learning methods. Among all those methods, majority research uses regression based methods (logistic regression), neural networks, and ensemble methods including bagging, boosting, random forest and gradient boosting. Due to complications of human diseases, some models are developed based on specific disease types, like heart failure [32], [94], [101], pneumonia [39], and organ transplantation [113].

Indeed, many predictive models haven been proposed and are reportedly effective under certain circumstances, but they often do not perform well as expected when being applied to new health records [114], [115]. Evidently, Medicare, under the HRRP plan, cut payments to 2,853 hospitals in 2019. Among 3,129 general hospitals being evaluated in the HRRP program, 83% of them received a penalty. Partially, this is because that a hospital readmission is a compounded outcome of many factors, and not all of them can be modeled by using computational approaches. On the other hand, most existing methods only focus on the

TABLE 1: A summary of predictive models used for hospital readmission prediction.

Predictive Model Types	Methods & Papers	Strength	Weakness
Clinical rule based methods	LACE Index, HOSPITAL score, B Score [8]–[27]	High transparency & interpretability	Low accuracy & limited discriminative power
Case-based reasoning	k-NN classifiers [28]	High interpretability. Easy to maintain & adapt to changes	Time-consuming with high-dimensional medical data
Regression based methods	Logistic Regress [6], [8], [9], [11], [14], [16]–[19], [22], [29]–[71]	Easy to interpret & efficient to train with large patient records	Ineffective for modeling nonlinear relationships
Decision tree methods	Decision tree, Classification and regression tree (CART) [9], [14], [50], [63], [67], [72]–[75]	Transparency & Interpretability. Automatic identifying important factors	Unstable & sensitive to data. Low accuracy
Bayesian methods	Naive bayes, Bayesian conditional probability [9], [51], [63], [72], [76], [77], [77]–[81]	Transparency & incorporate domain knowledge	Computationally expensive & low accuracy
Neural networks	multi-layer neural networks, RBF network [15], [40], [63], [77], [79], [82]–[86]	High accuracy & effective for high dimensional data	Low training efficiency. Poor transparency & interpretability
Margin classifiers & kernel machines	Support vector machines (SVM) [14], [16], [18], [63], [87]–[93]	Model complex high-dimensional patient records	Low interpretability for clinical decisions
Ensemble methods	Bagging, Boosting, Random forest, Gradient boosting [9], [14], [14], [18], [18], [32], [67], [78], [87], [89], [90], [90], [92], [94]–[96]	Improve accuracy over single models	High computational costs & lack of transparency due to combined decisions
Cost-sensitive classification	Bayesian optimal decision [28], [97]	Tackle imbalanced class distributions & consider medical costs in modeling	Time-consuming on enormous skewed data
NLP methods	Topic models [98]	Work with unstructured data like discharge summaries	Cannot handle structured data
Deep learning methods	CNN, LSTM, deep contextual embedding [92], [97], [99]–[105]	High accuracy. Modeling complex relationships across multiple visits	Long training time for parameter tuning & require large volumes of data
Clustering methods	Hierarchical clustering [106]–[110]	Do not require labels. Interpret data distributions for large volume data	Ineffective for high dimensional data

modeling and learning aspects of the problem, failing to address the underlying challenges for readmission prediction. A recent study [116] systematically reviews 41 readmission prediction models (including 17 models for all patient risk prediction and 24 models for patient specific populations). Their investigations suggest that using Electronic Medical Records (EMR) data have better predictive performance than those using administrative data. However, technical challenges and solutions of readmission prediction still remain unaddressed, and are largely unclear.

The above observations motivate our study to review challenges and solutions for hospital readmission prediction. In order to systematically address the challenges, we proposes a taxonomy to categorize selected modeling methods, and summarize how existing approaches handle different challenges. In addition to the review of methodologies, we also outline public datasets available for model building and evaluation. The survey provides a comprehensive review for researchers and practitioners to understand the state-of-the-art in the filed, as well as designing new approaches to tackle hospital readmission prediction.

The rest of the paper is organized as follows. Sec. II proposes a taxonomy of hospital readmission challenge. Secs. III to VI outline each challenge as well as methods used to tackle them, including algorithms, applications, and performance. Public datasets are summarized in Sec. VII, and we conclude the paper in Sec. VIII.

2 PROBLEM DEFINITION & TAXONOMY

2.1 Problem Definition

Formally, we use an m dimensional vector $\mathbf{x}_i^{[a_j, d_j]} \in \mathbb{R}^m$ to denote a patient i and his/her j^{th} hospital visit, where $[a_j, d_j]$ denote admission and discharge/disease time of the jth visit, respectively. The m dimensional vector includes symptoms, treatment, medical notes, medications, procedures, and a variety of electronic health record (EHR) information of the patient, carried out during the jth visit. A

hospital readmission of a patient i is referred to a visit j, whose admission time a_j is within a certain window, typically 30 days or 90 days, following the discharge of the previous visit d_{j-1}, i.e. $a_j - d_{j-1} \leq 30$. Given a number of patients and their visit records, hospital readmission prediction **aims** to accurately predict the readmission probability of a patient after being discharged from the current visit.

2.2 Taxonomy of Readmission Prediction Challenges

Accurate prediction of hospital readmission is a significant challenge, mainly because health records and diseases are inherently complex in nature. For example, not all medical records are organized in feature format, and medical treatments and procedures are complicated making feature engineering a daunting task. Meanwhile, HIPPA (Health Insurance Portability and Accountability Act) regulations and policies also raise challenges for data usage and sharing. In order to carry out systematic review of computational models for hospital readmission prediction, we propose a taxonomy in Fig. 1 to summarize main technical challenges into two types and four categories: data challenges and model challenges. Data challenges explain the nature of readmission prediction and data related issues for model learning. Model challenges, on the other hand, explain model training, interpretation, and implementation issues.

2.2.1 Data Challenges

Data imbalance, locality and privacy represent the first data challenge for hospital readmission prediction. Databases used for hospital readmission prediction usually contain a large number of patient visit records collected from healthcare providers across disease types, ages, as well as length of in-patient treatment [117], [118]. Although patients' medical records are gathered and presented in details, the class distributions are often imbalanced [89] meaning readmission visits are much fewer than normal visits. Such large-scale but imbalanced data is an obstacle for learning accurate
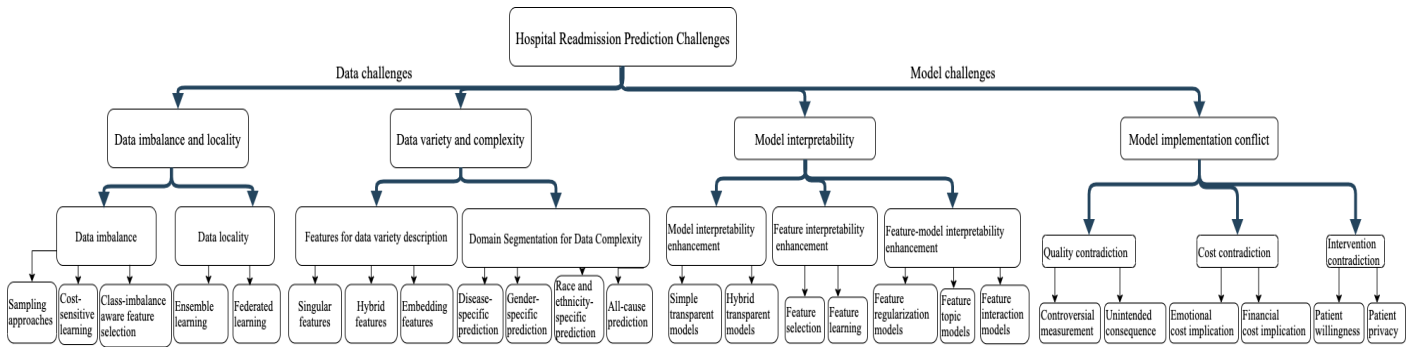
Fig. 1: A taxonomy of challenges of computational methods for hospital readmission prediction.

predictive models [119]. In addition to the data imbalance, the regional designation of hospitals adds additional complexity to this issue. Demographics of the hospital served regions often have great disparity across regions, where the top two characteristics reflecting the difference are race and average income of the residents. In other words, due to regional restrictions, some hospitals may have majority patients from certain ethnic groups, like Latino or Asian. Such difference greatly prompts the locality of hospital, and further complicates the data challenge. Medical information collected from those hospitals is biased for readmission research. Besides, the income gap between residents also affects medical expenses [120], [121]. Low income cohort may be vulnerable to medical services such as medical insurance, implying that there may exit a higher possibility that they are not willing to continue their medical treatment after discharge, even their conditions do not improve. This resistance to treatment makes the collected data lack integrity, and impose challenges to predictive modeling.

Data variety and complexity is another challenge to predict readmission. In order to train predictive models, it is necessary to collect information/features to characterize object of interest (such as patient, disease, hospitals, patient visits *etc.*). Due to inherent complexity of the healthcare systems, and dependency between readmission and other items, such as diseases, comorbidities, preexisting conditions, *etc.*, a predictive model should consider a variety of data objects for learning [101]. In addition to the data variety, a hospital readmission is a compounded outcome of many issues, including patient life styles, disease types, hospital treatments *etc.*, making accurate readmission prediction a complex task. Accordingly, data resources and learning may choose to focus on different medical conditions, by using domain segmentation, such as disease-specific prediction, gender-specific prediction *ect.*.

2.2.2 Model Challenges

At the model level, transparency and conflicts must be properly resolved before predictive models can be put into real-world usages. Interpretability of a model refers to comprehension of model decision process for human understanding [122]. A model resembling to human perception and decision logic, such as a decision rule or tree, is always preferred in medical domains. From practical aspect, it is also important to understand why some patients are predicted as high risk after being discharged from the hospital

while others are not. Many models perform well on benchmark datasets, but their decision logistic is intransparent. As a result, their performance may deteriorate after being applied to new data, due to the blind decision process.

The second model challenge is to resolve conflicts between models and implementation, such as qualify of services, cost reductions, patient emotion, *etc.*. The HRRP initiative intends to encourage hospitals to reduce their readmission rates [123]. It is expected that, as unnecessary readmission being reduced, the net income of hospitals can increase significantly [124]. However, the reduction of hospital readmission does not directly imply better qualify of services. A short-term cost reduction may result in a higher cost in the long run, and jeopardy the HRRP objective. Meanwhile, as more stakeholders are turning to predictive models, patients data are being collected for sharing and analysis [125], resulting in increasing privacy concerns [126].

3 DATA IMBALANCE, LOCALITY AND PRIVACY

Data imbalance and locality are two common biases in medical data, which are known to impose significantly challenge to predictive models [83].

3.1 Data Imbalance

Data imbalance refers to a phenomenon where datasets used to train a predictive model have an biased class distribution. In many cases, one type of samples (*i.e.* positive class) are significantly less than other types of samples. This is partially caused by the reality that disease samples are only a small percentage of the whole population, and naturally results in the class imbalance. Learning models with imbalanced class distributions is a significant challenge, because most algorithms are affected by frequency bias and pay more attention to majority class samples [127]. Data imbalance tends to force the classifier to classify all samples as normal, in order to satisfy the defined objective function, such as minimizing the classification errors [128]. Common solutions are to rebalance samples in different classes, by manipulating data populations (sampling approaches) or classification outcomes (cost-sensitive learning).

3.1.1 Sampling Approaches

Sampling approaches change data distributions to balance samples in different groups in order to tackle the data imbalance challenge. Common sampling solutions are to

either drop majority class samples, repeat samples from minority class, or create synthetic samples for minority class.

•Random Sampling Approaches Random Oversampling (ROS) and Undersampling (RUS) are the simplest ways broadly used in numerous domains, such as hospital readmission prediction and fraud detection. Fig. 2 describes the sampling methods in which Fig. 2a represents the original imbalanced data with blue dots as the prevalent class and orange rectangular as the minority class. Random undersampling, in Fig. 2b, involves randomly selecting examples from the majority class and down sampling them in the training dataset according to sampling strategy defined as the ratio of the minority class to majority class. Due to the loss of vast quantities of discarded data, a loss in classification performance can be resulted from the boundary ambiguity between the two classes. Random oversampling, in Fig. 2c, duplicates minority samples to form a balanced training set. Due to sample duplication, ROS may lead to over-fitting in the training.

A research study [92] using Medical Information Mart for Intensive Care III (MIMIC-III) database [129] shows that, by using undersampling, their model achieves 0.642 AUC score for ICU patient readmission. Another study [63] investigates RUS sampling and five supervised learning methods, decision trees, naive bayes, logistic regression, neural networks, and support vector machines (SVM) for risk modality and hospital readmission prediction. The results show that, overall, neural networks achieve best performance for both risk modality and hospital readmission prediction. In addition, using AdaBoost to change the weight of instances for learning results in 3% and 6% improvement for readmission and mortality predictions, respectively.

•Synthetic Samples For RUS and ROS sampling, the dropped/duplicated instances are part of the original training data, meaning that there is a potential risk that sampling will introduce information loss or bias. Synthetic sample generation, on the other hand, will generate new samples similar (but not identical) to the training data. Two common approaches to generate synthetic samples are Synthetic Minority Over-sampling Technique (SMOTE) [130] and Random Over Sampling Examples (ROSE) [131]. As shown in Figs. 2d and 2e, SMOTE is an improvement of Random over sampling approach. A minority class sample x is randomly selected, with its k nearest minority class neighbors being determined. Then the synthetic instance is created by choosing one of the k nearest neighbors b at random and connecting x and b to form a line segment in the feature space. The synthetic instances can be generated as a convex combination of the instances x and b. ROSE, on the other hand, generates synthetic samples using a conditional density estimating the positive and negative classes. A randomly selected minority sample is used as the center of a created density function, and synthetic samples are the ones generated from the estimated density functions.

Using SMOTE to generate synthetic instances to balance positive and negative samples for 30 day readmission prediction has been studied [102] by using a UCI hospital readmission dataset [132]. The experiments show exceptionally higher AUC values (0.974) than results from other studies (normally around 0.7 AUC range). One possible reason is that UCI readmission dataset has a relatively bal-

anced sample distributions because 11.2% samples belong to positive class (readmission), whereas in other dataset, such as National Readmission Database [117], the positive ratio is much smaller. By using different sampling approaches, including RUS, ROS, and ROSE, a method [90] comparatively studies the three methods using UCI readmission dataset [132], using different classifiers, such as SVM, random forest, gradient boosting, and regression and partition trees. The results show that ROSE is significantly worse than other approaches (including original data without any sampling). In addition, RUS and ROS have comparable performance, and both frequently outperform models trained from original imbalanced dataset.

•Random Class Balancing A new technique to tackle data imbalance called Random Balance is proposed inspired by the idea of randomly deciding class proportions [133]. In this approach, Data sampled from training dataset is used to train every member of the Random Balance ensemble and the augmentation is completed with synthetic samples created by SMOTE. The probabilities of selecting an instance from minority class and majority class are presented in Eq. (1), where N is the total samples with p positive instances and n negative instances.

$$P_{mi} = \frac{1}{N-3}(N - \frac{p+3}{2} - \frac{1}{p}); P_{ma} = \frac{1}{N-3}(N - \frac{n+3}{2} - \frac{1}{n})$$
(1)

Although preprocessing techniques are usually used to restore the balance of the class proportions to a given level, Random Balance relies on completely random proportions. The key step in this method is that both the size of majority and minority classes is randomly set, followed by SMOTE and Random Undersampling in order to increase or reduce the size to match the ideal class size. This methods alleviates the problem of deleting important examples by being repeated multiple times. Despite its simplicity, this methods outperforms other advanced ensemble methods.

3.1.2 Cost-Sensitive Learning

Data imbalance challenge can also be mitigated through the change of learning algorithms. The ultimate goal of machine learning is to minimize/satisfy the loss function. If the loss of misjudgment on minority samples is increased in the loss function, the model can be adjusted to better identify minority samples. Cost sensitive learning is one approach to adjust the loss function of the learning algorithm to make the model more sensitive to minority samples.

Cost-sensitive confusion matrix, in Table 2, is a common way to adjust the loss function. Instead of treating all misclassification equally, the cost matrix differentiates costs associated to different types of mistakes. For example, in Table 2, μ is defined as the cost of a single False Positive (FP) and λ is defined as the cost of a single False Negative (FN). In order to support cost-sensitive readmission prediction, one can either apply the cost matrix to the posterior probability (classification outcomes) or the learning objective function to minimize the misclassification costs.

•Posterior Probability Adjustment: In a research [28] predicting 30-day readmission for patients with Chronic Obstructive Pulmonary Disease (COPD), the cost matrix is used to adjust the posterior provability for prediction. Assuming
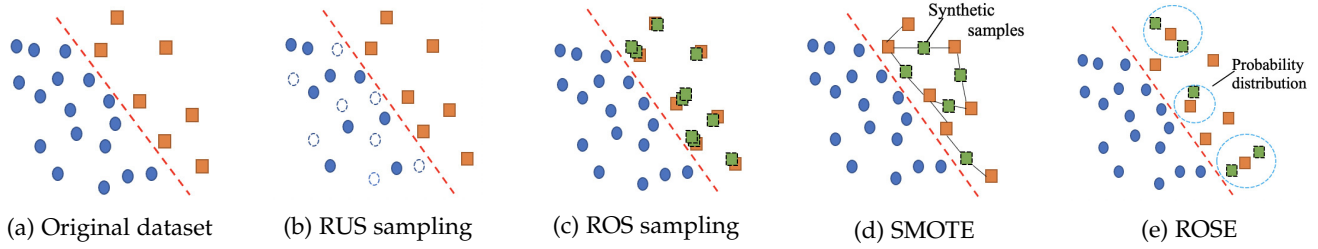
Fig. 2: (a) A dataset with imbalanced class distributions; (b) random under sampling (RUS); (c) random oversampling (ROS); (d)Synthetic Minority Over-sampling Technique (SMOTE); and (e) Random Over Sampling Examples (ROSE).

TABLE 2: Cost sensitive confusion matirx. λ and μ are positive values denoting misclassification costs between two groups (positive *vs.* negative).

	Predicted positive	Predicted negative
Actual positive	0	λ
Actual negative	μ	0

x denotes a patient (or a visit) and $P(True|x)$ denotes the posterior probability of a generic classifier classifying x as being a readmission (True), the prediction of readmission is based on the adjustment of posterior probability in Eq. (2) where the threshold of classification is presented as $\frac{\mu}{\lambda+\mu}$.

$$Prediction = \begin{cases} True, & If\ P(True|x) > \frac{\mu}{\lambda+\mu} \\ False, & Otherwise \end{cases} \quad (2)$$

By applying the above design to several generic classification models, including Naïve Bayes (NB), Random Forest (RF), Support Vector Machines (SVM), k-Nearest Neighbors (kNN), C4.5, Bagging with REPTree, and Boosting with Decision Stump, their experiments [28] show that cost-sensitive classification is effective in minimizing costs and the cost matrix is more desirable than commonly used AUC, when evaluating hospital readmission systems.

• Learning Objective Function Adjustment: In order to directly integrate the costs to the learning, a cost-sensitive formulation [97] is used to train a multi-layer perceptron fed by learnt features through convolutional neural networks (CNN) and statistical features via feature embedding to predict hospital readmission. To tackle the misclassification problems of the minority class, the cost sensitive deep neural network (CSDNN) consists of one input layer, one output layer, and multiple hidden layers with fully-connected neurons formulated using weight matrix $\mathbf{W}$. A modified cross entropy is set as the loss function shown in Eq. (3), where n is the total number of patients, $y_i \in \{T, F\}$ denotes the label of a patient x_i, where True (1) means a readmission or False (0) otherwise. $P(\hat{y}_i|\mathbf{x}_i, \mathbf{W})$ denotes the posterior probability of the ith patient, and $\mathbf{C}(\hat{y}_i, y_i)$ defines the classification cost of x_i with respect to the current prediction y_i.

$$\mathcal{L} = -\frac{1}{n}\sum_{i=1}^{n}\log[\sum_{\hat{y}_i\in\{T,F\}} P(\hat{y}_i|\mathbf{x}_i, \mathbf{W})\mathbf{C}(\hat{y}_i, y_i)] + \frac{\lambda}{2}\|\mathbf{W}\|_2^2 \quad (3)$$

The above approach is validated on two real-world medical datasets from Barnes-Jewish Hospital and the results prove that their prediction of readmission performs significantly better than several baselines with a much higher Area Under the ROC Curve (0.70 AUC score) than baselines.

3.1.3 Class-imbalance aware feature selection approach

A readmission prediction algorithm Joint Imbalanced Classification and Feature Selection (JICFS) is proposed to construct the loss function and applied sample weight to handle class-imbalance problem [134]. To be specific, this approach solves the readmission prediction problem with imbalance class by constructing an improved margin-based loss function, which involves two parameters α and γ to reduce the weight of loss assigned to easily classified samples.

$$\min_{w\in\mathbb{R}^m} f(w) = \sum_{n=1}^{N}(\alpha_n log(1+exp(-y_n\gamma(w^T z_n)))/\gamma)+\lambda\|w\|_1 \quad (4)$$

In Eq. (4), the class-imbalance aware feature selection approach contains α, γ and λ three parameters from objective function to obtain coefficient matrix w in order to realize feature selection from class-imbalance data. The method was compared with different class-imbalance learning algorithms based on six real-world readmission datasets and it always can achieve better performance on each dataset.

3.2 Data Locality and Privacy

While data imbalance is concerning the learning target (or class labels), data locality, on the other hand, is associated to the sample distributions (or independent variables). At population level, data for readmission prediction might be collected from a local/regional hospital, where the demographics of the patient body naturally introduce bias. At individual level, when collecting data for each patient, the hospital visits used to characterize the patient may also introduce bias. At the ministration level, regularizations also impose restriction for data sharing across hospitals, making it difficult to learn good models from local data.

A study [64] considers two types of discharge sampling, first time discharge *vs.* all discharges. In other words, the research compares using a single discharge *vs.* using all discharges of patient visits to represent each patient for learning. Experiments show that using the first discharge per patient underestimates the readmission rate, and may result in misleading measures of model performance.

Common approaches to tackle data locality and privacy are to employ ensemble learning or federated learning. The former trains multiple models from local datasets and combine them for prediction, whereas the latter trains one model from multiple decentralized/localized datasets.

3.2.1 Ensemble Learning

Ensemble learning combines multiple base models for prediction. Typical approaches include bagging, boosting, and stacking [135]. Bagging trains base models separately (often in parallel), and then combines them using weighted (or unweighted) majority voting. Boosting, on the other hand, trains base models in a sequential and progressive manner, so a later trained base model is improved based on an earlier trained base model. Stacking is a meta learning approach, which uses base classifiers to generate outputs, and then retrains another model from the outputs for prediction.

In [103], a localized sampling approach is proposed to allow sampling process to focus on instances difficult to classify. By using localized sampling to generate balanced datasets, this approach is validated using data collected from several South Florida regional hospitals. A joint ensemble-learning model [96] combines weight boosting algorithm with stacking algorithm, and compares three major baseline (1) the LACE index, (2) RandomForest-Lasso-SMOTE, and (3) SMOTE (which uses SMOTE to replace bagging for data samping) on national Hosptial Quality Monitoring System (HQMS) database (including 651,816 records after data processing). The results show that LACE (which is commonly hospital score systems) has the least performance, confirming that machine learning is useful for hospital readmission prediction. Meanwhile, bagging with weight boosting and stacking shows clear benefits on high dimensional medical data with imbalance class distributions and imbalanced misclassification costs.

3.2.2 Federated Learning

Different from ensemble learning which focuses on combining models trained from local datasets, federated learning tackles data locality challenge by allowing multiple data holders to collaboratively train a model, and keep participant data in private without exchanging raw data. A research [93] proposes to use federated learning to build a global model to predict hospitalizations due to heart diseases using patient electronic health records (whether a patient will be hospitalized within one or two years, prior to the time of prediction). To tackle the problem, they formulate the problem as a sparse support Vector Machine (SVM) learning problem, with the following objective function:

$$\min_{\mathbf{w},w_0} \sum_{i=1}^{n} f_i(\mathbf{w}, w_0, x_i) + 0.5\tau \|\mathbf{w}\|_2^2 + \rho \|w_0\|_1 \quad (5)$$

In Eq. (5), $\mathbf{w} \in \mathbb{R}^d$ and $w_0 \in \mathbb{R}$ are weight vectors (parameters) defining the classifier. $f_i(\mathbf{w}, w_0, x_i) = \max\{0, 1 - \ell(\mathbf{w}^T x_i + w_0)\}$ defines a hinge loss for instance x_i. τ and ρ are penalty coefficients enforce L_2 norm and L_1 norm constraints on the parameters. In federated learning setting, multiple agents (or hospitals) each hold their own private data, so $x_i, i = 1, \cdots, n$ are not presented to any single agent. The key is to use local data from each single agent, to learn $(\mathbf{w}, w_0)$, which optimizes the objective function in Eq. (5). To solve their own $(\mathbf{w}, w_0)$, based on local data, all agents will combine learned solutions to create a global $(\mathbf{w}, w_0)$. In addition to the predictive model, their method also find important factors/features associated to hospi-

talizations, such as "Age" and "Admission due to Other Circulatory System Diagnoses" etc.

Similarly, another research [107] studies federated learning for patient mortality and hospital stay time prediction, using distributed electronic medical records. They propose a distributed clustering to separate patients into clinically meaningful communities (both communities and data are local). Experiments show that this approach results in a higher predictive accuracy and lower communication cost, comparing to other federated learning methods.

4 DATA VARIETY AND COMPLEXITY

A hospital readmission is the outcome of numerous compounding factors, involving patients, diseases, care providers etc.. Collecting representative training data is an important step for modeling. Therefore, the second major data challenge for readmission prediction is to properly characterize training samples and learning tasks.

4.1 Features for Data Variety Description

Patient records from electronic medical records (EMRs) and administrative databases usually include various information from basic personal demographics to professional medical diagnosis [136]. In this section we summarize features used to describe samples into three major categories: singular features, hybrid features, and latent features. The detailed feature types are listed in Table 3, including demographics, admission and discharge, clinical, hospital etc.

4.1.1 Singular Features

Singular features include the first five types of features listed in Table 3. They are single factors, indicators, statistics or textual information, used to describes patients, diseases, medical procedures, hospitals etc.. Demographics is defined as features including basic patient information such as age, race, contact information in case emergency condition happens, insurance information indicating whether a patient has insurance or not. Such demographics provides general health information representing patients' health condition at the time of the hospital visit.

After a patient is admitted to hospital, admission and discharge information will record administrative features related to the visit, such as the dates of admission/discharge and whether the patient is admitted through emergent department and so on. Clinical information records patients' symptoms and procedures for in-patient treatment. The information of the hospitals to which patients are admitted are summarized as hospital information, which includes statistic features such as percent readmission within a time period, hospital locations, and ownership.

Textual information is another type of singular features consisting of literal summary like discharge summary, physician notes. Such information is often stored in unstructured format, and provide comprehensive information not detailed in administrative and clinical features, which are often stored in structured format (such as a table). For most methods, they simply combine single features into one feature set to train predictive models.

Singular features are ideal for training logistic regression classifiers because features are rather independent and

TABLE 3: A summary of features used to describe patients, hospitals, hospital visits for readmission prediction.

Feature Types & Reference	Feature Subtypes	Feature Information	Structured
Demographics features [8]–[11], [13]–[22], [29]–[32], [34]–[41], [43]–[62], [65], [66], [68]–[77], [79], [80], [82], [87]–[91], [94], [95], [100], [101], [109], [113], [137]–[144]	Basic demographics information	Age, Gender, Race, Education, Marital status	✓
		Language, Income/Financially issues	✓
	Contact information	Family size/members, Zip code/address, Family doctor	✓
	Insurance information	Insurance provider, Mode of payment	✓
	General health information	Medical conditions, Allergies, Current medications	✗
		Completed outpatient appointment rate, Nursing home needed	✓
	Social history	Smoking, Alcohol, Living situation, Employment	✗
Admission and discharge information [8]–[22], [29]–[32], [34]–[47], [49]–[62], [65], [66], [68]–[73], [75]–[77], [79], [80], [82], [87]–[91], [94], [95], [99]–[101], [109], [113], [137]–[144]	Admission information	Admission date, First hospital visit, Elective	✓
		Number of admissions in a past time period, Cost-weight of previous admission, Diagnosis of last admission	✓
	Discharge information	Total charge, Discharge date	✓
		Transfer, Discharge disposition	✓
Clinical information [9]–[11], [13], [14], [16]–[20], [22], [29]–[31], [33]–[38], [40]–[62], [65], [66], [68]–[80], [82], [87]–[91], [94], [95], [95], [101], [101], [106], [109], [113], [139]–[145]	Payment code information	ICD-10codes, ICD-9 codes	✓
		APR-DRG codes, DRG codes	✓
	In-hospital symptom	Vitals and lab values	✓
	Rhythmic features	Mean 10 most active hours, Total sleep time, Sedentary time	✓
	Medical images	Ultrasound exam	✗
Hospital information [34], [53], [65], [71], [77], [87], [106]	Hospital statistics	Total number of admissions, Number of patients	✓
		Percent readmission within a time period (30 days etc.)	✓
	Hospital characteristics	Ownership of hospital, Rural/Urban	✓
Textual information [99], [101]		Discharge summary, Physician note, Date the note was initiated, Subject of the note, Prescription medication, Date of the prescription update, Dose and strength of the drug	✗
Hybrid information [8]–[27], [70]		LACE, HOSPITAL score, Charlson comorbidity score, Elixhauser Comorbidity Index, Baltimore Score	✓
Latent features: [40], [104]		Clinical nodes embedding features, Patient embedding features	✓

easy to interpret. A research [50] uses conditional logistic regression, combined with patient demographics, clinical information and categorical data, for model development. After correcting the data imbalance using undersampling, the applicability of conditional logistic regression is tested and compared with several standard classifiers, including standard logistic regression, random forests, SVM and stepwise logistic regression. The models are performed on two different prediction variable sets, original variables and selected ones, in order to achieve efficient readmission management with identified features.

Another similar study [54] also employs demographics, clinical data, and textual notes from administrative data to predict 30-day hospital readmission for maintenance hemodialysis patients. Data are collected from University of North Carolina Hospital (2008 to 2013) and are recorded as means and SDs for continuous variables whereas textual variables are presented as frequencies and percentages. The study trains models using univariate and multivariable binary logistic regressions which only consider variables with a univariate P value <0.20. Results show that two multivariable logistic regression models outperform univarible models with an AUC of 0.79 (95% CI, 0.73 to 0.85).

4.1.2 Hybrid Features

Apart from features representing single factors, hybrid features have a long history being used in the medical field to combine multiple factors to form a feature indicator for prediction. LACE index [36], HOSPITAL score [13], Charlson comorbidity score [25], Elixhauser Comorbidity Index [24] and Baltimore Score (B Score) [23], are the commonly used hybrid features. In many predictive models, hybrid features are further integrated as independent variables to train classifiers for prediction.

•Charlson Comorbidity Index [25] was first proposed in 1987 to classify prognostic comorbidity in longitudinal study. In the original version, 19 types of comorbidities are identified and assigned weights according to the adjusted relative risk of one-year mortality. The final single comorbidity score is the sum of weights. ICD-9-CM codes are assigned for categories of the Charlson Comorbidity Index and the number of categories is reduced from 19 to 17 by combining 'Leukemia' and 'Lymphomas' into 'Any tumor' [146]. Later on, a multi-step process [26] is conducted to develop ICD-10 codes to define the index and Table 4 shows the revised version of Charlson Comorbidity Index.

A research [31] combines Charlson Comorbidity Index with patients demographics as indicators for readmission risk prediction. A retrospective multiple regression analysis of 958 non-pregnant adults is conducted based on the data abstracted from hospital administrative sources and electronic medical records. After comparing with other models, using AUC values for validation, the study concludes that poly-pharmacy and higher Charlson score at admission can predict readmission risk as good as or better than published risk prediction models.

TABLE 4: Latest version of Charlson comorbidity index

Score	Comorbidity
1	Previous myocardial infarction Cerebrovascular disease Peripheral vascular disease Diabetes without complications
2	Congestive heart failure Diabetes with end organ damage Chronic pulmonary disease Mild liver or renal disease Any tumor (including lymphoma or leukemia)
3	Dementia; Connective tissue disease
4	HIV infection Moderate or severe liver or renal disease
6	Metastatic solid tumor

•LACE Index In order to create an easy-to-use index to quantify patients' risk of readmission or death, after being discharged from hospital, a secondary analysis of a multi-centre prospective cohort study is performed based on 48 patient-level and admission-level variables [36]. Multivariable logistic regression, fractional polynomial function and points system [147] are applied to derive and validate the

LACE Index in Table 5 which is the sum of the scores from four aspects: length of stay (L), acuity of the admission (A), Comorbidity of the patient (measured with the Charlson Comorbidity Index) (C) and emergency department visit in the six months before admission to hospital (E). $C-$statistic with 95% confidence intervals is used to evaluate the accuracy of the index. The results indicate that LACE Index has moderate discrimination for early death or readmission, therefore it can be used as a measurement to predict risk of patient death and unexpected readmission.

The feasibility and strength of LACE index in predicting 30-day hospital readmission is further studied using anonymised patient data obtained from hospital information system [8]. A positive correlation is observed by univariate logistic regression meaning the increment of LACE index score could greatly result in larger possibility of readmission than elective or day case readmissions (12.8%, 5.7% and 1.8% respectively). In addition, the significance of LACE index is also proved by multivariate logistic regression with an AUC of 0.773 (95% CI 0.768 to 0.779) and R^2 of 0.180.

TABLE 5: LACE index

Attribute	Value	Points	Attribute	Value	Points
	<1	0		1	1
	1	1	C	2	2
	2	2		3	3
L	3	3		≥ 4	5
	4-6	4		0	0
	7-13	5		1	1
	≥ 14	7	E	2	2
A	Yes	3		3	3
C	0	0		≥ 4	7

•HOSPITAL Score, as shown in Table 6, consists of seven independent variables, including hemoglobin at discharge, discharge from an oncology service, sodium level at discharge, procedure during the index admission, index type of admission, number of admissions during the last 12 months, and length of stay. It is derived and validated as a model in a dataset with a total of 12,383 patients discharged from the medical services of the Brigham and Women's Hospital [13]. The seven factors are determined by a series of models: A multi-variable regression model, followed by a regression coefficient-based scoring method and finally a backward multi-variable logistic regression analysis. The internal validation confirms that HOSPITAL score has potential to easily identify patients with a risk of readmission.

An international study validates the HOSPITAL score based on 117,065 adult patients from 9 hospitals across 4 countries [10]. The score is verified in the logistic regression from three aspects: overall performance, discriminatory power and calibration. Overall, the study confirms the discriminatory power of the HOSPITAL score in predicting avoidable readmission, with a C-statistic of 0.72 (95% CI, 0.72-0.72) and its prediction of potentially readmission matches the observed proportion with an excellent calibration (Pearson χ^2 test P=0.89).

•Elixhauser Comorbidity Index Because hospital readmission is often tied to complication of disease comorbidity, Elixhauser Comorbidity Index [24] was proposed in 1998 to measure comorbidities in large administrative inpatient datasets. This comorbidity algorithm is developed and tested on the data consisting of 1,779,167 adult patients from

TABLE 6: HOSPITAL score

Attributes	Score
Hemoglobin < 12g/ml or 120g/L at discharge	1
Discharge from an oncology service	2
Sodium < 135 mEql/L at discharge	1
Procedure performed during hospital stay (any ICD coded procedure)	1
Index admission type: nonelective	1
# of hospital admission during the previous year	
0	0
1-5	2
>5	5
Length of stay ≥ 5 days	2

438 acute care hospitals in California. Although original Elixhauser Comorbidity index addresses some limitations from previous measures, it requires 30 binary variables, which limits its application. Later on, it is revised as a single score for administrative data using backward stepwise multivariate logistic regression [27].

By using Elixhauser Comorbidity index, combined with age and gender date, a hierarchical logistic regression models [84] predicts 30-day readmission for patients with acute myocardial infarction (AMI), congestive health failure (HF) and pneumonia (PNA). The comparisons using four models, including 1) the hierarchical logistic regression model; 2) XGBoost model with binary ICD-9 codes; 3) a feed-forward ANN model trained on dummy variable representation of ICD-9 code; and 4) an ANN models trained based on latent ICD-9 codes variables, show that the AUC for hierarchical logistic regression is 0.68 (95% CI 0.678, 0.683) and it is improved by the fourth model to 0.72 (95% CI 0.718, 0.722).

•Baltimore Score (B Score) Readmission risk-assessment tools including LACE index, HOSPITAL score and Elixhauser Comorbidity index have been proven effective in predicting patient readmission risk. However, they mainly make predictions based on a portion of patients' features such as length of stay in hospitals and comorbidities. As a result, they may not be able to consider a large number of factors, and some important characteristics may be ignored.

Baltimore score (B score) is a learning based model using thousands of health data variables [23]. The Baltimore score (B score), invented by researchers from the University of Maryland Medical System (UMMS), is an easily implemented machine learning score to calculate standard readmission risk-assessment scores in real time to predict 30-day unplanned readmissions. This model is individualized for each of three University of Maryland Medical System hospitals in different settings and 382 variables are drew for the final model including demographics, lab test results, *etc.*, from more than 8,000 possible data variables. The research compares the B-score readmission risk level with the actual readmission rates of the three hospitals and the predictions derived from other plans. In the three hospitals, despite the different settings, the overall B score is better than other scores in identifying patients at risk of re-admission, and is the most accurate among the highest-risk patients. The 10% of patients with the highest risk of B score at discharge have a 30.7% chance of unplanned readmission. Similarly, the 5% of patients with the highest B scores at discharge have a 43.1% chance of being misdiagnosed again. The AUROC of the B score is 0.72 (95% CI, 0.70-0.73) and it increases to 0.78

(95% CI, 0.77-0.79) at discharge (all P < .001) compared with the 0.63 (95% CI, 0.59-0.69) AUROC for HOSPITAL, 0.64 (95% CI, 0.61-0.68) for Maxim/RightCare, and 0.66 (95% CI, 0.62-0.69) for modified LACE score. As a result, the study concludes that clinicians can use patient data to calculate B Score and further reduce adverse events and improve patient safety.

4.1.3 Embedding Features

Different from singular or hybrid features which measure one or multiple characteristics of objects as one score for evaluation, embedding features (also called latent features) use a feature vector to represent each object. This provides a much more comprehensive capability in describing different types of objects, such as patient, hospitals, diseases *etc.*, than singular or hybrid features.
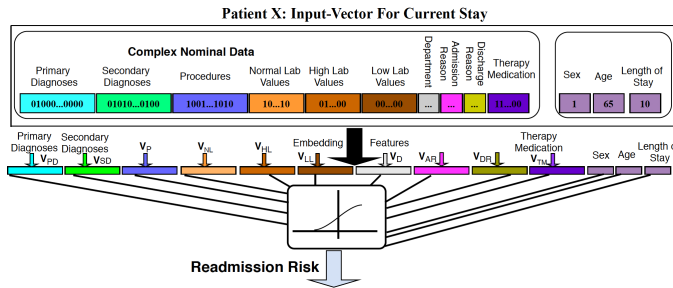

Fig. 3: Clinical object embedding based readmission risk prediction framework [40]. Each categorical clinical object, such as primary diagnose, is embedded as a feature vector. Numerical features, such as, age, remain original form. All features are fed into a regression framework to estimate readmission risk.

•Clinical Object Embedding Clinical objects have a variety of types, such as diseases, medical procedures, abbreviations, jargon from doctors, hospitals and other care providers. After being admitted to hospital, medical information like diagnoses and procedure codes are recorded as individual features, resulting in a large feature space ineffective for machine learning algorithms to train a model. Alternatively, embedding features can downsize the feature space and simplify the dependencies between input features and target variables.

Instead of using clinical data as features, a latent embedding based framework, as shown in Fig. 3, is proposed to embed nominal clinical data for hospital readmission prediction [40]. The framework uses latent embeddings of the nominal parts of data, such as diagnosis and procedure codes (ICD-10). An $n \times m$ matrix is organized using clinical data where each of the n rows contains data of a patient during his/her stay in the hospital and each of the m columns denotes one feature. For numerical features, such as age and length-of-stay, the feature are numerical values. For categorical feature, such as diagnose code, one-hot encoding is used to indicate whether a specific code appears in the current visit (1), or not (0).

To calculate latent embeddings, a data covariance matrix, in Eq. (6), is first calculated to find relationship between features. In this equation, $X \in \mathbb{R}^{n \times m}$ is an $n \times m$ sparse binary matrix standing for nominal data from the data

matrix, μ_X is the means vector and σ_X is the deviations, and $\otimes$ denotes the outer vector product.

$$X_{Cov} = \frac{X^T X - m \cdot (\mu_X \otimes \mu_X)}{(m-1) \cdot \sigma_X \otimes \sigma_X} \tag{6}$$

Given covariance matrix X_{Cov}, latent embedding is calculated by applying Singular Value Decomposition to decompose X_{Cov} in Eq. (7), where $V \in \mathbb{R}^{n \times k}$ is a dense matrix with n rows and k column, meaning each row of the V matrix represents a k dimensional dense feature vector.

$$X_{Cov} = V \Sigma^2 V^T \tag{7}$$

In order to tackle the large number of diagnose code in medical domains, a "max-pooling" procedure is applied to select the most responsive latent features from the latent embeddings, so each type of diagnose, such as "Primary Diagnose", is denoted by a k-dimensional vector. The performance comparison between latent logistic regression and binary logistic regression indicates that latent embedding improves the AUC score of readmission prediction from 0.779 to 0.790.

•Patent Embedding In addition to embedding each clinical object, research also proposes to represent a patient's entire raw electronic health record (EHR) as a single vector [104], or embed a single visit as a feature vector [85]. By using data from the University of California San Francisco (UCSF) from 2012 to 2016, and the University of Chicago Medicine (UCM) from 2009 to 2016, a research [104] develops a single data structure representing the EHR dataset in temporal order for each patient. All available data for each patient, from the beginning of a patient's record until the point of prediction, form the patient's personalized input to the model to predict unplanned 30-day readmission. A deep learning model is trained and compared with existing EHR and achieves high accuracy for tasks in predicting 30-day unplanned readmission (AUROC 0.75–0.76).

In order to learn efficient expression of medical concepts, Med2Vec [85] is proposed to learn healthcare concept at two levels: code-level and visit-level. When learning from code-level information, Skip-gram algorithm is employed to train a non-negative weight ReLU which guarantees sparse code representation production and provides much more convenience to interpret codes. A multi-layer perception network is trained to exploit sequential information of visits V_t, presented by a binary vector $\mathbf{x}_t$, and enables prediction of past and future visits. The binary vector can be converted into visit representation as shown in Eq. (8) where $\mathbf{W}$ is the code weight matrix and $\mathbf{b}$ is the bias vector.

$$\mathbf{u}_t = ReLU(\mathbf{W}_c \mathbf{x}_t + \mathbf{b}_c) \tag{8}$$

A visit vector is generated by concatenating visit demographics $\mathbf{d}_t$ with clinical feature vector in Eq. (9). As a result, a patient is represented by concatenating corresponding visit vectors. The performance of Med2Vec is evaluated using data from Children's Healthcare of Atlanta (CHOA), and shows significant improvement in prediction accuracy.

$$\mathbf{v}_t = ReLU(\mathbf{W}_c[\mathbf{u}_t, \mathbf{d}_t] + \mathbf{b}_v) \tag{9}$$

4.2 Domain Segmentation for Data Complexity

A hospital readmission visit is tied to many factors, including patients, diseases, care facilities, *etc.* As a result, it is often ineffective to rely on a one-for-all predictive model for accurate prediction. High readmission rates of hospitals at certain geographical locations might be because of (1) majority patients sharing similar demographics, (2) hospitals mainly practicing certain diseases, or (3) policy and management issues. To tackle such data complexity, one effective way is to rely on domain segmentation to segment readmission learning into multiple small tasks with better data coherence. The purpose is to use patient cohort with less variance (reduced data complexity) to train reliable models. Because readmission rates share rather distinct patterns with respect to different patient groups and diseases, a common approach is to use disease, gender, or race and ethnicity for segmentation.

4.2.1 Disease-Specific Prediction

Disease segmentation is the most common way to tackle the data complexity for readmission prediction. Intuitively, patients suffering from same/similar diseases are likely sharing similar symptoms, treatments, and post-discharge complications. Therefore, using disease to segment patient visits will minimize the data discrepancy for better prediction. Many models are proposed to predict readmission for leading cause of death diseases such as hear failure [94], stroke [71], COPD [28], AMI and pneumonia [84], *etc.*

In a research [94] to determine the risk of unplanned cardiovascular readmission for patients with chronic heart failure, a modified Cox's proportional hazards model (taking into account the competing risk of death) is used to develop a multivariate prediction model followed by boot-strap methods to identify data factors for the final model. Variables are chosen by a backward-deletion method with a p-value threshold for retention. Based on patients information selected from WHICH? trials (Which Heart Failure Intervention is most Cost-effective & consumer friendly in reducing Hospital care), the C-statistic of the final trained model reaches 0.80.

4.2.2 Gender-Specific Prediction

Despite the effectiveness of hospital readmission rate control policies and interventions targeted at the whole nation, a gender difference with regard to the rate of 30-day hospital readmissions has been pointed out and this observation triggered a knowledge exploration about gender different in readmission [148]. Compared with men, women are reported to have a higher readmission risk after being discharged [69], especially young women aged under 65 years old [144]. This indicates that policies and interventions aimed at preventing readmissions may need to consider both biological sex and gender in their designs and implementations. A psychiatric administrative dataset with women (n=33,353) and men (n=32,436) patients are used to identify predictors of 30-day readmission. Multi-variable logistic regression models with demographics, clinical information etc. are conducted and the results show female patients have 0.2% higher risk than male patients to be readmitted within 30 days and suggest that "Certain key predictors of psychiatric readmission differ by sex" [68].

4.2.3 Race and Ethnicity-Specific Prediction

Race and ethnicity are known factors of health disparity. When it comes to the implementation of HRRP in ethnics, a similar drop in the 30-day readmission rate of Myocardial Infarction patients in both Black and non-Black patients has been previously observed. Meanwhile, although the race-specific 30-day readmission rate differs, research also concerns that it might be attributed to patient-level factors such as incoming, age. In other words, the implementation of HRRP has nothing to do with the improvement or deterioration of racial differences in readmission rates and mortality [149]. In a mixed-effects logistic regression cohort study of 272,758 adults with diabetes, black patients are at a much higher risk of unscheduled readmission to the hospital within 30 days than other racial/ethnic groups with a 12.2% among black patients, 10.2% for white individuals, 10.9% among Hispanic patients and 9.9% for Asians [70]. The higher hospital readmission risk is also confirmed by a research based on 4,784 Blacks and 33,684 Whites with stroke admissions in year 2000 using a truncated negative binomial (TNB) model. The readmission risk among black patients age 65-74 is 40% higher than white patients [71].

4.2.4 All-Cause Prediction

In contrast to the above domain specific approaches, many predictive models still address the hospital readmission using an all-cause approach, which considers all disease types, gender groups, races *etc.* For example, a study compare different models used for readmission risk prediction of hospitalized primary care patients [34]. The study includes four classifiers including Walraven's LACE index, LACE+ index, Donze's HOSPITAL score and logistic regression based classifiers, to predict 30-day readmission risk, using a dataset at Mayo Clinic Department of Family Medicine. The results show that logistic regression based classifiers yield only moderate performance in predicting readmission with a C-statistic of 0.666 compared with LACE (0.680), LACE+ (0.662) and HOSPITAL (0.675) [34].

5 MODEL INTERPRETABILITY

Model interpretability defines the degree to which the behavior of a model can be explained and understood by human perception. For medical applications, an interpretable and transparent model is almost always preferred. Only interpretable predictions allow care providers to understand the decision patterns to support clinical actions. Therefore, the first model level challenge is to design clinically interpretable and practically usefulness models to provide actionable insights to the decision makers.

In Fig. 4, we outline general relationships between model interpretability *vs.* model accuracy of commonly-used machine learning algorithms. The x-axis denotes the model interpretability and the y-axis denotes the prediction accuracy. In summary, models like decision trees, linear/logistic regressions, case based reasoning *etc.* are relatively easy to interpret, mainly because that the model itself outputs parameters associated to some features or samples. For example, a decision tree will specify features (and their values) used to derive decision, and logistic regress will specify the weight of features for prediction. Case based

reasoning finds samples (or prototypes) similar to the query instance to derive decision. As models are moving towards relying on general computing machines (such as support vector machines, neural networks, Bayesian networks) or ensemble models, the interpretability will decrease, mainly because that the models do not directly specify how features/samples are used to form decisions.
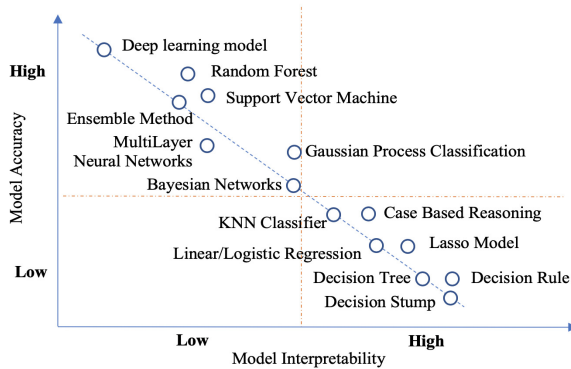

Fig. 4: An inverse relationship between model interpretability *vs.* model accuracy for common machine learning algorithms. A model is the output of the referred classifiers (or learning algorithms). In general, a model with a higher accuracy tends to have a lower interpretability, whereas models easier to interpreter tend to be less accurate.

5.1 Model Interpretability Enhancement

5.1.1 Simple Transparent Models

Simple transparent models are the ones with strong interpretability, and the models can directly inform decision actions. Decision trees/rules and linear/logistic regressions, in Fig. 4, are the ones falling into this category and frequently being used in the medical domains.

In a recent study, researchers from MIT [67] propose to build interpretable predictive models for inpatient flow prediction, which predicts flow of hospitalized patients into two major categories: flows out of the hospital, *i.e.*, discharges, and flows between units of the hospital. A collection of machine learning techniques, including linear models and decision trees, are applied to address four length of stay-related tasks: same-day and next-day discharges, and more-than-7 and more-than-14-day hospital stays. Their study shows that the prediction accuracy and scalability cannot be hindered by attentions on modeling and interpretability. Instead, it brings more interpretable functions and thus clinicians and care providers are more involved, less data as well as computing resources is required.

5.1.2 Hybrid Transparent Models

Machine learning algorithms with high accuracy, such as neural networks and support vector machines, often have very little interpretability [33], imposing significant obstacles for clinical decisions and model implementation across health systems [116]. There is a trade-off between performance and interpretability: complexity models are untraceable black boxes, while classic interpretable models are usually simplified with lower accuracy. This trade-off narrows

the application of the latest machine learning models in hospital readmission prediction area, which requires both high predictive performance and an understanding of the contribution of each attribute to the model results.

In order to deliver an interpretable system with higher accuracy, hybrid transparent models combine different types of predictive models to ensure high interpretability and accuracy. A mixture of two types of classifiers, a Boosted C5.0 tree and a support vector machines (SVM), is proposed [88] for hospital readmission prediction. The main objective is to take the advantage of high sensitivity of SVM prediction, as well as the transparency of C5.0 tree model. In their algorithm, two separated routines (optimizing SVM and C5.0 independently, and optimizing mixed ensemble) are carried out in order to determine optimal parameters for SVM and C5.0. A unique feature of the mixed ensemble is that the model delivers a hospital readmission prediction with balanced accuracy and reasonable transparency.

5.2 Feature Interpretability Enhancement

To enhance model interpretability, another effective way is to select/create effective features. By using a small set of important features, the model can inform factors playing important roles for readmission prediction.

5.2.1 Feature Selection

Feature selection selects or extracts significant variables from a set of given features to explain model decisions. Such approaches can ensure that the variables included in the final model have clinical significance, can be identified and understood, and can lead to new insights and hypotheses. Most importantly, interpretable machine learning supports clinicians and patients' decision-making by clearly indicating the nature and characteristics of the most important variables for algorithms to make a prediction.

SHAP (Shapley Additive exPlanations) value [150] is a feature selection based approach to predict personalized patients during and after hospitalization. The interactions between variables are examined by SHAP, followed by Gradient Boosting Machine (GBM) being used to train predictive models based on selected features. The most significant features include primary diagnosis, length of stay, and so on. The final predictive performance is: 30-day readmission (AUC 0.76/BSL 0.11); LOS $\geq$ 5 days (AUC 0.84/BSL 0.15); death within 48–72h (AUC 0.91/BSL 0.001). Similarly, information-based (Gini indexing) and frequency-based feature reduction methods are used to improve the understandability of determining readmissions based on a number of variables. Their research shows that prescription drugs, diagnostics, and information about operations during admission can help predict the admission rate [72].

5.2.2 Feature Learning

Feature learning intends to transform complex, redundant, and high dimensional features into a new representation, such as a dense feature vector, that can be effectively exploited in machine learning tasks. The strength of feature learning is that it extracts useful features or representations from raw data, and also learns the predictive task at the

same time. Feature learning can be divided into two categories: supervised feature learning in which features are learned with labeled input data and unsupervised feature learning, which learns features from unlabeled input data.

For most feature learning methods, the output features are dense vectors suitable for machine learning, but difficult for human understanding. A recent study [105] proposes to embed 971 features into 100 sparse dimensions through a k-sparse auto-encoding using health data records from Yale New Haven Health system. k-sparse auto-encoder is a sparse embedding function which carries the most vital information for each patient and considers the embedded vectors as a concentration of the original features. Given a dataset $\mathcal{D} = [X_1, X_2, \cdots, X_n] \in \mathbb{R}^{n \times m}$ with n instances and m original features, and each instance $X_i \in \mathbb{R}^m$ is denoted by an m dimensional features, the classical autoencoder objective function is defined as Eq. (10). ϕ denotes encoder transition converting X_i to a new k dimensional feature space $\widetilde{X}_i \in \mathbb{R}^k$, and ψ is the decoder transition converting $\widetilde{X}_i$ back to the original input space. $(\phi \circ \psi)X_i$ denotes the composite function, encoding and decoding to X_i.

$$\phi : X \to \widetilde{X} \qquad \psi : \widetilde{X} \to X$$
$$\underset{\phi, \psi}{\arg\min} \frac{1}{|\mathcal{D}|} \sum_{i=1}^{n} \|X_i - (\phi \circ \psi)X_i\|_2^2 \qquad (10)$$

Then a k-sparse activation constraint in Eq. (11) is used to penalize the deviation between observed average activation value of neurons $\rho_{h,\mathcal{D}}$ from pre-defined average activation value $\rho_{h,\mathcal{D}}^*$ on all instances in $\mathcal{D}$.

$$\underset{\phi, \psi}{\arg\min} \sum_{h \in \phi, \psi} \|max(0, \rho_{h,\mathcal{D}} - \rho_{h,\mathcal{D}}^*)\|^2 \qquad (11)$$

Finally, a secondary constraint is applied to limit each dimension of the sparse embedding to either 1 or 0. This binarization forces the sparse embedding to carry the most significant information for each patient.

$$\underset{\phi, \psi}{\arg\min} \frac{1}{|\mathcal{D}|} \sum_{i=1}^{n} \sum_{h=1}^{H} (\widetilde{X}_{i,h} \times (1 - \widetilde{X}_{i,h})) \qquad (12)$$

Based on the above design, a hierarchical structure in an interpretable framework is embedded with the sparse vectors. The patient data distribution shows that sparse embedding can make the hidden phenotype in the admission cohort more prominent. Therefore, the method shows capability of maintaining the integrity of important information.

5.3 Feature-Model Interpretability Enhancement

Medical data often have high dimensions, resulting in low learning efficiency, high complexity models with poor interpretability [33]. To enhance the interpretability, feature and model combined enhancement approaches integrate feature learning and model training together to learn features easy to interpret and models with a better prediction accuracy.

5.3.1 Feature Regularization Models

Sparsity regularization is commonly used to regularize the feature space by using different weight values for the predictive models to leverage feature weights for prediction.
•Sparsity Regularization Sparsity regularization uses L_1 norm normalization to enforce the sparsity, so majority

features have small weight values except that a few important features have relatively large values. In the research aiming to identify unplanned readmission risks in a high dimensionality electronic health dataset, a tree-lasso logistic regression [33] is proposed to integrate ICD-9-CM hierarchy for pediatric hospital readmission. The learning objective is to minimize a penalized likelihood defined in Eq. (13).

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} (1 + \exp(-y_i(\mathbf{x}_i^T \theta + c))) + \lambda P(\theta) \qquad (13)$$

In Eq. (13), θ is the model parameter (weight values of features), $\mathbf{x}_i$ defines feature vector of a patient (such as ICD-9-CM codes of each visit), and $y_i \in \{-1, 1\}$ is the label of $\mathbf{x}_i$. c and λ are predefined parameters. $P(\theta)$ defines the regularization term, and for Lasso regularization it is defined as $P(\theta) = \sum_{j=1}^{m} |\theta_j|$ which encourages to select "sparse" features. In reality, Lasso regularization still selects a rather large number of features hard to interpret, so Tree-Lasso regularization uses another regularization term defined in Eq. (14). The key idea is to utilize the ICD-9-CM hierarchy of diagnoses, which represents groups of diagnostic does. For example, code between 001 and 139 represent infectious and parasitic diseases, and 001 to 009 further represent intestinal infectious diseases.

$$P(\theta) = \sum_{j,k} w_j^k \|\mathbf{x}_{G_k}\|_2 \qquad (14)$$

In Eq. (13), G_k defines the group of features selected by the kth node in the hierarchy. $\| \cdot \|_2$ is the L_2 norm, and w_j^k is the weight assigned to each node.

The above model was evaluated using 66,000 pediatric hospital records, in terms of model interpretability and selecting features on different hierarchical levels of disease classes. Comparing to traditional Lasso logistic, the integration of ICD-9-CM hierarchy diseases provides better model interpretability and AUC values (0.779) with less information than traditional Lasso model.

Domain hierarchical structure of the diseases code (*e.g.* ICD-9-CM hierarchy) is also used in another tree based sparsity-inducing regularization [110] which exploits domain induced hierarchical structure for disease codes to improve the comprehension of hospital readmission topic. In their study, four regularizaition methods are compared in respect to logistic regression model. Sparse Group Regularizer (SGL) and Tree Structures Group Regularizer (TSGL) outperform L_1 and L_2 norm regularization. SGL assumes the input features can be set into k groups while TSGL applies the hierarchical structure presented on the features. Also, the TSGL method is proved with better sparsity at higher levels of the hierarchy.
•Attention Regularization Attention is a neural network mechanism to assign weight values to features/instances, by taking their correlation into consideration. An attention based neural network model (MLP_attention) is proposed to generate interpretable prediction results to determine the contribution of each feature. The final input feature representation is calculated by element-wise multiplication of input feature vector X and the attention signal α generated by a fully connected layer [86].

A two-level neural attention model [151] proposes to consider hospital visits in a reverse time order so that recent

clinical visits receive higher attention weights. The model, validated on 14 million visits, delivers better accuracy and interpretability compared to generic deep learning methods. It can also detect influential past visits and significant clinical variables within those visits (e.g. key diagnoses).

5.3.2 Feature Topic Models

Topic modeling is a machine learning approach that explores high level semantics from text documents. In a general setting, a set of documents are used to find clusters of words with similar semantics, which are interpreted as topics. Intuitively, feature vectors representing patients' in-hospital treatment can be very sparse and the relationships between features with similar meanings are also ignored. Topic modelling is one effective way to solve this challenge, by assigning features to different topics. This will essentially enhance the interpretability for both model and features, because users can understand which features, such as ICD codes, are tied to certain disease (topics).

Latent Semantic Analysis (LSA), Hierarchical Dirichlet Process (HDP), and Latent Dirichlet Allocation (LDA) are commonly used topic models, but they normally cannot take label information into consideration to learn topics. In a research to predict post-ICU mortality [81], labeled-LDA (Latent Dirichlet Allocation) [152] is used to incorporate domain knowledge into topic modeling, by using ICD-9-CM codes as labels (domain knowledge). Given a dataset with D documents and N unique words ($\omega_1, \omega_2, \cdots, \omega_N$), a list of binary topic presence/absence indicators $\Lambda^{(d)} = (l_1, ..., l_K)$ are used to denote topic(s) the current document d is associated to. The learning of labeled-LDA [152] is to train a generative model in Fig. 5. Unlike traditional LDA model in which a multinomial mixture distribution $\theta^{(d)}$ is drawn over a K topics for each document d from a Dirichlet prior $\alpha = (\alpha_1, \cdots, \alpha_K)$, in Labeled-LDA, the multinomial mixture distribution $\theta^{(d)}$ is restricted only over topics that equivalent to its labels $\Lambda^{(d)}$ as shown in Fig. 5.
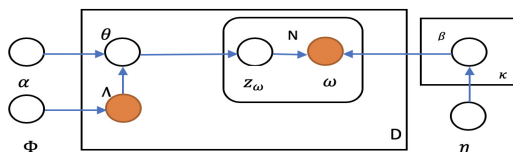

Fig. 5: Labeled-LDA topic model. Colored nodes are observable variables, where ω is a list of word indices (there are N words in total), and Λ is a list of binary topic presence/absence indicators. α is the parameter of the Dirichlet topic prior; ϕ is the labeling prior probability; θ is the mixture distribution over all topics; z_ω is the word-topic assignments; β is the multi-nomial topic distributions over vocabulary; η is the parameters of the word prior.

Table 7 lists examples of ICD-9-CM codes, their definitions along with the corresponding keywords learned from Labeled-LDA. For example, "Anemia" topic includes words "tablet", "mg", "blood", "po", *etc.* There are two advantages of using the ICD-9-CM code as a label in Labeled-LDA. First, the clinical notes from a given patient's record only contribute to a subset of topics corresponding to that patient's ICD-9-CM code assignment. Secondly, by combining the ICD-9-CM code definition and the keywords of a given theme, the interpretability of the theme can be realized.

TABLE 7: Examples of feature topic models using Labeled-LDA to find keywords associated to ICD-9-CM codes. "Definition" shows diseases, injuries, symptoms and conditions defined by ICD-9-CM codes. "Keywords" are the words learned to correspond to the ICD-9-CM.

ICD-9-CM	Definition	Keywords
280-285	Anemia	pt tablet mg blood po hct sig daily discharge pm doctor namepattern md patient pain day history gi admission hospital
420-429	Other forms of heart disease	pt mg patient hr chest resp left lasix gi po stable pain gu neuro gtt bp bs day cv plan
270-279	Other metabolic and immunity disorders	patient mg pt chest day left artery pain po stable coronary cabg status discharge history date post namepattern clip examination
317-319	Mental retardation	pt tube noted chest cc resp patient retardation thick secretions care cont plan trach abd hr ct neuro telemetry coarse

By using textual information, such as medical notes and other structured data from MIMIC II database, the model is validated to achieve transparent outcome prediction and high AUC scores (0.835 and 0.829) for 30-days and 6-months readmission prediction, respectively [81].

5.3.3 Feature Interaction Models

Using interactions between variables can interpret models by explaining factors playing important roles in the decision. Accordingly, generalized additive models with pairwise interactions (GA^{2}Ms) is proposed using standard Generalized additive models (GAMs), which usually model the dependent variable as a sum of univariate models [153]. Pairwise interactions is added into GAMs in order to improve model accuracy and interpretability. Eq.(15) shows how GA^{2}Ms works: $x_i = (x_{i1}, ..., x_{ip})$ is the feature vector with p features and y_i is the target. g is the link function, for each item f_j, $E[f_j] = 0$. For GA^{2}Ms, it builds the best GAM at first and then all possible pairs of interactions in the residuals are detected and ranked. It is tested on a readmission prediction for pneumonia case study and GA^{2}M models present an excellent readmission prediction accuracy while at the meantime, the interpretability of the decisions is improved. GA^{2}Ms is proved to be useful in accurately explaining the prediction decisions made by model for each patient focusing on the most relevant terms for each patient [154].

$$g(E[y]) = \beta_0 + \sum_j f_j(x_j) + \sum_{i \neq j} f_{ij}(x_i, x_j) \qquad (15)$$

Using deep neural networks, INPREM [155] is recently proposed to enhance deep learning interpretability for interpretable and trustworthy healthcare prediction. To ensure model interpretability, INPREM employs a $CM \in \mathbb{R}^{T \times |C| \times l}$ matrix to characterize contributions of different factors, where T denotes the number of hospital visits, $|C|$ denotes number of medical/procedure codes, and l denotes the number of classes. $CM[i,j][k]$ captures the contribution of the jth medical event in the ith visit, with respect to the class k. By using deep neural networks to learn weight matrices

$\mathbf{w}_v \in \mathbb{R}^{g\times|C|}$ and $\mathbf{w}_c \in \mathbb{R}^{g\times l}$ where g is the dimension of the embedding space. It also uses $\alpha \in \mathbb{R}^{1\times T}$ to capture non-linear dependencies between visits, and uses $\beta \in \mathbb{R}^{g\times T}$ to encode non-linear dependencies between medical events within each visit, INPREM calculates CM matrix using Eq. (16), where $\odot$ denotes element-wise multiplication.

$$CM[i,j] = \mathbf{W}_c^T(\alpha[i]\beta[:,i] \odot \mathbf{W}_v[:,j]) \qquad (16)$$

6 MODEL IMPLEMENTATION CONFLICT

HRRP is evidently effective in encouraging hospitals, especially the ones with high readmission rates, to reduce readmission rates [123]. Ideally, it is expected that the program will motivate hospitals to improve treatments, enforce discharge evaluation and follow-up procedures, and eventually lead to medical cost reduction and improve the patient satisfaction and qualify of services [156]. In reality, however, the implementation of the models is challenged by many conflicts, partially because predictive models create a layer of separation between patients and care providers.

6.1 Quality Contradiction

6.1.1 Controversial Measurement

Concerns about the HRRP program, as a healthcare service quality measurement, have been rising due to the complexity of the healthcare ecosystem. One one hand, for many patients, even if they do receive high-quality care, there still exists possibility the patients will return to the hospital due to the nature of disease development. An inverse relationship between a hospital's readmission rate and its mortality rate has been observed, and shows that low-mortality hospitals tend to have higher readmission rates [157]. On the other hand, prediction models usually fail to include data elements that have perverse incentives such as prior utilization. In order to determine whether the healthcare expenses of patients with a 30-day readmission differ from those without a 30-day readmission, a retrospective study is designed to measure the consequences. After comparing the 12 month genuine costs of two groups of patients: patients with a 30-day readmission *vs.* patients without 30-day readmission, the study show that readmitted patients are often "sicker" than non-admitted patients. Therefore, using hospital readmission model for profiling hospitals may "systematically underestimate performance of hospitals with high rates of observed readmissions", and jeopardy the true "saving" to the healthcare system [158].

Another research [159] also concludes that certain patient groups, such as cardiac arrest, have high readmission rates and inpatient costs. A two-part model, including a logistic model and a gamma regression model, is constructed to determine adjusted costs and cost ratios based on a cohort from GWTG-Resuscitation submitted by 523 acute-care hospitals with a total number of 19,373 patients. The overall amount of 30-day readmission is 2005 in which cardiovascular disease contributes to 35.9% followed by pulmonary disease (17.1%). For inpatient resource usages, the mean cost of in-hospital cardiac arrest is $35,808 \pm $38,230, much higher than the average cost $7,741 \pm $2,323 of the whole cohort.

6.1.2 Unintended Consequence

After the HRRP initiative, the readmission rate has shown an annually downward trend. Yet, the true factors and the consequence of this reduction have not been clearly understood so far. One possible reason is that hospitals choose to avoid readmissions by raising the threshold to readjust patients recently discharged from the hospital. A recent research [160] studies the HRRP policy on the emergency (ER) visit, and observes that HRRP is associated with a decreasing likelihood "that hospitals would readmit recently discharged patients returning to the ED within 30 days of discharge". Such observations raise concerns about unintended consequence of the HRRP program, and its potential long term impact to both hospitals and patients.

6.2 Cost Contradiction

6.2.1 Emotional Cost Implication

Being readmitted into hospital after a recent discharge is a complicated experience for many patients. Researchers find that patients often harbor different opinions towards their own readmission from healthcare providers [161]. The research, from 178 interviews of readmitted patients compared with the perspective of the admission provider, concludes that 58% patients believe that system issues, controlled by the discharge process, are the contributors to their readmission. Meanwhile, patients concern that their readmission could have been avoided if the system were modified. In NSW Australia, in order to explore the readmission experience of aged population, three elderly patients discharged from a large tertiary referral hospital are interviewed. Their feedback suggests that nurses are more proactive in people-centered care of older patients to alleviate the patient being left out and let down [162].

In order to study relationship between patient experience and predictive modeling of readmission, a research [163] uses patient-level Hospital Consumer Assessment of Healthcare Providers and Systems (HCHAPS) and Press Ganey data to understand whether a patient being readmitted is a driver of poor experience scores (reverse causation). The study uses multivariable logistic regression to analyze how HCHAPS and Press Ganey data 30-readmission status are correlated. The results point out that "poor patient experience may be due to being readmitted, rather than being predictive of readmission". In other words, predictive models do not directly lead to poor patient experience [163].

6.2.2 Financial Cost Implication

Despite of the cost reduction motivation of the HRRP initiative, the total costs to hospitals and patients are contracting. At very minimum level, the implementation of predictive models in hospitals, prior to the discharge of patients for preventable readmission, is time consuming and expensive. The long term accumulated costs can even be more profound. A study using hierarchical regression analyses to examine the cumulative costs of short-term readmissions after percutaneous coronary intervention (PCI) shows that 30 days of readmission and cumulative expenses are positively correlated [164]. To reduce readmission, a study suggests hospitals to consider providing better medication reconciliation, giving instruction to both patients and their caregivers

on the follow-up care and also implementing post-discharge care coordination [116]. The post-discharge education and follow-up intervention include complimentary home visit within 48 hours after a patient is discharged to review the education provided before discharge and a minimum of four nursing telephone meetings. Despite of the fact that the intervention [165] on a patient after discharge is effective in reducing readmission rate, it imposes significant extra costs to medical staff and healthcare systems.

A research [166] investigates 8 million Medicare beneficiary hospitalizations from 2005 to 2015, and finds that the implementation of the HRRP is associated with a significant increase in 30-day postdischarge mortality among beneficiaries hospitalized for heart failure and pneumonia. In other words, there is an association that 30-day readmission policy may increase the mortality for patients beyond the 30 day window. On the other hand, another study [167] shows that transitional care interventions, such as home visits by nurses, can reduce death rates and hospital readmissions by more than 30%. As a result, they suggest that transitional care services should become the standard of care for post-discharge management of patients with heart failure.

6.3 Intervention Contradiction

6.3.1 Patient Willingness

The implementation of predictive models implies an intervention process which may contradict to patients' willingness. For example, a model may predict a patient as high readmission risk and requires further hospitalization, but the patient feels fine and is unwilling to cope. Discharge against medical advice (AMA) means a patient chooses to leave the hospital earlier than the discharge recommendation. Leaving the hospital without following the doctor's advice may put the patient at risk of insufficient medical care and result in the need for readmission. A retrospective cohort analysis based on 2014 National Readmission Database shows that patients discharged against medical advice (AMA) are of larger possibility of 30-day readmission compared with non-AMA individuals, which has caused huge loss to the healthcare system and higher hospital mortality. Patients discharged from the AMA are also more likely to have an earlier rebound and readmission, which may reflect their dissatisfaction with the initial care [168].

6.3.2 Patient Privacy

As EMRs are becoming common, the automatic collection, use and storage of patient medical information have been facilitated, allowing studies on hospital readmission and stimulating care providers for automated decision making. However, the applications of large amounts of medical data also raise public concerns about patients' medical information leakage. Although HIPAA and other institutions prohibit researchers to use identified patient-level data and patients' personal information are protected properly by this policy [126], [169], indeed, there are always risks of information breach. Developers need to find a balance between collecting sufficient representative patients' medical data and minimizing the risk of leaking patients' privacy. A governance structure that includes patients and other stakeholders early in the development of a model is recommended to be implemented by model developers at the earliest stage of development [125].

7 DATASETS AND RESOURCES

In this section, we summarize a list of datasets publicly available for hospital readmission prediction in Table 9.

7.1 All Cause Readmission Datasets

All cause datasets contain information about all hospitalized patients of typical diseases within a specific period of time.

Among all data sources, the Healthcare Cost and Utilization Project (HCUP) contributes significantly to the nationwide databases in healthcare readmission. HCUP is developed as a combination of healthcare databases, software tools, and products through federal-state-industry partnerships and are jointly developed by healthcare research and the Agency for Healthcare Research and Quality (AHRQ). Table 8 lists the National Inpatient Dataset (NIS), State Inpatient Databases (SID), and Nationwide Readmissions Database (NRD), which all belong to the HCUP databases.

The HCUP database brings together the data collection efforts of state data organizations, hospital associations, private data organizations, and the federal government to create a national information resource for encounter-level medical data as shown in Table 8. HCUP includes the largest collection of longitudinal hospital care data in the United States, starting from 1988 with full payment, encounter level information. These databases can be used to study a wide range of health policy issues, including the cost and quality of medical services, medical practices at the national, state and local market levels, access to health care plans, and treatment outcomes.

TABLE 8: HCUP Databases

Dataset	Description	#of records
National (Nationwide) Inpatient Sample (NIS) https://www.hcup-us.ahrq.gov/db/nation/nis/nisdbdocumentation.jsp	U.S. regional and national estimates of inpatient utilization, access, charges, quality, and outcomes	7 million each year
Kids' Inpatient Database (KID) https://www.hcup-us.ahrq.gov/db/nation/kid/kiddbdocumentation.jsp	The largest publicly-available all-payer pediatric inpatient care database in the US	3 million each year
Nationwide Ambulatory Surgery Sample (NASS) https://www.hcup-us.ahrq.gov/db/nation/nass/nassdbdocumentation.jsp	The largest all-payer ambulatory surgery database in the United States	9.9 million each year
Nationwide Emergency Department Sample (NEDS) https://www.hcup-us.ahrq.gov/db/nation/neds/nedsdbdocumentation.jsp	The largest all-payer emergency department (ED) database in the United States	30 million each year
Nationwide Readmissions Database (NRD) https://www.hcup-us.ahrq.gov/db/nation/nrd/nrddbdocumentation.jsp	Support various types of analyses of national readmission rates for all patients	18 million each year
State Inpatient Databases (SID) https://www.hcup-us.ahrq.gov/db/state/siddbdocumentation.jsp	Inpatient discharge records from community hospitals in that State	State-specific
State Ambulatory Surgery and Services Databases (SASD) https://www.hcup-us.ahrq.gov/db/state/sasddbdocumentation.jsp	Encounter-level data for ambulatory surgeries and may also include various types of outpatient services	State-specific
State Emergency Department Databases (SEDD) https://www.hcup-us.ahrq.gov/db/state/sedddbdocumentation.jsp	Emergency visits at hospital-affiliated emergency dept. (EDs) not resulting in hospitalization	State-specific

TABLE 9: Public datasets used for hospital readmission prediction (Demographics (Demo.), Hospital Information (Hosp.), Clinical Information (CL), Healthcare utilization (HC)

Dataset & Paper	Type	Population	#of patients	#of admin.	Demo.	Hosp.	CL.	HC.
Medicare Provider Analysis & Review (MedPAR) [6], [73]	All cause	Medicare	10E5	NA	✓	✓	✓	✓
Medicare Current Beneficiary Survey (MCBS) [158]	All cause	Medicare	10E5	NA	✓	✓	✓	✓
Health Care's Enterprise Data Warehouse (EDW) [29]	All cause	18+	10E5	NA	✓	✓	✓	✓
Cerner's Millennium® EHR software system [29]	All cause	NA	10E5	NA	✗	✗	✓	✗
New Zealand National Minimum Dataset [30], [87]	All cause	All patients	10E5	NA	✗	✗	✓	✗
National Inpatient Dataset (NIS) [32]	All cause	All patients	10E5	NA	✓	✓	✓	✓
State Inpatient DB (SID) [33], [50], [68], [75], [76], [145]	All cause	All patients	10E5	NA	✓	✗	✓	✓
Nationwide Readmissions Database (NRD) [84]	All cause	All patients	10E5	NA	✓	✓	✓	✓
Resource and Patient Management System (RPMS) [45]	All cause	All patients	10E5	NA	✓	✗	✓	✓
Nationally Rep. Health and Retirement Study (HRS) [47]	All cause	50+	20000	NA	✓	✗	✓	✗
Queensland Hosp. Admit. Patient Data (QHAPDC) [62]	All cause	All patients	10E5	NA	✓	✗	✓	✗
MIMIC III database [92]	All cause	16+	38,597	49,785	✓	✗	✓	✓
MIMIC II database [81]	All cause	16+	26,000	NA	✓	✗	✓	✓
Cerner Health Facts database [100]	All cause	All patients	17,880,231	74,036,643	✓	✓	✓	✓
Natl. Surgical Quality Imp. Program [42], [58], [106]	Surgery	All patients	NA	NA	✓	✓	✓	✓
PREVENT III database [60]	Limb ischemia	18+	1404	24.4%	✓	✗	✓	✓
WHICH? Trial [94]	Heart Failure	18+	280	13%	✓	✗	✓	✓
UCI Machine Learning Repository"Diabetes-130US" [9]	Diabetes	All patients	NA	69,984	✓	✗	✓	✓

7.2 Specialized Readmission Datasets

Specialized readmission datasets focus on certain type of diseases, such as diabetes or heart disease, special medical conditions, or a certain type of medical procedures, such as surgery procedure intervention.

The National Surgical Quality Improvement Program (NSQIP) is a surgical result database of the American College of Surgeons (ACS), which aims to measure the results of risk-adjusted surgical interventions in order to compare results between hospitals. It is based on 135 variables collected before surgery to 30 days after surgery such as demographics, surgical profile and infection at the surgical site. Through this risk adjustment completed by logistic regression model, the results of hospitals of different sizes that serve different patient populations can be fairly compared. ACS NSQIP data enhances the hospital's ability to zero preventable complications. Because it was developed by surgeons understanding the actual conditions in the operating room, ACS NSQIP can help hundreds of hospitals across the country evaluate the quality of their surgical plans with unparalleled accuracy and significantly improve the surgical results.

7.3 Data Fields Related to Readmission

7.3.1 Dataset Statistics

Table 9 includes simple statistics and brief features the dataset. The type of patient indicates whether the dataset is collected from all causes or from specific cohorts. Population specifies the patient groups, and number of patients/admissions specify the dataset size. Because some patients may return to hospital (readmission) for multiple times, the number of patients and number of readmission often vary for each dataset.

7.3.2 Patient Information

The check boxes in Table 9 list whether a dataset has patient information includes demographics (Demo), patient clinical information (CL) and healthcare utilization (HC). Patients age, gender, insurance information, marital status, education, medical history, income etc, are considered as demographics. Clinical information including patients' in-hospital information such as the treatment he/she received, lab values are categorized into patient information. The last part of patient information is the healthcare utilization, which consists of patients' length of stay in the hospital, admission and discharge date, medical notes etc.

7.3.3 Hospital Information

Hospital information (Hosp) specifies whether the dataset has information about the hospitals to which patients were admitted. In most cases, such hospital information includes hospital teaching status, location (rural/urban), hospital ownership (private/public), hospital bed sizes, and the annually hospital total discharges *etc*.

8 DISCUSSION

8.1 Summary of Data & Model Challenge Solutions

Data and model challenges represent the two most critical factors for predictive modeling of hospital readmission. In Table 10, we outline strength and weakness of solutions proposed to address these two types of challenges. Indeed, no single approach can perfectly address both challenges, and an effective solution should be customized based on the real-world domain settings.

8.2 Future Research

This survey opens many opportunities for future study. First, our taxonomy organizes challenges into two main categories (data challenges *vs.* model challenges). Health and medical domains typically rely on specific measures, in addition to common measures, such as accuracy, AUC value etc. Therefore, other challenges, such as performance metrics, may also be considered for future research. Second, our survey emphasizes on academic publications, yet many commercial systems are available and they are focused on the system and engineering aspects of the problem. Third, we are focused on the English literature, and inherently overlooked high-quality publications from non-English venues.

TABLE 10: Summary of strength vs. weakness of methods addressing data & model challenges

Challenges	Solutions	Strengths	Weaknesses
Data challenges	Sampling approaches	low cost, high understandability	low representativeness
	Cost-sensitive learning	can integrate medical costs in modeling	time-consuming finding accurate loss function
	Ensemble learning	better performance than single classifier	time-consuming with large medical record
	Federated learning	better patient privacy protection	expensive communication
	Singular features	high representative features can be used	sparse, may ignore some informative features
	Hybrid features		
	Embedding features	dense features with low dimensionality	hard to understand feature meaning
	Disease specific prediction	provide better predictions for targeted patients	not applicable to different disease types prediction
	Gender-specific prediction		
	Race, ethnicity-specific prediction		
	All-cause prediction	rich data from all disease types	less accurate than domain-specific models
Model challenges	Simple transparent models	simple, easy to understand	less accurate
	Hybrid transparent models	high interpretability	large computational requirement
	Feature selection	important, indicative clinical features can be selected	do not guarantee improved performance
	Feature learning		
	Feature regularization models	fair weight values distribution	can not select groups of correlated features
	Feature topic models		

9 CONCLUSIONS

In this paper, we provided a comprehensive review of predictive models for hospital readmission. The survey first proposes a taxonomy to summarize challenges into four main categories: (1) data imbalance, locality and privacy; (2) data variety and complexity; (3) model interpretability; and (4) model implementation. We further organized challenges into subgroups in which the main problems of these challenges, the hindrance to hospital readmission prediction research, and improvement or solutions are discussed. Popular predictive model types, according to methods used, and public available datasets for hospital readmission prediction are summarized to provide materials for creative modeling approaches. The survey, including summary and analysis, not only provides a thorough understanding of existing challenges and methods in this field, but also lists available resources to advance the research for accurate hospital readmission prediction and modelling.

REFERENCES

[1] Hospital readmissions. [Online]. Available: https://www.americashealthrankings.org/explore/senior/measure/hospital_readmissions_sr/state/ALL

[2] H. Felix, B. Seaberg, Z. Bursac et al., "Why do patients keep coming back? Results of a readmitted patient survey." Soc Work Health Care, vol. 54, no. 1, pp. 1–15, 2016.

[3] A. Forster, H. Murff, J. Peterson et al., "The incidence and severity of adverse events affecting patients after discharge from the hospital." Ann Intern Med, vol. 138, no. 3, p. 161, 2003.

[4] A. Forster et al., "Adverse events among medical patients after discharge from hospital." CMAJ, vol. 170, no. 3, p. 345, 2004.

[5] L. Wier, M. Barrett, C. Steiner, and H.Jiang, All-Cause Readmissions by Payer and Age, 2008: Statistical Brief #115. Agency for Healthcare Research and Quality (US), Rockville (MD), 2006.

[6] T. Tsai, E. Orav, and K. Joynt, "Disparities in Surgical 30-Day Readmission Rates for Medicare Beneficiaries by Race and Site of Care." Annals of Surgery, vol. 259, no. 6, p. 1086–1090, 2014.

[7] M. Bailey et al. Characteristics of 30-day all-cause hospital readmissions, 2010-2016. [Online]. Available: https://www.ahrq.gov/sites/default/files/wysiwyg/data/infographics/readmission-rates.pdf

[8] S. Damery and G. Combes, "Evaluating the predictive strength of the LACE index in identifying patients at high risk of hospital readmission following an inpatient episode: a retrospective cohort study." BMJ Open, vol. 7, no. 7, p. e016921, 2017.

[9] K. Hempstalk and D. Mordaunt, "Improving 30-day readmission risk predictions using machine learning," Health Informatics New Zealand (HiNZ) Conference 2016, 2016.

[10] J. Donze, M. Williams, E. Robinson et al., "International validity of the hospital score to predict 30-day potentially avoidable hospital readmissions," JAMA Internal Medicine, vol. 176, 2016.

[11] L. Low, N. Liu, M. Ong et al., "Performance of the lace index to identify elderly patients at high risk for hospital readmission in singapore," Medicine, vol. 96, no. 19, p. e6728, 2017.

[12] M. Hakim et al., "Performance of the lace index to predict 30-day hospital readmissions in patients with chronic obstructive pulmonary disease," Clin. Epidemiol., vol. 10, pp. 51–59, 2017.

[13] J. Donze, D. Aujesky, D. Williams, and J. Schnipper, "Potentially Avoidable 30-Day Hospital Readmissions in Medical Patients Derivation and Validation of a Prediction Model." JAMA Intern Med, vol. 173, no. 8, pp. 632–638, 2013.

[14] S. Sushmita, G. Khulbe, A. Hasan et al., "Predicting 30-day risk and cost of "all-cause" hospital readmissions," AAAI Workshop: Expanding the Boundaries of Health Informatics Using AI, 2016.

[15] M. Jamei, A. Nisnevich, E. Wetchler, S. Sudat, and E. Liu, "Predicting all-cause risk of 30-day hospital readmission using artificial neural networks," PLOS ONE, vol. 12, p. e0181173, 2017.

[16] S. Yu, F. Farooq, A. Esbroeck, G. Fung, V. Anand, and B. Krishnapuram, "Predicting readmission risk with institution-specific prediction models." Artif Intell Med, vol. 65, no. 2, pp. 89–96, 2015.

[17] P. Cotter, V. Bhalla, S. Wallis, and R. Biram, "Predicting readmissions: Poor performance of the lace index in an older uk population," Age and ageing, vol. 41, pp. 784–9, 2012.

[18] X. Min, B. Yu, and F. Wang, "Predictive modeling of the hospital readmission risk from patients' claims data using machine learning: A case study on copd," Scientific Reports, vol. 9, p. 2362, 2019.

[19] G. Garrison, M. Mansukhani, and B. Bohn, "Predictors of thirty-day readmission among hospitalized family medicine patients," J. of the Am. Board of Family Medicine, vol. 26, pp. 71–77, 2013.

[20] R. Robinson and T. Hudali, "The HOSPITAL score and LACE index as predictors of 30 day readmission in a retrospective study at a university-affiliated community hospital." PeerJ, vol. 5, p. e3137, 2017.

[21] E. Ritt and P. Nathan, "The identification of post-acute patients at risk for hospital readmission: Clinical implications of the LACE index." Clinical Nursing Studies, vol. 4, no. 3, 2016.

[22] L. Spiva, M. Hand, L. Vanbrackle, and F. Mcvay, "Validation of a predictive model to identify patients at high risk for hospital readmission," J. for Healthc. Qual., vol. 38, no. 1, pp. 34–41, 2014.

[23] D. Morgan, B. Bame, P. Zimand et al., "Assessment of machine learning vs standard prediction rules for predicting hospital readmissions," JAMA Network Open, vol. 2, p. e190348, 2019.

[24] A. Elixhauser, C. Steiner et al., "Comorbidity measures for use with administrative data," Medical care, vol. 36, pp. 8–27, 1998.

[25] M. Charlson, P. Pompei, K. Ales et al., "A new method of classifying prognostic comorbidity in longitudinal studies: development and validation." J Chronic Dis, vol. 40, no. 5, pp. 373–383, 1987.

[26] H. Quan, V. Sundararajan, P. Halfon et al., "Coding algorithms for defining comorbidities in icd-9-cm and icd-10 administrative data," Medical care, vol. 43, pp. 1130–1139, 2005.

[27] C. Walraven, P. Austin et al., "A modification of the elixhauser comorbidity measures into a point system for hospital death using administrative data," Medical care, vol. 47, pp. 626–33, 2009.

[28] C. Baechle, A. Agarwal, R. Behara, and X. Zhu, "A cost sensitive approach to predicting 30-day hospital readmission in copd patients," 2017 IEEE-EMBS BHI Conf., pp. 317–320, 2017.

[29] S. Choudhry, J.Li, D. Davis, C. Erdmann, R. Sikka, and B. Sutariya, "A public-private partnership develops and externally validates a 30-day hospital readmission risk prediction model." Online J Public Health Inform, vol. 5, no. 2, p. 219, 2013.

[30] R. Vaithianathan, N. Jiang, and T.Ashton, "A Model for Predicting Readmission Risk in New Zealand." Auckland University of Technology, Department of Economics, Tech. Rep., 2012.

[31] E. Logue *et al.*, "Admission Data Predict High Hospital Readmission Risk." *J Am Board Fam Med*, vol. 29, no. 1, pp. 50–59, 2016.

[32] K. Zolfaghar, N. Meadem, A. Teredesai *et al.*, "Big data solutions for predicting risk-of-readmission for congestive heart failure patients," *IEEE Intel. Conf. on Big Data*, pp. 64–71, 2013.

[33] M. Jovanovic, S. Radovanovic *et al.*, "Building interpretable predictive models for pediatric hospital readmission using tree-lasso logistic regression," *Artif. Intell. in Med.*, vol. 72, pp. 12–21, 2016.

[34] G. Garrison, P. Robelia, J. Pecina *et al.*, "Comparing performance of 30-day readmission risk classifiers among hospitalized primary care patients." *J. of Eva. in Clinical Practice*, vol. 23, 2016.

[35] J. Pederson, S. Majumdar, M. Forhan, J. Johnson, and F. McAlister, "PROACTIVE Investigators. Current depressive symptoms but not history of depression predict hospital readmission or death after discharge from medical wards: a multisite prospective cohort study." *Gen Hosp Psychiatry*, vol. 39, pp. 80–85, 2016.

[36] C. Walraven, I. Dhalla, C. Bell *et al.*, "Derivation and Validation of an Index to Predict Early Death or Unplanned Readmission after Discharge from Hospital to the Community." *CMAJ*, vol. 33, no. 6, pp. 551–557, 2010.

[37] C. Hebert, C. Shivade *et al.*, "Diagnosis-specific readmission risk prediction using electronic health data: a retrospective cohort study." *BMC Med Inform Decis Mak*, vol. 14, p. 65, 2014.

[38] E. Yian, H.Zhou, A. S. A *et al.*, "Early Hospital Readmission and Mortality Risk after Surgical Treatment of Proximal Humerus Fractures in a Community-Based Health Care Organization." *Perm J*, vol. 20, no. 1, pp. 47–52, 2016.

[39] G. Schaefer *et al.*, "Evaluation of the Yale New Haven Readmission Risk Score for Pneumonia in a General Hospital Population." *Am J Med*, vol. 130, no. 9, p. 1107-1111, 2017.

[40] D. Krompass, C. Esteban, V. Tresp *et al.*, "Exploiting latent embeddings of nominal clinical data for predicting hospital readmission," *KI - Künstliche Intelligenz*, vol. 29, pp. 153–159, 2014.

[41] B. Ruiz, M. García, U. Aguirre, and C. C. Aguirre, "Factors predicting hospital readmissions related to adverse drug reactions," *European J. of clinical pharmacology*, vol. 64, no. 7, p. 715—722, 2008.

[42] L. Glance *et al.*, "Hospital readmission after noncardiac surgery the role of major complications," *JAMA surgery*, vol. 149, 2014.

[43] O. Hasan, D. Meltzer, S. Shaykevich *et al.*, "Hospital readmission in general medicine patients: a prediction model." *J Gen Intern Med*, vol. 25, no. 3, pp. 211–219, 2010.

[44] C. Feudtner, J. Levin *et al.*, "How well can hospital readmission be predicted in a cohort of hospitalized children? a retrospective, multicenter study," *Pediatrics*, vol. 123, pp. 286–293, 2009.

[45] C. King, S. Atwood, and M. Lozada, "Identifying risk factors for 30-day readmission events among American Indian patients with diabetes in the Four Corners region of the southwest from 2009 to 2016." *PLoS One*, vol. 13, no. 8, p. e0195476, 2018.

[46] M. Tsuchihashi, H. Tsutsui *et al.*, "Medical and socioenvironmental predictors of hospital readmission in patients with congestive heart failure," *American heart journal*, vol. 142, p. E7, 11 2001.

[47] M. Barnett, J. Hsu, and J. McWilliams, "Patient characteristics and differences in hospital readmission rates," *JAMA internal medicine*, vol. 175, 09 2015.

[48] K. Erickson, W. Winkelmayer, G. Chertow *et al.*, "Physician visits and 30-day hospital readmissions in patients receiving hemodialysis," *J. of Am. Society of Nephrology : JASN*, vol. 25, 2014.

[49] M. Shulan *et al.*, "Predicting 30-day all-cause hospital readmissions," *Health care manag. sci.*, vol. 16, no. 2, pp. 167–175, 2013.

[50] K. Zhu, Z. Lou, J. Zhou *et al.*, "Predicting 30-day Hospital Readmission with Publicly Available Administrative Database. A Conditional Logistic Regression Modeling Approach." *Methods Inf Med*, vol. 54, no. 6, pp. 560–567, 2015.

[51] M. McGirt *et al.*, "Prediction model for outcome after low-back surgery: individualized likelihood of complication, hospital readmission, return to work, and 12-month improvement in functional disability." *Neurosurg Focus*, vol. 39, no. 6, p. E13, 2015.

[52] R. Wallmann, J. Llorca *et al.*, "Prediction of 30-day cardiac-related-emergency-readmissions using simple administrative hospital data." *Int J Cardiol*, vol. 164, no. 2, pp. 193–200, 2013.

[53] E. Philbin *et al.*, "Prediction of hospital readmission for heart failure: development of a simple risk score based on administrative data." *J Am Coll Cardiol*, vol. 33, no. 6, pp. 1560–1566, 1999.

[54] J. Flythe, S. Katsanos, Y. Hu, A. Kshirsagar, R. Falk, and C. Moore, "Predictors of 30-day hospital readmission among maintenance hemodialysis patients: A hospitals perspective," *Clinical Journal of the American Society of Nephrology*, vol. 11, 2016.

[55] J. Hamner and K. Ellison, "Predictors of hospital readmission after discharge in patients with congestive heart failure," *Heart & lung : the journal of critical care*, vol. 34, pp. 231–9, 2005.

[56] P. Coventry, I. Gemmell, and C. Todd, "Psychosocial risk factors for hospital readmission in copd patients on early discharge services: A cohort study," *BMC pulm. med.*, vol. 11, p. 49, 2011.

[57] M. Silverstein, H. Qin, S. Mercer *et al.*, "Risk factors for 30-day hospital readmission in patients >= 65 years of age." *Proc (Bayl Univ Med Cent)*, vol. 21, no. 4, pp. 363–372, 2008.

[58] M. Kassin, R. Owen, S. Perez *et al.*, "Risk factors for 30-day hospital readmission among general surgery patients." *J Am Coll Surg*, vol. 215, no. 3, pp. 322–330, 2012.

[59] E. Graboyes, T. Liou, B. Nussenbaum, and J. Diaz, "Risk factors for unplanned hospital readmission in otolaryngology patients," *Otolaryngology–head and neck surgery*, vol. 149, pp. 562–571, 2013.

[60] J. Mcphee, L. Nguyen, K. Ho, C. Ozaki, M. Conte, and M. Belkin, "Risk prediction of 30-day readmission after infrainguinal bypass for critical limb ischemia," *Journal of vascular surgery*, vol. 57, 2013.

[61] C. Baillie, C. VanZandbergen, G. Tait *et al.*, "The readmission risk flag: Using the electronic health record to automatically identify patients at risk for 30-day readmission," *Journal of hospital medicine*, vol. 8, no. 12, pp. 689–695, 2013.

[62] S. Howell, M. Coory, J. Martin, and S. Duckett, "Using routine inpatient data to identify patients at risk of hospital readmission," *BMC health services research*, vol. 9, p. 96, 2009.

[63] M. Almardini and Z. Ras, "A supervised model for predicting the risk of mortality and hospital readmissions for newly admitted patients," *23rd Intl. Sym. on Foundations of Intel. Sys. (ISMIS)*, 2017.

[64] H. Zhao, S. Tanner, S. Golden *et al.*, "Different sampling and modeling strategies affect analyses of readmission risk among diabetes patients," *Diabetes*, vol. 67, 2018.

[65] D. He, S. Mathews, A. Kalloo, and S. Hutfless, "Mining high dimensional claims data to predict early hospital readmissions," *J. of the American Medical Informatics Association*, vol. 21, 2013.

[66] J. Friedmann, G. Jensen *et al.*, "Predicting early nonelective hospital readmission in nutritionally compromised older adults." *Am J Clin Nutr*, vol. 65, no. 6, pp. 1714–1720, 1997.

[67] D. Bertsimas, J. Pauphilet *et al.*, "Predicting inpatient flow at a major hospital using interpretable analytics," *medRxiv*, 2020.

[68] L. Barker, A. Gruneir, K. Fung *et al.*, "Predicting psychiatric readmission: sex-specific models to predict 30-day readmission following acute psychiatric hospitalization," *Soc Psychiatry Psychiatr Epidemiol*, vol. 53, pp. 139–149, 2018.

[69] L. Lam *et al.*, "Gender differences in the rate of 30-day readmissions after percutaneous coronary intervention for acute coronary syndrome," *Womens Health Issues*, vol. 29, no. 1, pp. 17–22, 2018.

[70] R. Rodriguez-Gutierrez, J. Herrin *et al.*, "Racial and ethnic differences in 30-day hospital readmissions among us adults with diabetes," *JAMA Netw Open*, vol. 2, no. 10, p. e1913249, 2019.

[71] B. Kennedy, "Does race predict stroke readmission? an analysis using the truncated negative binomial model," *J Natl Med Assoc*, vol. 97, no. 5, p. 699–713, 2005.

[72] A. Hosseinzadeh, M. Izadi, A. Verma *et al.*, "Assessing the predictability of hospital readmission using machine learning," *Proc. AAAI Conf.*, pp. 1532–1538, 2013.

[73] S. Fisher, J. Graham, S. Krishnan, and K. Ottenbacher, "Predictors of 30-day readmission following inpatient rehabilitation for patients at high risk," *Physical therapy*, vol. 96, 2015.

[74] N. Flaks-Manov, E. Srulovici, R. Yahalom *et al.*, "Preventing hospital readmissions: Healthcare providers' perspectives on "impactibility" beyond ehr 30-day readmission risk prediction," *J. of General Internal Medicine*, vol. 35, 2020.

[75] J. Hilbert, S. Zasadil *et al.*, "Using decision trees to manage hospital readmission risk for acute myocardial infarction, heart failure, and pneumonia," *Appl Health Econ Hea*, vol. 12, 2014.

[76] M. Vukicevic, S. Radovanovicand, A. Kovacevic, S. G, and Z. Obradovic, "Improving hospital readmission prediction using domain knowledge based virtual examples," *Lecture Notes in Business Information Processing*, vol. 224, pp. 695–706, 2015.

[77] P. Wolff, M. Graña, S. Ríos, and M. Yarza, "Machine learning readmission risk modeling: A pediatric case study," *BioMed Research International*, vol. 2019, pp. 1–9, 2019.

[78] A. D. A. Dey, G. Kao, and C. Low, "Modeling biobehavioral rhythms with passive sensing in the wild: A case study to predict readmission risk after pancreatic surgery," *Proc. of the ACM on Interac., Mobile, Wearable & Ubiquitous Tech.*, vol. 3, pp. 1–21, 2019.

[79] R. Duggal, S. Shukla, S. Chandra *et al.*, "Predictive risk modelling for early hospital readmission of patients with diabetes in India." *Int J Diabetes Dev Ctries*, vol. 36, pp. 519–528, 2016.

[80] K. Shameer, K. Johnson, A. Yahi *et al.*, "Predictive modeling of hospital readmission rates using electronic medical record0wide machine learning: A case study using mount sinai heart failure cohort," *Pac Symp Biocomput*, vol. 22, pp. 276–287, 2017.

[81] Y. Luo and A. Rumshisky, "Interpretable topic features for post-icu mortality prediction," *AMIA Annual Symposium Proc.*, vol. 2016, pp. 827–836, 2017.

[82] K. Zolfaghar *et al.* (2013) Predicting risk-of-readmission for congestive heart failure patients: A multi-layer approach. [Online]. Available: https://arxiv.org/abs/1306.2094v1

[83] M. Mazurowski, P. Habas, J. Zurada *et al.*, "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," *Neural Networks*, vol. 21, pp. 427–436, 2008.

[84] W. Liu, K. Singh, A. Ryan *et al.*, "Predicting 30-day hospital readmissions using artificial neural networks with medical code embedding," *PLoS One*, vol. 15, no. 4, 2020.

[85] E. Choi, T. Bahadori *et al.*, "Multi-layer representation learning for medical concepts," *Workshop track - ICLR*, pp. 1495–1504, 2016.

[86] P. Chen, W. Dong, J. Wang *et al.*, "Interpretable clinical prediction via attention-based neural network," *BMC Med Inform Decis Mak*, vol. 20, no. Suppl 3, p. 131, 2020.

[87] J.Futoma, J. Morris, and J. Lucas, "A comparison of models for predicting early hospital readmissions." *J Biomed Inform*, vol. 56, pp. 229–238, 2015.

[88] L. Turgemanand *et al.*, "A mixed-ensemble model for hospital readmission." *Artif. Intell. in Med.*, vol. 72, pp. 72–88, 2016.

[89] S. Jiang, K. Chin, G. Qu, and K. Tsui, "An integrated machine learning framework for hospital readmission prediction." *Knowledge-Based Systems*, vol. 146, pp. 73–90, 2018.

[90] A. Choudhury and C. Greene, "Evaluating Patient Readmission Risk: A Predictive Analytics Approach." *American Journal of Engineering and Applied Sciences*, vol. 11, no. 4, pp. 1320–1331, 2018.

[91] J. Zhang, S. Yoon, M. Khasawneh *et al.*, "Hospital readmission prediction using swarm intelligence-based support vector machines," *IIE Annual Conf. and Expo*, pp. 1522–1531, 2013.

[92] A. Junqueira *et al.*, "A machine learning model for predicting icu readmissions and key risk factors: analysis from a longitudinal health records," *Health and Technology*, vol. 9, pp. 297–309, 2019.

[93] T. Brisimi, R. Chen, T. Mela *et al.*, "Federated learning of predictive models from federated electronic health records," *Intl. J. of Informatics*, vol. 112, pp. 59–67, 2018.

[94] V. Betihavas, S. Frost, P. Newton *et al.*, "An Absolute Risk Prediction Model to Determine Unplanned Cardiovascular Readmissions for Adults with Chronic Heart Failure." *Heart, Lung and Circulation*, vol. 24, pp. 1068–1073, 2015.

[95] S. Bae, A. K. Dey, and C. A. Low, "Using passively collected sedentary behavior to predict hospital readmission," *Proc. of the ACM Intl. Joint Conf. on Pervasive and Ubiquitous Computing*, 2016.

[96] K. Yu and X. Xie, "Predicting hospital readmission: A joint ensemble-learning model," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, pp. 447–456, 2020.

[97] H. Wang, Z. Cui, Y. Chen *et al.*, "Predicting hospital readmission via cost-sensitive deep learning," *IEEE/ACM Trans. on Computational Biology and Bioinformatics*, vol. 15, no. 6, pp. 1968–1978, 2018.

[98] A. Rumshisky, M. Ghassemi *et al.*, "Predicting early psychiatric readmission with natural language processing of narrative discharge summaries," *Translational Psychiatry*, vol. 6, no. e921, 2016.

[99] C. Xiao, T. Ma, A. Dieng, D. Blei, and F. Wang, "Readmission prediction via deep contextual embedding of clinical concepts." *PLoS One*, vol. 13, no. 4, p. e0195024, 2018.

[100] B. Reddy and D. Delen, "Predicting hospital readmission for lupus patients: An RNN-LSTM-based deep-learning methodology." *Comput Biol Med*, vol. 101, pp. 199–209, 2018.

[101] S. Golas, T. Shibahara, S. Agboola *et al.*, "A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: A retrospective analysis of electronic medical records data," *BMC Med. Inform. Decis.*, vol. 18, p. 44, 12 2018.

[102] S. Jauhari, S. Shukla *et al.* (2020) Embpred30: Assessing 30-days readmission for diabetic patients using categorical embeddings. [Online]. Available: https://arxiv.org/abs/2002.11215

[103] X. Zhu, J. Hurtado, and H. Tao, "Localized sampling for hospital re-admission prediction with imbalanced sample distributions," *Proc. of International Joint Conference on Neural networks*, 2017.

[104] A. Rajkomar, E. Oren *et al.*, "Scalable and accurate deep learning for electronic health records," *npj Digital Medicine*, vol. 1, 2018.

[105] Z. Huo, H. Sundararajhan, N. Hurley *et al.*, "Sparse embedding for interpretable hospital admission prediction," *IEEE Eng Med Biol Soc*, vol. 2019, pp. 3438–3441, 2019.

[106] R. Merkow, M. Ju, J. Chung *et al.*, "Underlying reasons associated with hospital readmission following surgery in the united states," *JAMA*, vol. 313, pp. 483–95, 2015.

[107] L. Huang, A. Shea, H. Qian *et al.*, "Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records," *J. of Biomedical Informatics*, vol. 99, 2019.

[108] C. Olson, S. Dey, V. Kumar *et al.*, "Clustering of elderly patient subgroups to identify medication-related readmission risks," *Intl. J. of Medical Informatics*, vol. 85, no. 1, pp. 43–52, 2016.

[109] S. Hao, Y. Wang, B. Jin *et al.*, "Development, Validation and Deployment of a Real Time 30 Day Hospital Readmission Risk Assessment Tool in the Maine Healthcare Information Exchange." *PLoS One*, vol. 10, no. 10, p. e0140271, 2015.

[110] J. Jiang, S. Hewner, and V. Chandola, "Hospital readmission prediction - applying hierarchical sparsity norms for interpretable models," *ArXiv*, vol. abs/1804.01188, 2018.

[111] R. Fornango *et al.* Project evaluation activity in support of partnership for patients interim evaluation report, final. [Online]. Available: https://downloads.cms.gov/files/cmmi/pfp-interimevalrpt.pdf

[112] Hospital readmissions reduction program (hrrp). [Online]. Available: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program

[113] M. McAdams-DeMarc, A.Law, M. Salter *et al.*, "Frailty and Early Hospital Readmission After Kidney Transplantation." *American Journal of Transplantation*, vol. 13, pp. 2091–2095, 2013.

[114] D. Kansagara, H. Englander, A. Salanitro *et al.*, "Risk Prediction Models for Hospital Readmission: A Systematic Review." *JAMA*, vol. 306, no. 15, pp. 1688–1698, 2011.

[115] B. Goldstein, A. Navar, M. Pencina, and J. Ioannidis, "Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review." *J Am Med Inform Assoc*, vol. 24, no. 1, pp. 198–208, 2017.

[116] E. Mahmoudi, N. Kamdar, N. Kim *et al.*, "Use of electronic medical records in development and validation of risk prediction models of hospital readmission: systematic review," *British Medical Journal (BMJ)*, 2020.

[117] Overview of the nationwide readmissions database (nrd). [Online]. Available: https://www.hcup-us.ahrq.gov/nrdoverview.jsp

[118] Nis database documentation. [Online]. Available: https://www.hcup-us.ahrq.gov/db/nation/nis/nisdbdocumentation.jsp

[119] Z. Liu, W.Cao *et al.*, "Self-paced ensemble for highly imbalanced massive data classification," *CoRR*, vol. abs/1909.03500, 2019.

[120] C. Klabunde, D. Joseph, J. King *et al.*, "Vital Signs: Colorectal Cancer Screening Test Use—United States, 2012." Centers for Disease Control and Prevention, Tech. Rep., 2013.

[121] Health, united states, 2012: With special feature on emergency care. [Online]. Available: https://www.cdc.gov/nchs/data/hus/hus12.pdf

[122] T. Miller., "Explanation in artificial intelligence: Insights from the social sciences." *Artificial Intelligence*, vol. 267, pp. 1–38, 2019.

[123] J. Wasfy, C. Zigler, C. Choirat *et al.*, "Readmission Rates After Passage of the Hospital Readmissions Reduction Program: A Pre-Post Analysis." *Ann Intern Med*, vol. 166, no. 5, pp. 324–331, 2017.

[124] S. Upadhyay, A. Stephenson, and D. Smith., "Readmission Rates and Their Impact on Hospital Financial Performance: A Study of Washington Hospitals." *Inquiry*, vol. 56, pp. 1–10, 2019.

[125] I. Cohen, R. Amarasingham *et al.*, "The legal and ethical concerns that arise from using complex predictive analytics in health care." *Health Aff (Millwood)*, vol. 33, no. 7, pp. 1139–1147, 2014.

[126] Health insurance portability and accountability act of 1996 (hipaa). [Online]. Available: https://www.cdc.gov/phlp/publications/topic/hipaa.html

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2021.3089682, IEEE/ACM Transactions on Computational Biology and Bioinformatics

IEEE/ACM TRANS. ON COMPUTATIONAL BIOLOGY & BIOINFORMATICS, SEP., 2020 20

[127] H. He and E. Garcia, "Learning from imbalanced data," *IEEE Trans. on Knwl. & Data Eng.*, vol. 21, pp. 1263 – 1284, 10 2009.

[128] P. Branco *et al.*, "A survey of predictive modelling under imbalanced distributions," *CoRR*, vol. abs/1505.01658, 2015.

[129] A. Johnson, T. Pollard *et al.* Mimic-iii clinical database (version 1.4). [Online]. Available: https://doi.org/10.13026/C2XW26

[130] N. Chawla *et al.*, "Smote: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[131] G. M. N. Torelli, "Training and assessing classification rules with imbalanced data," *Data Mining and Knowledge Discovery*, 2014.

[132] U. M. L. Repository. Diabetes 130-us hospitals for years 1999-2008 data set. [Online]. Available: https://archive.ics.uci.edu/ml/index.php

[133] J.-F. Díez-Pastor, J. Rodríguez, C. García-Osorio, and L. Kuncheva, "Random balance: Ensembles of variable priors classifiers for imbalanced data," *Knwl.-Based Sys.*, vol. 85, 2015.

[134] G. Du, J. Zhang, Z. Luo, F. Ma, L. Ma, and S. Li, "Joint imbalanced classification and feature selection for hospital readmissions," *Knowl. Based Syst.*, vol. 200, p. 106020, 2020.

[135] D. Opitz and R. Marlin, "Popular ensemble methods: An empirical study," *J. of Artif. Intell. Research*, vol. 11, pp. 168–198, 1999.

[136] K. Tu, T. Mitiku, N. I. vers *et al.*, "Evaluation of electronic medical record administrative data linked database (emrald)," *Am J Manag Care*, vol. 20, no. 1, pp. e15–e20, 2014.

[137] B. Brooke, D. Martino, M. Girotti, J. Dimick, and P. Goodney, "Developing strategies for predicting and preventing readmissions in vascular surgery." *J Vasc Surg*, vol. 56, no. 2, pp. 556–562, 2012.

[138] S. Matthew and H. Kharrazi, "Feasibility of 30-day hospital readmission prediction modeling based on health information exchange data," *Intl. J. of Medical Informatics*, vol. 84, 09 2015.

[139] D. Rubin, "Hospital readmission of patients with diabetes," *Current diabetes reports*, vol. 15, p. 584, 2015.

[140] R. Kiran, C. Delaney, A. Senagore *et al.*, "Outcomes and prediction of hospital readmission after intestinal surgery," *J. of the Am. Col. of Surgeons*, vol. 198, pp. 877–83, 2004.

[141] D. Mcmanus, J. Saczynski, and D. Lessard, "Reliability of predicting early hospital readmission after discharge for an acute coronary syndrome using claims-based data," *The American Journal of Cardiology*, vol. 117, no. 4, pp. 501–507, 2015.

[142] F. Kareliusson, L. Geer, and A. Tibblin, "Risk prediction of icu readmission in a mixed surgical and medical population," *Journal of intensive care*, vol. 3, p. 30, 2015.

[143] P. Halfon, Y. Egglis *et al.*, "Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care," *Medical care*, vol. 44, pp. 972–81, 2006.

[144] R. Dreler, I. Ranasinghe, Y. Wang *et al.*, "Sex differences in the rate, timing and principal diagnoses of 30-day readmissions in younger patients with acute myocardial infarct," *Circulation*, vol. 132, no. 3, pp. 158–166, 2015.

[145] S. Radovanovic, M. Vukicevic, A. Kovacevic *et al.*, "Domain knowledge based hierarchical feature selection for 30-day hospital readmission prediction," *Springer, Cham*, vol. 9105, 2015.

[146] R. Deyo, D. Cherkin, and M. Ciol, "Adapting a clinical comorbidity index for use with icd-9-cm administrative databases," *J. of Clinical Epidemiology*, vol. 45, no. 6, pp. 613–619, 1992.

[147] L. Sullivan, J. Massaro, and R. D'Agostino, "Presentation of multivariate data for clinical use: the framingham study risk score functions," *Stat Med*, vol. 23, pp. 1631–1660, 2004.

[148] T. Ono, A. Tamai, D. Takeuchi, and Y. Tamai, "Factors related to readmission to a ward for dementia patients: Sex differences," *Psychiatry Clin Neurosci*, vol. 65, pp. 490–498, 2011.

[149] A. Pandey, N. Keshvani, R. Khera *et al.*, "Temporal trends in racial differences in 30-day readmission and mortality rates after acute myocardial infarction among medicare beneficiaries," *JAMA Cardiol*, vol. 5, no. 2, pp. 136–145, 2020.

[150] C. Hilton, A. Milinovich, C. Felix *et al.*, "Personalized predictions of patient outcomes during and after hospitalization using artificial intelligence," *npj Digit. Med*, vol. 3, no. 51, 2020.

[151] E. Choi, M. Bahadori, J. Kulas *et al.*, "Retain: An interpretable predictive model for healthcare using reverse time attention mechanism," *Proc. of the NIPS*, pp. 3504–3512, 2016.

[152] D. Ramage, D. Hall, R. Nallapati, and C. Manning, "Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora," *EMNLP Conf.*, vol. 1, 2009.

[153] Y. Lou, R. Caruana, J. Gehrke *et al.*, "Accurate intelligible models with pairwise interactions," *Proc. of KDD Conf.*, pp. 623–631, 2013.

[154] R. Caruana, Y. Lou, J. Gehrke *et al.*, "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission," *Proc. of the KDD Conf.*, pp. 1721–1730, 2015.

[155] X. Zhang *et al.*, "Inprem: An interpretable and trustworthy predictive model for healthcare," *Proc. of the KDD*, pp. 450–460, 2020.

[156] R. Zuckerman, S. Sheingold, E. Orav *et al.*, "Readmissions, observation, and the hospital readmissions reduction program," *New England Journal of Medicine*, vol. 374, 2016.

[157] A. Jha, "To fix the hospital readmissions program, prioritize what matters," *JAMA*, vol. 319, no. 5, pp. 431–433, 2018.

[158] S. Zheng, A. Hanchate, and M. Shwartz, "One-year costs of medical admissions with and without a 30-day readmission and enhanced risk adjustment," *BMC Health Serv. Res.*, vol. 19, 2019.

[159] P. Chan, B. Nallamothu *et al.*, "Readmission rates and long-term hospital costs among survivors of in-hospital cardiac arrest," *Circ Cardiovasc Qual Outcomes*, vol. 7, no. 6, pp. 889–895, 2014.

[160] C. Hsuan, B. Carr, R. Hsia, and G. Hoffman, "Assessment of hospital readmissions from the emergency department after implementation of medicare's hospital readmissions reduction program," *JAMA Netw Open*, vol. 3, no. 5, p. e203857, 2020.

[161] A. Smeraglio, P. Heidenreich, G. Krishnan, J. Hopkins, J. Chen, and L. Shieh, "Patient vs provider perspectives of 30-day hospital readmissions," *BMJ Open Quality*, vol. 8, p. e000264, 2019.

[162] S. Dilworth, I. Higgins, and V. Parker, "Feeling let down: An exploratory study of the experiences of older people who were readmitted to hospital following a recent discharge," *Contemporary nurse*, vol. 42, pp. 280–8, 2012.

[163] Z. Siddiqui, S. Berry, A. Bertram *et al.*, "Does patient experience predict 30-day readmission? a patient-level analysis of hcahps data," *Journal of hospital medicine*, vol. 13, pp. 681–687, 2018.

[164] A. Tripathi, J. Abbott, G. Fonarow *et al.*, "Thirty-day readmission rate and costs after percutaneous coronary intervention in the united states: A national readmission database analysis," *Circulation: Cardiovascular Interventions*, vol. 10, p. e005925, 2017.

[165] M. Huntington, A. Guzman, A. Roemen *et al.*, "Hospital-to-home: a hospital readmission reduction program for congestive heart failure," *South Dakota medicine*, vol. 66, pp. 370–373, 2013.

[166] R. Wadhera, M. Joynt, J. Wasfy *et al.*, "Association of the hospital readmissions reduction program with mortality among medicare beneficiaries hospitalized for heart failure, acute myocardial infarction, and pneumonia," *JAMA*, vol. 320, p. 2542, 2018.

[167] M. Blum, H. Øien, C. armichael *et al.*, "Cost-effectiveness of transitional care services after hospitalization with heart failure," *An. of internal medicine*, vol. 172, 2020.

[168] S. Tan, J. Fen, C. Joyce, J. Fisher *et al.*, "Association of hospital discharge against medical advice with readmission and in-hospital mortality," *JAMA Network Open*, vol. 3, p. e206009, 2020.

[169] K. Emam, *Guide to the De-Identification of Personal Health Information.* Auerbach Publications, 2013.

Shuwen Wang is a PhD student in the Dept. of Computer and Electrical Engineering & Computer Science, Florida Atlantic University, Boca Raton, FL, USA. Her research mainly focuses on data mining, machine learning, and medical data analysis.

Xingquan Zhu (SM'12) is a Full Professor in the Dept. of Computer and Electrical Engineering & Computer Science, Florida Atlantic University, Boca Raton, FL, USA. His research interests include data mining, machine learning, and bioinformatics. He is an Associate Editor of the IEEE TRANS. ON KNOWL. & DATA ENG. (2008 to 2012, and from 2014 to date), and an Associate Editor of the ACM TRANS. ON KNOWl. DISCOVERY FROM DATA (2017 to date).