Expert Systems With Applications 176 (2021) 114791

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Towards graph-based class-imbalance learning for hospital readmission

Guodong Du [a], Jia Zhang [b], Fenglong Ma [c], Min Zhao [d], Yaojin Lin [e,*], Shaozi Li [a,*]

[a] Department of Artificial Intelligence, Xiamen University, Xiamen 361005, People's Republic of China
[b] College of Information Science and Technology, Jinan University, Guangzhou 510632, People's Republic of China
[c] College of Information Sciences and Technology, Pennsylvania State University, University Park 16802, USA
[d] Informatics Center, The First Affiliated Hospital of Xiamen University, Xiamen 361003, People's Republic of China
[e] School of Computer Science, Minnan Normal University, Zhangzhou 363000, People's Republic of China

ARTICLE INFO

Keywords:
Hospital readmission
Graph embedding
Class-imbalance learning
Neural network model

ABSTRACT

Predicting hospital readmission with effective machine learning techniques has attracted a great attention in recent years. The fundamental challenge of this task stems from characteristics of the data extracted from electronic health records (EHR), which are imbalanced class distributions. This challenge further leads to the failure of most existing models that only provide a partial understanding for the learning problem and result in a biased and inaccurate prediction. To address this challenge, we propose a new graph-based class-imbalance learning method by fully making use of the data from different classes. First, we conduct graph construction for learning the pattern discrimination from between-class and within-class data samples. Then we design an optimization framework to incorporate the constructed graphs to obtain a class-imbalance aware graph embedding and further alleviate performance degeneration. Finally, we design a neural network model as the classifier to conduct imbalanced classification, i.e., hospital readmission prediction. Comprehensive experiments on six real-world readmission datasets show that the proposed method outperforms state-of-the-art approaches in readmission prediction task.

1. Introduction

Hospital readmission is one of crucial problems in the field of healthcare, which may happen due to doctor negligence in providing patients with necessary medicines, or patients failing to follow treatment instructions after discharge (Mardini & Raś, 2019; Wang et al., 2018; Arancibia et al., 2019). As we know, readmission has a significant financial influence on stakeholders, such as government agencies, hospitals and patients, which will increase the cost for patients, hospitals, and taxpayers. According to the survey in America hospitals, over 20 billion dollars are used on 30-day readmission every year (Agarwal, Baechle, Behara, & Zhu, 2018). In the United Kingdom, 35% of unplanned readmissions are existed and cost 11 billion pounds per year (Cui, Wang, Wang, Yu, & Jin, 2018). Furthermore, many countries, including Australia (Reischl, Eigner, Bodendorf, Schaffer, & Wickramasinghe, 2019), America (Jencks, Williams, & Coleman, 2009), United Kingdom (Ali, Loeffler, Aylin, & Bottle, 2019), China (Yu & Xie, 2020), Canada (Kouyoumdjian et al., 2020) and Brazil (Gama, Backman, & de Oliveira, 2020), have used the unplanned readmission as one of important indicators to evaluate the medical quality of hospitals. In light of this opportunity, it is necessary to perform more effective and intelligent learning algorithms for predicting the hospital readmission risk of the inpatients.

For achieving the purpose, some feasible and effective prediction methods have been proposed to identify the potential readmission of patients in recent years, which can be found in the machine learning, data mining, and healthcare community (Ma et al., 2019). However, most of them are simply designed either for hospital comparison or clinical purpose, such as the prediction directly estimated by some well-known methods (e.g., logistic regression (Jeejeebhoy et al., 2015) and support vector machine (Braga, Portela, Santos, & Rua, 2014)), which do not deliver a satisfactory performance. In addition, to model the problem of readmission prediction based on electronic health records (EHR) data, mining the inherent property is beneficial to guide the learning process (Ma et al., 2018), e.g., the issue of the class-imbalance. In fact, readmission prediction is such an intrinsically class-imbalance problem (Lee et al., 2019). For example, in congestive heart failure (CHF), the average 30-day readmission rate is about 24.7% for the

* Corresponding authors.
E-mail addresses: duguodong@stu.xmu.edu.cn (G. Du), jiazhang@jnu.edu.cn (J. Zhang), fenglong@psu.edu (F. Ma), zhaomin77@stu.xmu.edu.cn (M. Zhao), yjlin@mnnu.edu.cn (Y. Lin), szlig@xmu.edu.cn (S. Li).

https://doi.org/10.1016/j.eswa.2021.114791
Received 29 October 2020; Received in revised form 8 January 2021; Accepted 23 February 2021
Available online 8 March 2021

patients; in pneumonia (PNA) and acute myocardial infarction (AMI), the average 30-day readmission rates are more lower to 18.5% and 19.8% respectively (Du, Zhang, Li, & Li, 2021). The low prevalence of readmission brings the challenge in identifying the minority class, i.e., readmitted patients. Nevertheless, most of existing literature has limited studies of methods to handle imbalanced readmission prediction data. Not surprisingly, they may be in trouble to obtain better performance for hospital readmission.

Class-imbalance learning is the paradigm capable of solving the above problem, which naturally emerges for identifying the hard-to-classify samples, thus improving the classification performance. For implementation, widely used strategies include data sampling, ensemble learning, cost-sensitive learning, and dimension reduction, etc (Gu, Angelov, & Soares, 2020; Xu et al., 2019). Among them, data sampling approaches use sampling for the modification of an imbalanced dataset, and representative examples are over-sampling, under-sampling, and hybrid-sampling (Dumpala, Chakraborty, & Kopparapu, 2018; Peng et al., 2019; Seiffert, Khoshgoftaar, Van Hulse, & Napolitano, 2009). Ensemble learning methods consider to construct several complementary classifiers for obtaining a strong classifier, show its effectiveness in adaptability and generalization in dealing with the class-imbalance data (Liu, Wu, & Zhou, 2008; Liu et al., 2020; Díez-Pastor, Rodríguez, García-Osorio, & Kuncheva, 2015). Cost-sensitive learning methods achieve this purpose by assigning appropriate misclassification cost to minority samples (Castro & Braga, 2013). Finally, dimension reduction methods focus on obtaining a low-dimensional space (Zhang et al., 2020; Wang & Zhang, 2007; Wang, 2011), which usually needs to be tightly bound with class distribution, so that the generated space can make a good effect for imbalance classification (Xue & Titterington, 2008). The aforementioned methods have achieved great performance for handling class-imbalance problem, they can also be applied to readmission prediction tasks. However, the information contained in patients EHR data may not sufficient for building a good readmission risk prediction model. Thus, a pre-processing step is helpful to improve the performance of the final result.

To address the above challenge for readmission prediction with class-imbalance data, in this paper, we propose a new graph-based prediction method, namely Graph-CL. Towards a similar purpose as class-imbalance learning with dimension reduction, we are aimed at generating an embedding to achieve the purpose. To be specific, we extract the geometrical and discriminative information on readmission data locally. Based on positive samples, we construct a graph for generating the embedding sensitive to the minority class, and a graph based on negative samples is also constructed to balance the influence of the majority class. Furthermore, we enhance the separability of different classes with the information of the constructed between-class graph. Based on this, we then propose an optimization framework to incorporate these graphs. The optimization framework can be easily solved, hence the optimal solution, i.e., an embedding, can be learned to make positive and negative samples separable. Finally, we use the generated embedding as the input, and presented a fully-connected neural network as the classifier to identify readmitted patients, thereby achieving imbalanced classification for readmission prediction. Experimental results on six real-world readmission datasets, including All-cause, Lace-score, MIMIC, RA, T-carer and Diabetic, show that the proposed algorithm is capable of handling class-imbalance data for the performance improvement, especially suitable to the prediction task of hospital readmission.

In short, the major contributions of our work are as follows: (1) Graph embedding is utilized for the learning from class-imbalance patients data. For this purpose, we develop a unified optimization framework, thus carrying the most discriminative information for readmission prediction. (2) We present an end-to-end trainable prediction model, which is a fully-connected neural network, and employed to map a reliable output. (3) We use six readmission datasets to conduct the experiments, and the results show that the proposed method has the advantage of hospital readmission prediction.

The rest of this paper is organized as follows. In Section 2, we give a brief review of related work on readmission prediction. Section 3 describes the proposed algorithm in detail. Then, the experimental results are demonstrated in Section 4. Finally in Section 5, we conclude the work and discuss several issues for future work.

2. Related work

Hospital readmission prediction is the task for identifying the potential risk of a patient being unexpectedly readmitted to hospital. In general, by using EHR data, such as the information of patients' demographic, social and economic status, treatment and clinical, and healthcare utilization, a high-performance model can be constructed to determine whether or not a readmission is required (Du et al., 2020).

In some of the early work, logistic regression (LR) is frequently adopted to predict the risk with some related factors, which is a state-of-the-art classification model capable of that the target variable is binary. LR has achieved a great success in readmission prediction. For example, Jeejeebhoy et al. (2015) used the LR model to predict the length of hospital stay and the risk for readmission. In the light of that the LR model cannot be used to capture the complex model among different attributive characters, Tutun, Khanmohammadi, He, and Chou (2016) designed a modified LR based method, in which the authors utilized mixed evolutionary simulating annealing LASSO to achieve this purpose for readmission prediction. Moreover, Caruana et al. (2015) developed a generalized additive algorithm to generate the prediction result.

In recent years, from the respective of artificial intelligence in medicine, the research of readmission prediction based on EHR data has made a remarkable achievement. Numerous machine learning models have been applied in this task. Yu and Xie (2020) designed a joint ensemble learning method to solve the class-imbalance problem for readmission prediction, which combined the modified weight boosting algorithm with stacking. Golmohammadi and Radnia, 2016 introduced three machine learning algorithms in their study, including neural network, decision tree-based classification method, and Chi-squared automatic interaction detection, to reach the model with a high level of accuracy. Braga et al. (2014) used SVM, decision tree, and some other methods to evaluate the risk index of the discharged patient. These solutions can predict readmission risk accurately, and help to identify the patient which may need to readmit and return ICU. In addition, Liu, Zolfaghar, Chin, Roy, and Teredesai (2014) designed a principled solution by learning the structure and parameters of a multi-layer Bayesian network from the underlying high-dimensional readmission data. Caruana et al. (2015) proposed an interpretable readmission prediction model to estimate the probability that the event of pneumonia risk maybe occur. Duggal, Shukla, Chandra, Shukla, and Khatri (2016) used naive Bayes classifier and decision tree to achieve the purpose, which is proved to be effective in enhancing the overall performance. Moreover, Mesgarpour, Chaussalet, and Chahed (2019) proposed a generative ensemble risk model of hospital readmission, and reduced the cost of emergency admissions.

With the development of information technologies, many researchers also try to use some advanced machine learning methods, such as deep learning, to achieve the prediction of hospital readmission. For example, Wang et al. (2018) presented a cost-sensitive deep learning method via convolutional neural network and MLP for readmission prediction. Ashfaq, Sant'Anna, Lingman, and Nowaczyk (2019) designed a deep learning framework to predict 30-day unplanned readmission, which used long short-term memory (LSTM) neural network to modeling, also combined expert features with contextual embedding of clinical concepts. Xiao, Ma, Dieng, Blei, and Wang (2018) developed an end-to-end deep sequential prediction model. Specially, probabilistic topic model and recurrent neural network (RNN) were employed to predict the risk of patients with heart failure. Choi et al. (2020) applied a graph convolutional transformer method to hidden

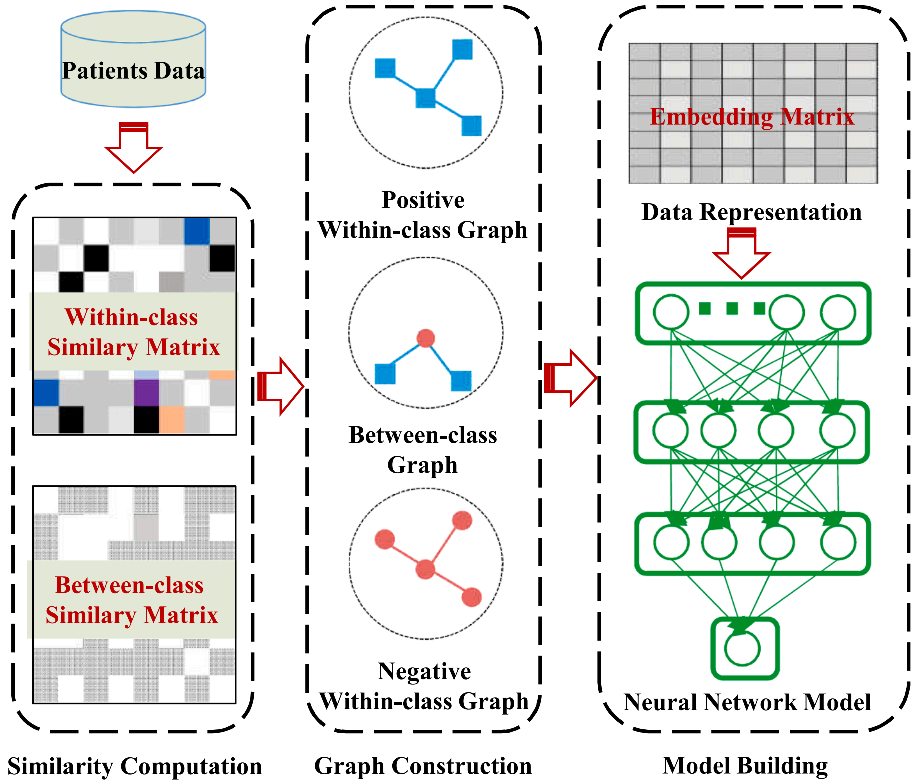


Fig. 1. The framework of the proposed method.

structure exploitation for predicting whether the patient will be admitted to the ICU again.

The aforementioned prediction methods provide many workable solutions to model hospital readmission, however, most of them have neglected to solve the class-imbalance problem, which results in that the performance is not satisfactory, and can be further improved (Lee et al., 2019). Several corrective strategies have been developed in an attempt to solve the class-imbalance problem in readmission prediction. For example, Hosseinzadeh, Izadi, Verma, Precup, and Buckeridge (2013) first utilized random down-sampling to the negative class (not readmitted) data, and then utilized different classification algorithms to predict whether a patient is in readmission. Besides, Zheng et al. (2015) used an over-sampling method to solve the class-imbalance problem, and proposed an SVM-RBF method with particle swarm optimization (PSO) to establish the predictive model. Furthermore, Jiang, Chin, Qu, and Tsui (2018) applied SMOTE algorithm (Chawla, Bowyer, Hall, & Kegelmeyer, 2002) to deal with the class-imbalance problem, and combined with feature selection algorithms to develop a machine learning model to predict hospital readmission, experimental results showed that the proposed method has a good result in several datasets. In addition, Turgeman and May (2016) used a mixed-ensemble model for the hospital readmission prediction, which can minimize the classification error. More recently, Wang, Elkin, and Zhu (2020) designed a random sampling approach to balance the sample distributions, which achieves a good performance in readmission prediction task. In our previous works, we have shown that performing gradient boosting decision tree (Du et al., 2019), feature selection (Du et al., 2020) and heterogeneous data learning (Du et al., 2021) can improve the performance of class-imbalance classification in readmission prediction task. Different to these existing imbalanced learning methods for achieving the readmission prediction task, we propose a graph-based class-imbalance learning method. This method incorporate multiple graphs into a unified framework, thereby carrying the most discriminative information for predict readmitted patients.

3. The framework

In this section, we describe the proposed graph-based class-imbalance learning framework named Graph-CL. This is because of the hospital readmission data has the characteristic with imbalanced class distribution. We are also aware that using traditional learning approaches can cause negative impacts. It is expected that with proper design of a new method, the negative impact of misclassification could be minimized and the performance of hospital readmission risk prediction can be significantly improved. Thus, this framework considers to exploit graph-based data structures for achieving the goal of accurately predicting hospital readmission, which mainly includes three steps, i.e., similarity computation, graph construction, and neural network model designing, as shown in Fig. 1. For similarity computation, some similarity measure methods are employed to capture the pattern discrimination for positive and negative classes respectively. Based on this, adjacent weights of graphs can be indicated by these similarity calculations for graph construction. Here, we construct (positive and negative) within-class graphs and between-class graph to obtain a representation (e.g., an embedding) for patients data. Finally, the generated data representation is used as input, and a neural network model is carefully designed to output the corresponding result, thus achieving the task of hospital readmission prediction.

To formulate the learning problem, some important symbols are first given as follow: We use the data collection $\mathscr{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$ to denote the readmission training dataset, where x_n $(n = 1, \ldots, N)$ denotes the feature information of the n-th patient (generally x_n can be represented by a feature vector), and y_n denotes the corresponding class label. $y_n = 1$ in case that the n-th patient is a readmitted patient, otherwise $y_n = 0$. According to this setting, our goal is to learn an embedding based on the feature information of patients, which is able to characterize the discriminant nature of patients data, hence positive and negative samples can be easily separable.

However, hospital readmission prediction suffers from the issue of the class-imbalance. Thus, achieving the purpose in a class-wise effective manner is crucial to distinguish potential readmitted patients. Aimed at designing an effective prediction method, we consider to apply graph embedding to modelling. In graph embedding, leveraging geometrical distributions of data samples to discover the essential discriminant structures of data is a feasible and popular way. By this way, the global structure of all data samples can be preserved to get a compact feature representation. In the light of that hospital readmission prediction is a special machine learning task based on class-imbalance patients data, such a globality preserving graph embedding may be not effective enough, and a more reasonable manner is that graph embedding should be in locality. It means that graph embedding should be helpful for predicting the minority class, so that readmission patients can be easily separated from many non-readmitted patients in the embedding space. Following this principle, we try to expand the influence of positive samples, balance the influence of negative samples, and increase the separability of different classes, and propose a unified optimization framework incorporating multiple graphs to achieve the purpose. In particular, suppose G_{wp} and G_{wn} denote the within-class graphs in terms of positive and negative samples respectively, and G_b denotes the between-class graph in terms of heterogeneous samples, we define the optimization framework as follow:

$$\underset{\mathbf{Z}}{\operatorname{argmax}} \mathscr{L}_{wp}\left(G_{wp}, \mathbf{Z}\right) + \alpha \mathscr{L}_{wn}\left(G_{wn}, \mathbf{Z}\right) + \beta \mathscr{L}_{b}\left(G_{b}, \mathbf{Z}\right), \tag{1}$$

where $\mathbf{Z}$ denotes the embedding induced from original feature space. $\mathscr{L}_{wp}$ is defined as the loss function to enforce the embedding affected by positive samples, which is implemented by using the graph information G_{wp}, $\mathscr{L}_{wn}$ is to control the influence of negative samples based on G_{wn}, and $\mathscr{L}_b$ is used to increase the local margin between heterogeneous samples based on G_b. In addition, α and β are tradeoff parameters. It can be observed in Eq. (1) that $\mathbf{Z}$ is involved in all the terms. Based on this, the discriminative information of class-imbalance patients data is expect to be well retained for imbalanced classification. In the next Section, we describe the definitions of all the terms in detail.

3.1. Graph based class-imbalance learning

3.1.1. Within-class graph embedding

Graph embedding contributes to the recovery of most of data structures by the appropriate graph construction. According to this property, we consider to exploit the discriminant information on class-imbalance patients data. Specially, positive and negative within-class graphs are designed (using positive and negative sample data) respectively, which enable the local discrimination preservation for different classes. By virtue of these constructed graphs, an embedding can be generated by controlling the influence of positive and negative samples, thereby achieving graph based class-imbalance learning.

Formally, we first provide the solution for embedding based on positive within-class graph G_{wp}, as follow:

$$\mathscr{L}_{wp} = \frac{1}{2}\sum_{i,j}\left\|\mathbf{z}_i - \mathbf{z}_j\right\|^2 W_{wp,ij} = \operatorname{tr}\left(\mathbf{Z}^T \mathbf{L}_{wp} \mathbf{Z}\right), \tag{2}$$

where $\mathbf{L}_{wp} = \mathbf{D}_{wp} - \mathbf{W}_{wp}$ is the graph Laplacian of G_{wp}, $\mathbf{D}_{wp}$ is the diagonal matrix while $D_{wp,ii} = \sum_j W_{wp,ij}$. Besides, $\mathbf{W}_{wp}$ is the weight matrix of data samples, and element $W_{wp,ij} \in \mathbf{W}_{wp}$ measures the similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$. In the light of that the adjacent weights of edges in G_{wp} are discriminatively defined for the local structures preservation among positive samples, naturally, we only keep the similarity information of positive samples in $\mathbf{W}_{wp}$, and implement $\mathbf{W}_{wp}$ by a heat kernel, as follow:

$$W_{wp,ij} = \begin{cases} s\left(\mathbf{x}_i, \mathbf{x}_j\right) & \mathbf{x}_j \in \mathscr{M}_{wp}(\mathbf{x}_i) \text{or} \mathbf{x}_i \in \mathscr{M}_{wp}\left(\mathbf{x}_j\right) \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

where $s\left(\mathbf{x}_i, \mathbf{x}_j\right) = \exp\left(-\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{\sigma}\right)$. Although, σ in the similarity calculation will affect the final results (Wang & Zhang, 2008; Zhang, Luo, Li, Zhou, & Li, 2019; Zhang, Li, Jiang, & Tan, 2021). Thus, Raducanu and Dornaika (2012) suggested us to set σ as the average of squared distances between all pairs. Therefore, our proposed method can automatically construct the graph without parameter tuning. Thus, we can makes the results more robust. $\mathscr{M}_{wp}(\mathbf{x}_i) = \left\{\mathbf{x}_j \middle| y_j = y_i = 1, s(\mathbf{x}_i, \mathbf{x}_j) > \overline{W}(\mathbf{x}_i)\right\}$ denotes the neighbor set of $\mathbf{x}_i$, and $\overline{W}(\mathbf{x}_i)$ is the threshold defined for adaptive search of $\mathbf{x}_i$'s neighbors. For simplicity, we use the average similarity of $\mathbf{x}_i$ with all the other data samples to measure $\overline{W}(\mathbf{x}_i)$, as follow:

$$\overline{W}(\mathbf{x}_i) = \frac{1}{N}\sum_{j=1}^{N} s\left(\mathbf{x}_i, \mathbf{x}_j\right). \tag{4}$$

In short, we attempt to conduct graph embedding based on the data with positive classes, as shown in Eq. (2). By optimizing this term, the influence of positive samples can be expanded for obtaining a class-imbalance aware embedding $\mathbf{Z}$. Next, we use negative within-class graph G_{wn} to regularize the optimization of $\mathbf{Z}$. Similar with the graph embedding regarding positive samples, we define a regularizer as follow:

$$\mathscr{L}_{wn} = \frac{1}{2}\sum_{i,j}\left\|\mathbf{z}_i - \mathbf{z}_j\right\|^2 W_{wn,ij} = \operatorname{tr}\left(\mathbf{Z}^T \mathbf{L}_{wn} \mathbf{Z}\right), \tag{5}$$

where $\mathbf{L}_{wn} = \mathbf{D}_{wn} - \mathbf{W}_{wn}$ for the construction of graph G_{wp}, and $\mathbf{D}_{wn}$ is the diagonal matrix, whose diagonal element is defined as $D_{wn,ii} = \sum_j W_{wn,ij}$. $W_{wn,ij}$ can be calculated with the following formula:

$$W_{wn,ij} = \begin{cases} s\left(\mathbf{x}_i, \mathbf{x}_j\right) & \mathbf{x}_j \in \mathscr{M}_{wn}(\mathbf{x}_i) \text{or} \mathbf{x}_i \in \mathscr{M}_{wn}\left(\mathbf{x}_j\right) \\ 0 & \text{otherwise} \end{cases}, \tag{6}$$

where $\mathscr{M}_{wn}(\mathbf{x}_i) = \left\{\mathbf{x}_j \middle| y_j = y_i = 0, s(\mathbf{x}_i, \mathbf{x}_j) > \overline{W}(\mathbf{x}_i)\right\}$. It is worth mentioning that only difference of the embedding using positive and negative samples is the graph construction. By alternating the calculation of W_{wn}, i.e., the adjacent weights of graph G_{wn}, we can optimize $\mathbf{Z}$ as the influence of negative samples is in control.

3.1.2. Between-class graph embedding

To globally reflect the inherent geometry between classes of patients data, we further consider to construct the graph of training samples from different classes. To be specific, we denote the between-class graph as G_b, and can get:

$$\mathscr{L}_{b} = \frac{1}{2}\sum_{i,j}\left\|\mathbf{z}_i - \mathbf{z}_j\right\|^2 W_{b,ij} = \operatorname{tr}\left(\mathbf{Z}^T \mathbf{L}_{b} \mathbf{Z}\right). \tag{7}$$

Similarly, in Eq. (7), $\mathbf{L}_b = \mathbf{D}_b - \mathbf{W}_b$ is the Laplacian of G_b, and $\mathbf{D}_b$ is the diagonal matrix, whose diagonal element $D_{b,ii} = \sum_j W_{b,ij}$. $W_{b,ij} \in \mathbf{W_b}$ represents the similarity between samples x_i and x_j. To maximize the local margin between heterogeneous samples for guiding the embedding process, we define the calculation of $W_{b,ij}$ in the following way, which is served as the adjacent weight of edges for the graph construction.

$$W_{b,ij} = \begin{cases} 1 & \mathbf{x}_j \in \mathscr{M}_{b}(\mathbf{x}_i) \text{or} \mathbf{x}_i \in \mathscr{M}_{b}\left(\mathbf{x}_j\right) \\ 0 & \text{otherwise} \end{cases}, \tag{8}$$

where $\mathscr{M}_b(\mathbf{x}_i) = \left\{\mathbf{x}_j \middle| y_j \neq y_i, s(\mathbf{x}_i, \mathbf{x}_j) > \overline{W}(\mathbf{x}_i)\right\}$. By this way, the pattern discrimination prepared for distinguishing between positive and negative samples is strengthened to increase the separability.

3.1.3. The objective function of graph-CL

As the above statement, the objective function of Eq. (1) can be clearly defined for hospital readmission prediction. For incorporating the constructed graphs into a unified framework for obtaining the embedding $\mathbf{Z}$, we use the scale constraints $\mathbf{Z}^T\mathbf{D}_{wp}\mathbf{Z} = \mathbf{I}$ and $\mathbf{Z}^T\mathbf{D}_{wn}\mathbf{Z} = \mathbf{I}$ to combine Eqs. (2), (5) and (7), in which $\mathbf{I} \in \mathbb{R}^{N\times N}$ is the identity matrix. In this way, we can avoid degenerated solutions (Wang & Sun, 2015). Thus, the final objective function of Eq. (1) can be rewritten as follow:

$$\underset{\mathbf{Z}}{\operatorname{argmax}} \quad \operatorname{tr}\left(\mathbf{Z}^T\mathbf{W}_{wp}\mathbf{Z}\right) + \alpha\operatorname{tr}\left(\mathbf{Z}^T\mathbf{W}_{wn}\mathbf{Z}\right) + \beta\operatorname{tr}\left(\mathbf{Z}^T\mathbf{L}_b\mathbf{Z}\right) \tag{9}$$

$$\text{s.t.}\mathbf{Z}^T\mathbf{D}_{wp}\mathbf{Z} = \mathbf{I}, \mathbf{Z}^T\mathbf{D}_{wn}\mathbf{Z} = \mathbf{I}.$$

Our proposed optimization framework aims to map patients data into an embedding space, so powerful predictive information can be well captured and makes readmission data separable. In the other words, the generated representation of patients data allows the classifier to easily find the boundary between different classes, hence test data belonging to the minority class can be found out with effect.

3.2. Solution

In this Section, we give the solution for solving the optimization problem, as shown in Eq. (9), from which we can see that the problem is a constrained quadratic programming process, and it can be solved efficiently. Specially, we use the Lagrangian multiplier method to obtain the optimal solution $\mathbf{Z}$, and define the Lagrangian of Eq. (9) as follow:

$$\begin{aligned} F(\mathbf{Z},\Lambda) = \left\{\operatorname{tr}\left(\mathbf{Z}^T\mathbf{W}_{wp}\mathbf{Z}\right) + \alpha\operatorname{tr}\left(\mathbf{Z}^T\mathbf{W}_{wn}\mathbf{Z}\right) + \beta\operatorname{tr}\left(\mathbf{Z}^T\mathbf{L}_b\mathbf{Z}\right)\right\} + \Lambda\left\{\left(\mathbf{Z}^T\mathbf{D}_{wp}\mathbf{Z}\right.\right. \\ \left.\left. - \mathbf{I}\right) + \left(\mathbf{Z}^T\mathbf{D}_{wn}\mathbf{Z} - \mathbf{I}\right)\right\}. \end{aligned} \tag{10}$$

In Eq. (10), Λ denotes the Lagrange multiplier. Let the partial derivative of the Lagrangian be equal to zero with respect to $\mathbf{Z}$, and we can get the following formula:

$$\frac{\partial F(\mathbf{Z})}{\partial \mathbf{Z}} = 2\left(\mathbf{W}_{wp} + \alpha\mathbf{W}_{wn} + \beta\mathbf{L}_b\right)\mathbf{Z} + 2\left(\mathbf{D}_{wp} + \mathbf{D}_{wn}\right)\mathbf{Z}\Lambda = 0. \tag{11}$$

Previous work has shown that the embedding matrix can be provided by an matrix of eigenvectors (Raducanu & Dornaika, 2012). To achieve this purpose, the search of the solution is transformed into a generalized eigenvectors problem, and we have:

$$\mathbf{AZ} = \mathbf{BZ}\Lambda, \tag{12}$$

where $\mathbf{A} = \mathbf{W}_{wp} + \alpha\mathbf{W}_{wn} + \beta\mathbf{L}_b, \mathbf{B} = -(\mathbf{D}_{wp} + \mathbf{D}_{wn})$. According to Eq. (12), $\mathbf{Z}$ is solvable with generalized eigenvalue decomposition (Gou et al., 2020). Specially, Λ can be represented by the eigenvalues of matrix $\mathbf{B}^{-1}\mathbf{A}$. As Λ is available, we can obtain the corresponding eigenvectors for generating $\mathbf{Z}$, i.e., the embedding for patients data.

3.3. Neural network model

Based on the proposed Graph-CL learning framework, such an embedding, i.e., $\mathbf{Z}$, is prepared for the mining of class-imbalance patients data. Further, we fed the embedding into a powerful classifier for achieving the task of readmission prediction. In consideration of that neural network based method is effective for the performance improvement, and has been used in many real applications successfully (Dumpala et al., 2018; Castro & Braga, 2013; Manju, Harish, & Nagadarshan, 2020), we simply implement it into the proposed learning framework.

Lin and Jegelka (2018) has suggested that the situations in which performance improves with a two or more hidden layer are very few, and one hidden layer is sufficient for the large majority of problems. Therefore, we design a feed-forward architecture, which is the neural network including three layers, i.e., one input layer, one output layer, and one hidden layer. Formally, the input unit is connected with the hidden layer unit $\boldsymbol{h}$, and it consists of $|Z|$ neurons, where $|\cdot|$ denotes the number of features. Afterwards, hidden layer has 100 neurons, by using weight $\boldsymbol{W}^{(1)}$ and bias term $\boldsymbol{b}^{(1)}$. Meanwhile, the hidden layer unit $\boldsymbol{h}$ is connected with the output layer unit $\boldsymbol{o}$ via weight $\boldsymbol{W}^{(2)}$ and bias term $\boldsymbol{b}^{(2)}$. The final output layer consists of 2 neurons as there are two classes. Based on this setting, the neural network can be described in a more concise form. Explanatorily, we construct a feed-forward neural network $f_\theta : \mathbf{Z}\rightarrow\boldsymbol{o}$, which is mixed with nonlinear functions, and the output varies from 0 to 1, as follow:

$$f_\theta(\mathbf{Z}) = f_o\left(\boldsymbol{W}^{(2)}f_h\left(\boldsymbol{W}^{(1)}\mathbf{Z} + \boldsymbol{b}^{(1)}\right) + \boldsymbol{b}^{(2)}\right), \tag{13}$$

where $\theta = \left\{\boldsymbol{W}^{(1)}, \boldsymbol{b}^{(1)}, \boldsymbol{W}^{(2)}, \boldsymbol{b}^{(2)}\right\}$, and f_o and f_h denote the activation functions of the hidden layer and the output layer, respectively.

For simplicity, f_θ can be further rewritten as follow:

$$s^{(1)} = \boldsymbol{W}^{(1)}\mathbf{Z} + \boldsymbol{b}^{(1)}, \boldsymbol{h} = f_h\left(s^{(1)}\right), \tag{14}$$

$$s^{(2)} = \boldsymbol{W}^{(2)}\boldsymbol{h} + \boldsymbol{b}^{(2)}, \boldsymbol{o} = f_o\left(s^{(2)}\right), \tag{15}$$

where $s^{(1)}$ and $s^{(2)}$ denote the sum of weights in terms of the input layer and the out layer, respectively. Considering that the Rectified Linear Unit (*ReLU*) can alleviate the problem of gradient dissipation, also help to accelerate the speed of training compared with some other activation functions, such as *sigmoid* and *tanh* (Goodfellow, Bengio, & Courville, 2016; Tsuchida, Roosta, & Gallagher, 2019). Thus, *ReLU* is employed to define the activation functions of the hidden layers, i.e., f_h.

$$f_h\left(s^{(1)}\right) = \max\left(0, s^{(1)}\right), \tag{16}$$

We use the *softmax* function as the loss function. Thus, the output layer can achieve the probabilistic score.

$$f_o\left(s^{(2)}\right) = softmax\left(s^{(2)}\right). \tag{17}$$

In addition, we use *Adam* optimization algorithm to search for the parameters, thus minimizing the loss between the generated output and the value of ground-truth label. To avoid overfitting, *dropout* is adopted in our network architecture to randomly disconnect nodes between layers.

In brief, we introduced a novel graph-based class-imbalance learning method for hospital readmission. The proposed method use of within-class closeness and between-class scatterness to obtain discriminative embedding for the data samples, which will be then fed into a simple MLP with three layers. This algorithm exploring local neighborhood discriminative structures to alleviate the class imbalance problem.

4. Experimental study

4.1. Datasets

We use six readmission datasets to conduct our comparison experiments. The detailed description of six hospital readmission datasets is demonstrated in Table 1.

Table 1
Characteristics of real-world datasets.

Data	#Feature	#Instance	#(+) Instance	#(-) Instance
All-cause	15	930	109	821
Lace-score	75	1893	193	1695
MIMIC	41	4778	598	4180
RA	30	7000	1,140	5860
T-carer	66	1021	48	972
Diabetic	28	71,515	6293	65,222

All-cause[1] contains 930 discharges electronic health record

[1] http://europepmc.org/articles/PMC5018668/bin/peerj-04-2441-s001.csv

collected from the SIU-SOM Hospitalist service, which collection period is from October 15, 2015 to March 16, 2016 (Robinson, 2016). The patients features we used are including age, sex, hospital scores and other information. This dataset contains 109 positive instances and 821 negative instances, and the imbalance ratio is 7.53.

Lace-score[2] contains 1893 samples and 75 features collected by a senior data engineer at Gray Matter Analytics. The patients feature we used in the experimental including race, gender, age, length of stay, visit times, diagnosis code, history information and other lab test results. In addition, the dataset contains 193 positive instances and 1,695 negative instances, and the imbalance ratio is 8.78.

MIMIC[3] is extracted from Beth Israel Deaconess Medical Center between 2001 and 2012 (Johnson et al., 2016). It contains physiological signals, various measurements captured from patient monitors, and comprehensive clinical data are obtained from hospital medical information systems. This dataset includes 41 features, and 4778 instances (598 positive instances and 4180 negative instances), and the imbalance ratio is 6.99.

RA[4] contains 7000 samples and 30 features, which is collected by a researcher from Columbia University. The dataset contains 1140 positive instances and 5860 negative instances, and the imbalance ratio is 5.14.

T-carer[5] contains 4998 samples and 243 features collected by a research team from Columbia University. This dataset contains the records from April 1999 to March 2009. It has 367 positive instances and 4631 negative instances.

Diabetic[6] is extracted from a national data warehouse that stores comprehensive clinical data across the United States hospitals (Strack et al., 2014). Specifically, the dataset includes encounter data, demographics, laboratory data, pharmacy data, in-hospital mortality and hospital characteristics. Moreover, this dataset has 28 features and 71,515 instances, and only 6293 instances are with the positive label.

4.2. Comparative studies

To demonstrate the advantage of our proposed algorithm, we select the following state-of-art class-imbalance learning methods as comparing methods:

s2s-MLP (Dumpala et al., 2018): Recently class-imbalance learning method using data representation to increases the number of training instances.

TU (Peng et al., 2019): Recently superior under-sampling method, which incorporates the optimization of evaluation metric into the data sampling procedure.

CSMLP (Castro & Braga, 2013): Promising cost-sensitive algorithm with MLP neural networks. CSMLP uses a single cost parameter to distinguish the importance of class errors.

EasyEnsemble (Liu et al., 2008): Class-imbalance learning method with an ensemble, which prevents the abandonment of important samples belonging to majority class by under-sampling.

RUSBoost (Seiffert et al., 2009): Boosting based class-imbalance learning method, which achieves imbalanced classification by random sampling, i.e., this method considers to delete the majority instance during each iteration.

RBBOOST (Díez-Pastor et al., 2015): Ensemble learning method capable of handling imbalance data, which employs random class proportions and instance re-weighting to balance the data.

Among these comparing methods, s2s-MLP and CSMLP are neural network based class-imbalance learning models. TU is a recent study to achieve class-imbalance leaning by data sampling strategy. EasyEnsemble, RUSBoost and RBBOOST are three representative ensemble learning methods, which are widely used as comparing methods for imbalanced classification performance evaluation. For the proposed method, the number of neighbors k for graph construction is set to 5, parameters α and β are searched in $\{0.1, 0.2, \ldots, 0.9\}$, and parameter γ is designed to identify the number of eigenvectors obtained from the embedding $\mathbf{Z}$ and it is searched in $\{10\%, 20\%, \ldots, 90\%\}$. For running the proposed classification model, we use 100 *batch size*, the *learning rate* is set to 0.3, and the *dropout rate* is 0.5. For the selected comparing methods, the parameter of each algorithm is set according to the corresponding reference suggested. We tune the parameters for each method, and report the best result on test data.

We implement 5-fold cross validation to conduct the experiment. Specifically, we randomly divide each dataset into 5 uniform folds, each fold is held-out in turn for test, while the remaining data is merged for training. The process is repeated 5 times until all folds have been used as test data. Finally, we obtain 5 different results to calculate the average result with standard deviation (Zhang et al., 2020; Dai et al., 2020).

Table 2
Confusion matrix.

	Actual Value	
Predicted Value	True Positive (TP) False Negative (FN)	False Positive (FP) True Negative (TN)

4.3. Evaluation metrics

For method comparison, the area under ROC curve (AUC), F1, Geometric mean (GM) and Matthews correlation coefficient (MCC) are employed as evaluation metrics, which are widely used to evaluate the performance of class-imbalance learning (Zhang, Li, Lin, Shao, & Li, 2017; Li et al., 2020). In order to give a clear definition for these metrics, we define a confusion matrix as shown in Table 2.

AUC evaluates the variation between true positive rate (TPR) and false positive rate (FPR). We obtain the ROC curve (Fawcett, 2006) by calculating TPR and FPR, in which $TPR = TP/(TP+FN)$ and $FPR = FP/(FP+TN)$ respectively, and then we can estimate the area under ROC curve as the final result.

F1 can be formulated with the computed results of Precision and Recall, as follow:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{18}$$

where $Precision = TP/(TP + FP)$, and $Recall = TP/(TP + FN)$.

G-mean evaluates the average sensitivity of all classes, especially reflects the degree of bias in minority classes, as follow:

$$GM = \sqrt{(TP/(TP + FN)) \times (TN/(TN + FP))}. \tag{19}$$

MCC is used to evaluate the quality of binary classification, which takes the balance ratios of the four confusion matrix categories into consideration, as follow:

$$\mathrm{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \tag{20}$$

The above evaluation metrics can appropriately evaluate the performance of imbalanced classification from different perspectives. For all the evaluation metrics, the larger the values the better the performance.

4.4. Method comparison

Hospital readmission prediction is essentially a class-imbalance

[2] https://github.com/ebaumeis/lace-score/tree/master/data
[3] https://mimic.mit.edu/
[4] https://github.com/shinnknight/Readmission-Analysis
[5] https://github.com/mesgarpour/T-CARER
[6] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3996476/bin/781670.f1.zip

Table 3
The comparative result (mean ± std. Deviation) with respect to AUC.

Method	All-cause	Lace-score	MIMIC	RA	T-carer	Diabetic
s2s-MLP	0.754 ± 0.056	0.631 ± 0.030	0.846 ± 0.015	0.826 ± 0.009	0.512 ± 0.037	0.636 ± 0.041
TU	0.781 ± 0.040	0.787 ± 0.040	0.858 ± 0.024	0.842 ± 0.008	0.617 ± 0.053	0.633 ± 0.038
CSMLP	0.607 ± 0.045	0.611 ± 0.098	0.779 ± 0.020	**0.897 ± 0.010**	0.506 ± 0.019	0.548 ± 0.011
EasyEnsemble	0.722 ± 0.033	0.747 ± 0.021	0.755 ± 0.011	0.852 ± 0.010	0.579 ± 0.052	0.632 ± 0.004
RUSBoost	0.685 ± 0.049	0.649 ± 0.024	0.733 ± 0.015	0.821 ± 0.007	0.529 ± 0.074	0.626 ± 0.003
RBBoost	0.679 ± 0.032	0.662 ± 0.019	0.741 ± 0.013	0.822 ± 0.011	0.531 ± 0.040	0.615 ± 0.041
Graph-CL	**0.812 ± 0.039**	**0.791 ± 0.071**	**0.865 ± 0.015**	0.886 ± 0.035	**0.776 ± 0.026**	**0.785 ± 0.056**

Table 4
The comparative result (mean ± std. Deviation) with respect to F1.

Method	All-cause	Lace-score	MIMIC	RA	T-carer	Diabetic
s2s-MLP	0.712 ± 0.009	0.736 ± 0.031	0.612 ± 0.020	0.822 ± 0.015	0.814 ± 0.030	0.619 ± 0.107
TU	0.772 ± 0.050	0.739 ± 0.028	0.618 ± 0.026	0.845 ± 0.011	0.859 ± 0.045	0.621 ± 0.024
CSMLP	0.625 ± 0.058	0.703 ± 0.035	0.564 ± 0.047	0.818 ± 0.020	0.715 ± 0.051	0.519 ± 0.040
EasyEnsemble	0.746 ± 0.052	0.697 ± 0.017	0.617 ± 0.017	0.853 ± 0.015	0.848 ± 0.027	0.580 ± 0.005
RUSBoost	0.762 ± 0.062	0.695 ± 0.025	0.614 ± 0.024	0.812 ± 0.018	0.853 ± 0.030	0.573 ± 0.006
RBBoost	0.757 ± 0.047	0.704 ± 0.021	0.615 ± 0.013	0.815 ± 0.020	0.851 ± 0.027	0.566 ± 0.030
Graph-CL	**0.832 ± 0.037**	**0.742 ± 0.119**	**0.627 ± 0.020**	**0.861 ± 0.076**	**0.861 ± 0.076**	**0.632 ± 0.044**

Table 5
The comparative result (mean ± std. Deviation) with respect to GM.

Method	All-cause	Lace-score	MIMIC	RA	T-carer	Diabetic
s2s-MLP	0.722 ± 0.089	0.523 ± 0.034	0.644 ± 0.016	0.825 ± 0.009	0.385 ± 0.085	0.630 ± 0.157
TU	0.719 ± 0.046	0.585 ± 0.001	0.653 ± 0.037	0.861 ± 0.008	0.572 ± 0.153	0.626 ± 0.179
CSMLP	0.535 ± 0.049	0.693 ± 0.078	0.646 ± 0.049	0.843 ± 0.009	0.509 ± 0.134	0.354 ± 0.060
EasyEnsemble	0.718 ± 0.031	0.546 ± 0.041	0.654 ± 0.041	0.833 ± 0.032	0.483 ± 0.059	0.622 ± 0.011
RUSBoost	0.668 ± 0.067	0.540 ± 0.029	0.630 ± 0.004	0.817 ± 0.010	0.423 ± 0.129	0.617 ± 0.024
RBBoost	0.668 ± 0.040	0.558 ± 0.021	0.639 ± 0.008	0.809 ± 0.005	0.442 ± 0.096	0.601 ± 0.053
Graph-CL	**0.868 ± 0.042**	**0.736 ± 0.110**	**0.765 ± 0.016**	**0.901 ± 0.009**	**0.775 ± 0.054**	**0.763 ± 0.073**

learning problem. To verify the effectiveness of our proposed method on imbalanced classification, we compare with six class-imbalance learning algorithms. Tables 3–6 summarize the experimental result in terms of AUC, F1, GM and MCC, respectively. In these tables, the best performance among all the algorithms is highlighted in boldface.

From Tables 3–6, we have a couple of observations. (1) The proposed method outperforms the selected comparing algorithms in most of the cases. For example, the proposed method achieves a good performance on the All-cause dataset, where the average AUC, F1, GM and MCC are 0.812, 0.832, 0.868 and 0.814, respectively. Moreover, a similar phenomenon occurs on some other datasets, such as Lace-score, MIMIC, T-carer and Diabetic. This indicates that the proposed method is a competitive algorithm on readmission prediction task, which has the advantage on dealing with class-imbalance data. (2) The experimental result in terms of AUC reveals that our proposed method is better than the other comparing methods on 5 out of 6 datasets, while on the RA dataset, CSMLP can obtain the best performance, but it has the unsatisfactory result on the other five hospital readmission datasets. (3) From the respective of algorithm level in dealing with the class-imbalance patients data, the proposed method compares favorably with the neural network models like s2s-MLP and CSMLP, also has the advantage compared with some ensemble learning methods (e.g., EasyEnsemble, RUSBoost and RBBOOST) and data sampling methods (e.g., TU). Based on this, we can come to the conclusion that our proposed method benefits to the performance on class-imbalance learning, and can achieve the better result in readmission prediction compared with some other well-established methods.

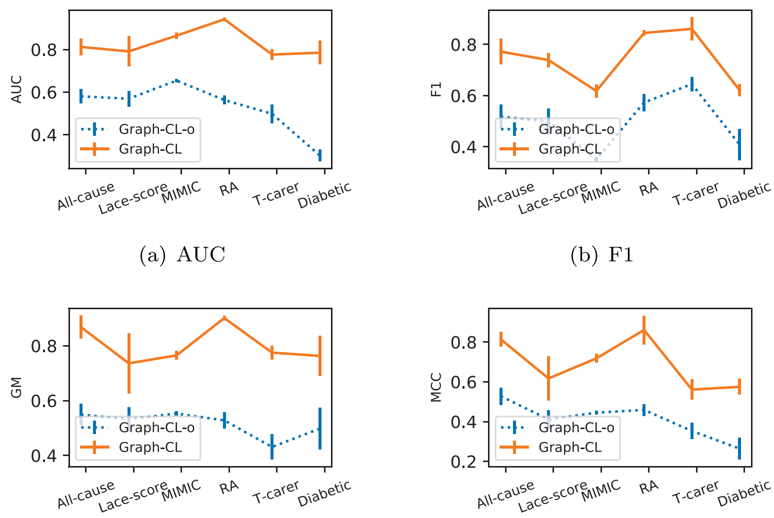
To further analyze the performance among all the algorithms, two popular statistical tests, i.e., Welch's t test and Mann–Whitney U test (Corder & Foreman, 2011) as the favorable statistical significance tests, are employed for the method comparison on the six datasets. The statistical tests with a significance level at 0.05 are given in Table 7. As described in Table 7, the values of significance are lower than 0.05. Based on the experimental result, the following observations can be

Table 6
The comparative result (mean ± std. Deviation) with respect to MCC.

Method	All-cause	Lace-score	MIMIC	RA	T-carer	Diabetic
s2s-MLP	0.715 ± 0.089	0.539 ± 0.038	0.704 ± 0.023	0.751 ± 0.018	0.525 ± 0.040	0.523 ± 0.053
TU	0.708 ± 0.058	0.605 ± 0.051	0.627 ± 0.031	0.779 ± 0.013	0.529 ± 0.060	0.515 ± 0.071
CSMLP	0.639 ± 0.069	0.444 ± 0.043	0.598 ± 0.043	0.803 ± 0.024	0.473 ± 0.059	0.464 ± 0.021
EasyEnsemble	0.686 ± 0.046	0.555 ± 0.025	0.508 ± 0.017	0.791 ± 0.018	0.466 ± 0.046	0.517 ± 0.007
RUSBoost	0.676 ± 0.066	0.560 ± 0.029	0.588 ± 0.022	0.748 ± 0.018	0.523 ± 0.056	0.407 ± 0.006
RBBoost	0.663 ± 0.049	0.576 ± 0.024	0.595 ± 0.016	0.742 ± 0.022	0.531 ± 0.038	0.489 ± 0.067
Graph-CL	**0.814 ± 0.037**	**0.617 ± 0.111**	**0.718 ± 0.022**	**0.861 ± 0.072**	**0.561 ± 0.051**	**0.575 ± 0.040**

Table 7
Summary of the Welch's *t* test and the Mann–Whitney U test while significance level is 0.05.

Method	Welch's *t* test				Mann–Whitney *U* test			
	AUC	F1	GM	MCC	AUC	F1	GM	MCC
s2s-MLP	4.348e−06	3.211e−06	2.609e−07	3.083e−04	2.999e−05	3.353e−06	1.799e−05	2.166e−06
TU	1.191e−05	6.127e−04	1.499e−06	1.916e−06	3.397e−06	1.250e−06	2.609e−05	6.660e−04
CSMLP	1.002e−06	2.317e−06	1.088e−05	1.649e−04	7.037e−06	3.833e−06	3.066e−06	2.226e−06
EasyEnsemble	5.864e−06	3.387e−06	1.813e−06	4.990e−06	1.034e−06	3.833e−06	1.312e−04	1.341e−06
RUSBoost	1.574e−05	3.644e−06	1.920e−06	4.372e−06	2.624e−05	1.271e−06	5.551e−06	2.465e−06
RBBoost	4.372e−06	2.793e−06	1.166e−05	4.231e−06	2.186e−06	2.637e−06	1.495e−05	2.166e−06



(a) AUC (b) F1

(c) GM (d) MCC

Fig. 2. Comparison of the proposed method and Graph-CL-o on AUC, F1, GM, and MCC.

Table 8
Experimental results of Graph-CL and its variant in terms of each evaluation metric on six readmission datasets.

Dataset	Method	AUC	F1	GM	MCC
All-cause	Graph-CL	0.812 ± 0.039	0.832 ± 0.037	0.868 ± 0.042	0.814 ± 0.037
	w/o-Graph-CL	0.607 ± 0.026	0.312 ± 0.047	0.494 ± 0.057	0.261 ± 0.040
Lace-score	Graph-CL	0.791 ± 0.071	0.742 ± 0.119	0.736 ± 0.110	0.617 ± 0.111
	w/o-Graph-CL	0.512 ± 0.013	0.102 ± 0.027	0.282 ± 0.042	0.028 ± 0.029
MIMIC	Graph-CL	0.865 ± 0.015	0.627 ± 0.020	0.765 ± 0.016	0.718 ± 0.022
	w/o-Graph-CL	0.595 ± 0.018	0.305 ± 0.033	0.500 ± 0.040	0.209 ± 0.029
RA	Graph-CL	0.886 ± 0.035	0.861 ± 0.076	0.901 ± 0.009	0.861 ± 0.072
	w/o-Graph-CL	0.643 ± 0.008	0.508 ± 0.011	0.742 ± 0.008	0.392 ± 0.013
T-carer	Graph-CL	0.776 ± 0.026	0.861 ± 0.076	0.775 ± 0.054	0.561 ± 0.051
	w/o-Graph-CL	0.555 ± 0.025	0.182 ± 0.057	0.365 ± 0.080	0.145 ± 0.048
Diabetic	Graph-CL	0.785 ± 0.056	0.632 ± 0.044	0.763 ± 0.073	0.575 ± 0.040
	w/o-Graph-CL	0.505 ± 0.015	0.018 ± 0.058	0.033 ± 0.104	0.021 ± 0.066

made: (1) Our proposed method is significantly superior to the selected comparing algorithms on the Welch's t test. (2) The experimental result also shows a significant difference between the proposed method and all the comparing methods among all the evaluation metrics on the Mann–Whitney U test. In summary, the result does display the effectiveness of our proposed method, and it helps to further improve the classification performance for predicting hospital readmission.

4.5. Influence of graph embedding

Our work focuses on designing a graph-based learning method to obtain an embedding for solving the class-imbalance problem. For revealing the influence of graph embedding, we run a variant of the proposed method without considering local structures of positive and negative data samples respectively. Instead, the variant utilizes the global structure of all data samples to achieve globality preserving graph embedding, which can be viewed as a degenerated version of the proposed method, and we call it as Graph-CL-o.

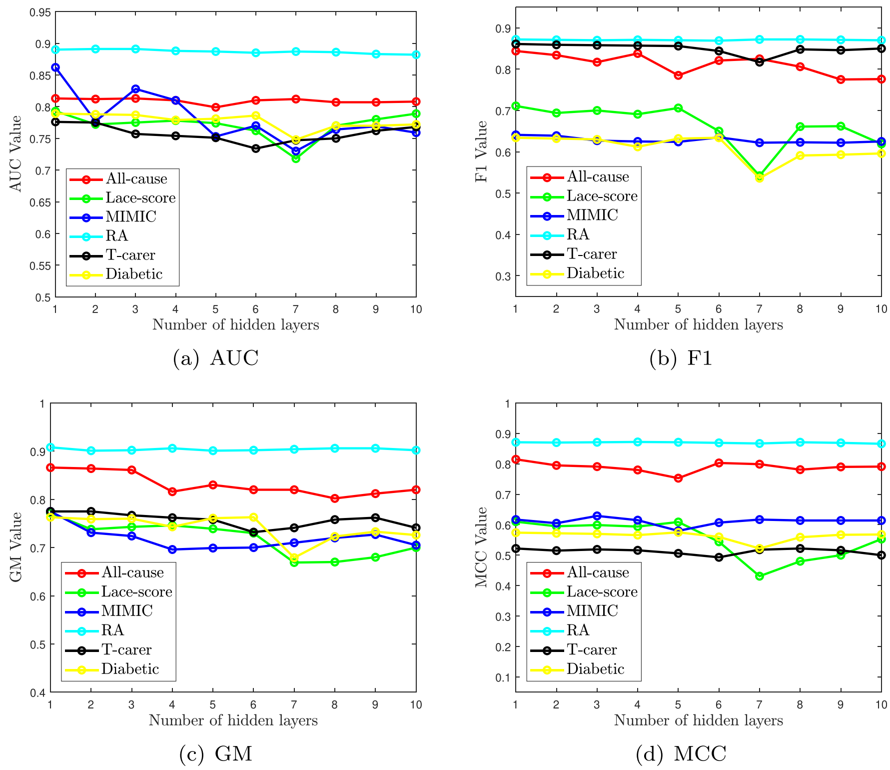


Fig. 3. Difference evaluation metric value versus the number of hidden layers for neural network model.

Figs. 2(a)–(d) show the comparison result of the proposed method with Graph-CL-o in terms of AUC, F1, GM and MCC, respectively. From Fig. 2(a), we can observe that the six hospital readmission datasets in the experiments are in favor of our proposed method on AUC, and the experimental result from Figs. 2(b)–(d) also shows that our proposed method achieves a better performance than Graph-CL-o on the other metrics. Thus, we can obtain that the proposed method, which uses local discrimination information of positive and negative data samples to conduct graph embedding, contributes to the prediction for the minority class, thereby achieving class-imbalance learning for hospital readmission.

Further, a same network without pre-processing method is designed to verify the effectiveness of our proposed method, which called w/o-Graph-CL. This method is implemented using the same setting as Graph-CL. Table 8 shows the comparison results of the proposed method with w/o-Graph-CL in terms of six readmission datasets, respectively.

We can see from Table 8 that the proposed method can always achieve superior performance on each dataset. Thus, we can conclude that the proposed graph-based learning method can obtain an effective embedding for solving the class-imbalance problem.

4.6. Influence of neural network

To study the efficiency of neural network, we conduct a series of experiments that assess the classification performance by using more deep hidden layers. Moreover, we present a detailed report of experimental results in Fig. 3.

In Fig. 3, the classification performance of difference evaluation metrics are displayed. For AUC in Fig. 3(a), we can observe that the higher AUC values are often achieved when the neural network model with less hidden layers. Notably, this change is not always monotonous with the varying of hidden layer number, and the performance may keep stable, or even worsen, as the number of hidden layers increases. Since it is easier to overfit the training data when we use more deep hidden layers. Based on our discussion above, it seems that smaller neural networks can be preferred if the data is not complex enough to prevent overfitting. Meanwhile, we can also find the similar phenomenon on F1, GM and MCC. Thus, these observations authentically and reliably confirm the effectiveness of the neural network. Unlike other application domains such computer vision and natural language processing, the deeper models have shown to be very powerful, medical problem are much more complicated (Min, Yu, & Wang, 2019; Rajkomar et al., 2018). Our studies verify the importance of careful design a suitable learning method can effectively improve the performance.

As further proof the the efficiency of the neural network, we apply multiple base classifier to compare in our experiments, such as k-nearest neighbor(kNN), decision tree (DT), support vector machines (SVM) and random forest (RF). These methods use the data representation, generated by our proposed method, as input, and we call these method as Graph-CL-KNN, Graph-CL-DT, Graph-CL-SVM and Graph-CL-RF.

The experimental results of these methods are presented in Table 9. It can be observed that the proposed algorithm performs much better than these compared methods. This is because the neural network model is a powerful data-driven, self-adaptive, flexible classification approach having the capability of capturing nonlinear and complex underlying characteristics of any physical process with a high degree of accuracy.

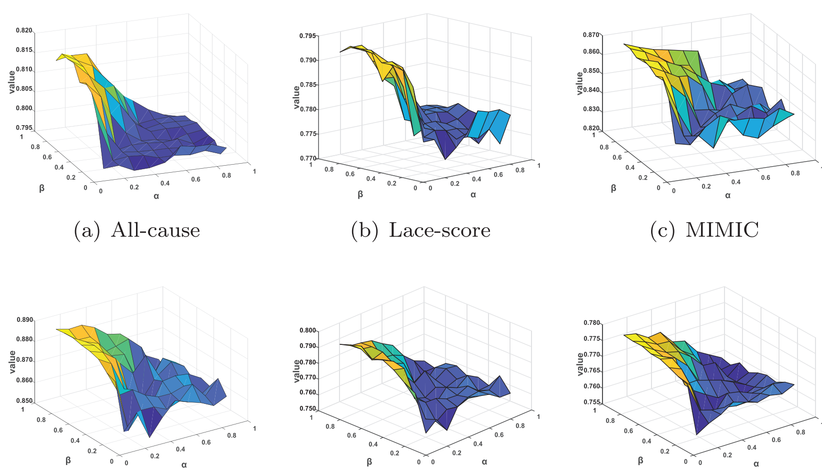
4.7. Parameter sensitivity analysis

Three parameters are involved in the experiments. Among these parameters, α is the parameter to balance the influence of the graph embedding using negative data samples, β is to balance the influence of between-class graph embedding, and γ is used to denote the percent of eigenvectors for generating an embedding.

We conduct the experiments on the six hospital readmission datasets, and vary the value of one parameter while the other two parameters are fixed with the best value. The experiment result is illustrated in Fig. 4 and Fig. 5. Fig. 4(a) shows the performance of AUC on the All-cause dataset by using different pairs of α and β. From Fig. 4(a), we observe that AUC is going to change dramatically in some cases. Thus, the

Table 9
Experimental results of different base classifiers on six readmission datasets.

Dataset	Metrics	Graph-CL	Graph-CL-KNN	Graph-CL-DT	Graph-CL-SVM	Graph-CL-RF
All-cause	AUC	0.812 ± 0.039	0.778 ± 0.014	0.802 ± 0.035	0.794 ± 0.027	0.807 ± 0.018
	F1	0.832 ± 0.037	0.400 ± 0.031	0.626 ± 0.055	0.624 ± 0.031	0.546 ± 0.042
	GM	0.868 ± 0.042	0.768 ± 0.016	0.789 ± 0.041	0.778 ± 0.036	0.803 ± 0.020
	MCC	0.814 ± 0.037	0.364 ± 0.027	0.578 ± 0.067	0.575 ± 0.033	0.495 ± 0.038
Lace-score	AUC	0.791 ± 0.071	0.595 ± 0.018	0.774 ± 0.027	0.783 ± 0.021	0.702 ± 0.031
	F1	0.742 ± 0.119	0.313 ± 0.046	0.561 ± 0.133	0.525 ± 0.054	0.480 ± 0.085
	GM	0.774 ± 0.022	0.442 ± 0.039	0.757 ± 0.034	0.736 ± 0.110	0.660 ± 0.033
	MCC	0.617 ± 0.111	0.358 ± 0.042	0.529 ± 0.145	0.473 ± 0.058	0.429 ± 0.111
MIMIC	AUC	0.865 ± 0.015	0.635 ± 0.003	0.738 ± 0.011	0.721 ± 0.018	0.751 ± 0.010
	F1	0.627 ± 0.020	0.282 ± 0.006	0.418 ± 0.026	0.570 ± 0.028	0.392 ± 0.007
	GM	0.765 ± 0.016	0.536 ± 0.006	0.734 ± 0.012	0.673 ± 0.027	0.744 ± 0.009
	MCC	0.718 ± 0.022	0.203 ± 0.005	0.344 ± 0.017	0.543 ± 0.024	0.335 ± 0.011
RA	AUC	0.886 ± 0.035	0.584 ± 0.008	0.785 ± 0.010	0.820 ± 0.006	0.819 ± 0.004
	F1	0.861 ± 0.076	0.317 ± 0.007	0.670 ± 0.010	0.579 ± 0.017	0.600 ± 0.010
	GM	0.901 ± 0.009	0.425 ± 0.022	0.765 ± 0.013	0.818 ± 0.006	0.819 ± 0.004
	MCC	0.861 ± 0.072	0.170 ± 0.010	0.619 ± 0.010	0.508 ± 0.017	0.526 ± 0.010
T-carer	AUC	0.776 ± 0.026	0.718 ± 0.096	0.750 ± 0.060	0.775 ± 0.048	0.745 ± 0.040
	F1	0.861 ± 0.076	0.349 ± 0.090	0.367 ± 0.061	0.404 ± 0.053	0.464 ± 0.062
	GM	0.775 ± 0.054	0.662 ± 0.155	0.723 ± 0.081	0.757 ± 0.060	0.708 ± 0.060
	MCC	0.561 ± 0.051	0.340 ± 0.095	0.355 ± 0.070	0.394 ± 0.052	0.448 ± 0.064
Diabetic	AUC	0.785 ± 0.056	0.696 ± 0.016	0.719 ± 0.017	0.734 ± 0.010	0.748 ± 0.007
	F1	0.632 ± 0.044	0.480 ± 0.031	0.593 ± 0.030	0.616 ± 0.016	0.585 ± 0.013
	GM	0.763 ± 0.073	0.680 ± 0.014	0.667 ± 0.025	0.692 ± 0.015	0.730 ± 0.007
	MCC	0.575 ± 0.040	0.356 ± 0.044	0.476 ± 0.026	0.486 ± 0.014	0.499 ± 0.018



(a) All-cause (b) Lace-score (c) MIMIC

(d) RA (e) T-carer (f) Diabetic

Fig. 4. The result of AUC when different pairs of α and β are employed.

proposed method is sensitive to α and β. Fig. 5(a) shows the influence of parameter γ on the All-cause dataset, and it reflects that the performance is going to a remarkable improvement while γ becomes larger. However, the performance slowly decreases when γ large than 40%. This suggests that some intermediate values of γ help to the graph-based class-imbalance learning. In the similar way, we can find out that there exists a similar phenomenon on the other five readmission datasets.

5. Discussion and conclusions

Improving the quality of medical care is an ongoing challenge for health care organizations. Unfortunately, the cost of care is increasing at a rate that is unaffordable in the current economy (Mesgarpour et al., 2019). Thus, it is an important and practical task to improve the patient's healthcare management by mining medical knowledge in readmission data. Meanwhile, predicting hospital readmission is indeed needed as this presents a high cost for both individual and countries. However, the readmission data have the characteristic with imbalanced class distribution. Due to the problem, using traditional learning approaches is difficult to unleash their predictive power. In recent years, there are some methods which have been well designed for readmission prediction, but these methods often cannot achieve a ideal performance in this task (Ma, Zhong, Gao, & Bian, 2019).

In this paper, we have presented an interesting graph based technique for tackling the class-imbalance problem in hospital readmission. Specially, we first constructed (positive and negative) within-class graphs and between-class graph to extract the most discriminative information from class-imbalance readmission data. Then an optimization

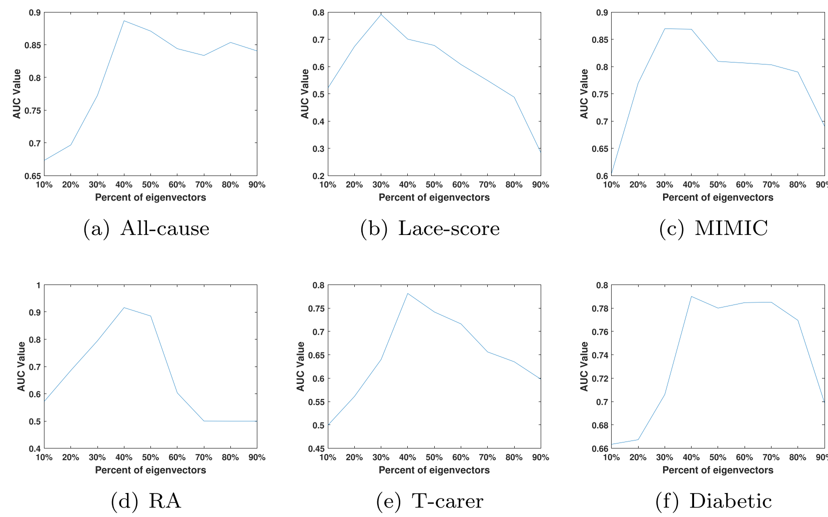


Fig. 5. The result of AUC with different percentage of eigenvectors.

framework was introduced to incorporate these graphs for obtaining an embedding, which can make positive and negative samples separable. Finally, based on the generated embedding, a fully-connected neural network was designed to achieve the prediction of hospital readmission. Extensive experiments on six real-world datasets demonstrated that our proposed method can obtain a superior performance on this task.

Although our proposed method has an excellent performance on six real-world readmission datasets, there still are some limitations which need to study further. In this paper, we choose the graph-based class-imbalance learning method to obtain a separable representation, but it is undeniable that this embedding has some defects. For example, it cannot identify the key factors in patient readmission through the learned embedding. In the other words, the proposed dimension reduction method cannot achieve the feature selection purpose, thus preserving the initial meaning of features.

In our future research, we plan to improve our proposed method to achieving hospital readmission prediction via class-imbalance learning. Note that most of the existing dimension reduction methods focus on local neighborhood discriminative structures. Thus, one direction for future research is to improve these method to further alleviate the class imbalance problem. We also intend to extend our proposed method to other healthcare scenarios, such as traditional Chinese medicine (TCM) clinical efficacy evaluation, early hospital mortality prediction and rare disease prediction.

CRediT authorship contribution statement

Guodong Du: Conceptualization, Methodology, Investigation, Writing - original draft, Software. **Jia Zhang:** Validation, Formal analysis, Visualization, Software. **Fenglong Ma:** Resources, Writing - review & editing, Data curation. **Min Zhao:** Writing - review & editing. **Yaojin Lin:** Resources, Writing - review & editing, Data curation. **Shaozi Li:** Supervision, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is supported by the National Nature Science Foundation of China (No. 61876159, No. 61806172, 62076116 & No. U1705286), Natural Science Foundation of Fujian Province, China (No. 2019J01571), Fujian Province 2011 Collaborative Innovation Center of TCM Health Management, Advanced Space Medico-Engineering Research Project of China (No. 020104), the Fundamental Research Funds for the Central Universities (Xiamen University, No. 20720200030).

References

Agarwal, A., Baechle, C., Behara, R. S., & Zhu, X. (2018). A natural language processing framework for assessing hospital readmissions for patients with COPD. *IEEE Journal of Biomedical and Health Informatics, 22*, 588–596.

Ali, A. M., Loeffler, M. D., Aylin, P., & Bottle, A. (2019). Predictors of 30-day readmission after total knee arthroplasty: Analysis of 566,323 procedures in the United Kingdom. *The Journal of Arthroplasty, 34*, 242–248.

Arancibia, J. C. N., Sanchez, F. J. M., rey Mejías, A. L. D., del Castillo, J. G., Vilaplana, J. C., Briñon, M. A. G., Suárez-Cadenas, M. M., Mayol, J., & Seara, G. (2019). Evaluation of a diagnostic decision support system for the triage of patients in a hospital emergency department. *International Journal of Interactive Multimedia and Artificial Intelligence, 5*, 60–67.

Ashfaq, A., Sant'Anna, A. P., Lingman, M., & Nowaczyk, S. (2019). Readmission prediction using deep learning on electronic health records. *Journal of Biomedical Informatics, 97*.

Braga, P., Portela, F., Santos, M. F., & Rua, F. (2014). Data mining models to predict patient's readmission in intensive care units. In *Proceedings of the 6th International Conference on Agents and Artificial Intelligence, Angers, Loire Valley, France* (pp. 604–610).

Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia* (pp. 1721–1730).

Castro, C. L., & Braga, A. P. (2013). Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data. *IEEE Transactions on Neural Networks and Learning Systems, 24*, 888–899.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357.

Choi, E., Xu, Z., Li, Y., Dusenberry, M. W., Flores, G., Xue, E., & Dai, A. M. (2020). Learning the graphical structure of electronic health records with graph convolutional transformer. In *Proceedings of the Thirty-Fourth AAAI conference on artificial intelligence, AAAI*, Hilton New York Midtown, New York, USA.

Corder, G. W., & Foreman, D. I. (2011). *Nonparametric statistics for non-statisticians*. John Wiley & Sons Inc.

Cui, S., Wang, D., Wang, Y., Yu, P., & Jin, Y. (2018). An improved support vector machine-based diabetic readmission prediction. *Computer Methods and Programs in Biomedicine, 166*, 123–135.

Dai, L., Du, G., Zhang, J., Li, C., Wei, R., & Li, S. (2020). Joint multilabel classification and feature selection based on deep canonical correlation analysis. *Concurrency and Computation: Practice and Experience, 32*, 1–13.

Díez-Pastor, J. F., Rodríguez, J. J., García-Osorio, C., & Kuncheva, L. I. (2015). Random balance: Ensembles of variable priors classifiers for imbalanced data. *Knowledge-Based Systems, 85*, 96–111.

Du, G., Ma, L., Hu, J., Zhang, J., Xiang, Y., Shao, D., et al. (2019). Prediction of 30-day readmission: An improved gradient boosting decision tree approach. *Journal of Medical Imaging and Health Informatics, 9*, 620–627.

Du, G., Zhang, J., Li, S., & Li, C. (2021). Learning from class-imbalance and heterogeneous data for 30-day hospital readmission. *Neurocomputing, 420*, 27–35.

Du, G., Zhang, J., Luo, Z., Ma, F., Ma, L., & Li, S. (2020). Joint imbalanced classification and feature selection for hospital readmissions. *Knowledge-Based Systems, 200*, Article 106020.

Duggal, R., Shukla, S., Chandra, S., Shukla, B., & Khatri, S. K. (2016). Impact of selected pre-processing techniques on prediction of risk of early readmission for diabetic patients in India. *International Journal of Diabetes in Developing Countries, 36*, 469–476.

Dumpala, S. H., Chakraborty, R., & Kopparapu, S. K. (2018). A novel data representation for effective learning in class imbalanced scenarios. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13–19, 2018, Stockholm, Sweden* (pp. 2100–2106).

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*, 861–874.

Gama, C. S., Backman, C., & de Oliveira, A. C. (2020). Impact of surgical checklist on mortality, reoperation, and readmission rates in brazil, a developing country, and canada, a developed country. *Journal of PeriAnesthesia Nursing, 35*, 508–513.e2.

Golmohammadi, D., & Radnia, N. (2016). Prediction modeling and pattern recognition for patient readmission. *International Journal of Production Economics, 171*, 151–161.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Gou, J., Yang, Y., Yi, Z., Lv, J., Mao, Q., & Zhan, Y. (2020). Discriminative globality and locality preserving graph embedding for dimensionality reduction. *Expert Systems with Applications, 144*, Article 113079.

Gu, X., Angelov, P. P., & Soares, E. A. (2020). A self-adaptive synthetic over-sampling technique for imbalanced classification. *International Journal of Intelligent Systems, 35*, 923–943.

Hosseinzadeh, A., Izadi, M. T., Verma, A., Precup, D., & Buckeridge, D. L. (2013). Assessing the predictability of hospital readmission using machine learning. In *Proceedings of the Twenty-Fifth Innovative Applications of Artificial Intelligence Conference, IAAI, Bellevue, Washington, USA*.

Jeejeebhoy, K. N., Keller, H., Gramlich, L., Allard, J. P., Laporte, M., Duerksen, D. R., Payette, H., et al. (2015). Nutritional assessment: Comparison of clinical assessment and objective variables for the prediction of length of hospital stay and readmission. *The American Journal of Clinical Nutrition, 101*, 956–965.

Jencks, S. F., Williams, M. V., & Coleman, E. A. (2009). Rehospitalizations among patients in the medicare fee-for-service program. *New England Journal of Medicine, 360*, 1418–1428.

Jiang, S., Chin, K., Qu, G., & Tsui, K. (2018). An integrated machine learning framework for hospital readmission prediction. *Knowledge-based Systems, 146*, 73–90.

Johnson, A. E., Pollard, T. J., Shen, L., Li-wei, H. L., Feng, M., Ghassemi, M., et al. (2016). Mimic-III, a freely accessible critical care database. *Scientific Data, 3*, Article 160035.

Kouyoumdjian, F. G., Lee, J. Y., Orkin, A. M., Cheng, S. Y., Fung, K., O'Shea, T., et al. (2020). Thirty-day readmission after medical-surgical hospitalization for people who experience imprisonment in Ontario, Canada: A retrospective cohort study. *PLoS one, 15*, Article e0227588.

Lee, H., Jin, R., Feng, Y., Bain, P. A., Goffinet, J., Baker, C., et al. (2019). An analytical framework for TJR readmission prediction and cost-effective intervention. *IEEE Journal of Biomedical and Health Informatics, 23*, 1760–1772.

Li, Y., Du, G., Xiang, Y., Li, S., Ma, L., Shao, D., et al. (2020). Towards Chinese clinical named entity recognition by dynamic embedding using domain-specific knowledge. *Journal of Biomedical Informatics, 106*, Article 103435.

Lin, H., & Jegelka, S. (2018). Resnet with one-neuron hidden layers is a universal approximator. In S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), it Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3–8, 2018, Montréal, Canada (pp. 6172–6181).

Liu, R., Zolfaghar, K., Chin, S., Roy, S. B., & Teredesai, A. (2014). A framework to recommend interventions for 30-day heart failure readmission risk. In *2014 IEEE International Conference on Data Mining, Shenzhen, China* (pp. 911–916).

Liu, X.-Y., Wu, J., & Zhou, Z.-H. (2008). Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Cybernetics, 39*, 539–550.

Liu, Z., Cao, W., Gao, Z., Bian, J., Chen, H., Chang, Y., & Liu, T. (2020). Self-paced ensemble for highly imbalanced massive data classification. In *36th IEEE International Conference on Data Engineering, ICDE, Dallas, TX, USA* (pp. 841–852).

Ma, F., Gao, J., Suo, Q., You, Q., Zhou, J., & Zhang, A. (2018). Risk prediction on electronic health records with prior medical knowledge. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD, London, UK* (pp. 1910–1919).

Ma, F., Wang, Y., Xiao, H., Yuan, Y., Chitta, R., Zhou, J., et al. (2019). Incorporating medical code descriptions for diagnosis prediction in healthcare. *BMC Medical Informatics and Decision Making, 19*, 267–280.

Ma, F., Zhong, S., Gao, J., & Bian, L. (2019). Influenza-like symptom prediction by analyzing self-reported health status and human mobility behaviors. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, Niagara Falls, NY, USA* (pp. 233–242).

Manju, N., Harish, B. S., & Nagadarshan, N. (2020). Multilayer feedforward neural network for internet traffic classification. *International Journal of Interactive Multimedia and Artificial Intelligence, 6*, 117–122.

Mardini, M. T., & Raś, Z. W. (2019). Extraction of actionable knowledge to reduce hospital readmissions through patients personalization. *Information Sciences, 485*, 1–17.

Mesgarpour, M., Chaussalet, T., & Chahed, S. (2019). Temporal comorbidity-adjusted risk of emergency readmission (t-carer): A tool for comorbidity risk assessment. *Applied Soft Computing, 79*, 163–185.

Min, X., Yu, B., & Wang, F. (2019). Predictive modeling of the hospital readmission risk from patients' claims data using machine learning: A case study on copd. *Scientific reports, 9*, 2362.

Peng, M., Zhang, Q., Xing, X., Gui, T., Huang, X., Jiang, Y., Ding, K., & Chen, Z. (2019). Trainable undersampling for class-imbalance learning. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27–February 1, 2019* (pp. 4707–4714).

Raducanu, B., & Dornaika, F. (2012). A supervised non-linear dimensionality reduction approach for manifold learning. *Pattern Recognition, 45*, 2432–2444.

Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., et al. (2018). Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine, 1*, 18–27.

Reischl, D., Eigner, I., Bodendorf, F., Schaffer, J. L., & Wickramasinghe, N. (2019). *Initial phase of developing a model to predict unplanned readmissions in the australian private healthcare sector before they occur: the case of hysterectomy*. Washington, DC, USA: In American Medical Informatics Association Annual Symposium.

Robinson, R. (2016). The hospital score as a predictor of 30 day readmission in a retrospective study at a university affiliated community hospital. *PeerJ, 4*, Article e2441.

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2009). Rusboost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems, Man, and Cybernetics, 40*, 185–197.

Strack, B., DeShazo, J. P., Gennings, C., Olmo, J. L., Ventura, S., Cios, K. J., & Clore, J. N. (2014). Impact of hba1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records. BioMed Research International, 2014.

Tsuchida, R., Roosta, F. F., & Gallagher, M. (2019). Exchangeability and kernel invariance in trained mlps. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI, Macao, China* (pp. 3592–3598).

Turgeman, L., & May, J. H. (2016). A mixed-ensemble model for hospital readmission. *Artificial Intelligence in Medicine, 72*, 72–82.

Tutun, S., Khanmohammadi, S., He, L., & Chou, C.-A. (2016). A meta-heuristic lasso model for diabetic readmission prediction. In *Proceedings of the 2016 Industrial and Systems Engineering Research Conference, Anaheim, California, USA* (pp. 1–6).

Wang, F. (2011). Semisupervised metric learning by maximizing constraint margin. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 41*, 931–939.

Wang, F., & Sun, J. (2015). PSF: A unified patient similarity evaluation framework through metric learning with weak supervision. *IEEE Journal of Biomedical and Health Informatics, 19*, 1053–1060.

Wang, F., & Zhang, C. (2007). Feature extraction by maximizing the average neighborhood margin. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, Minnesota, USA* (pp. 1–8).

Wang, F., & Zhang, C. (2008). Label propagation through linear neighborhoods. *IEEE Transactions on Knowledge and Data Engineering, 20*, 55–67.

Wang, H., Cui, Z., Chen, Y., Avidan, M., Abdallah, A. B., & Kronzer, A. (2018). Predicting hospital readmission via cost-sensitive deep learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 15*, 1968–1978.

Wang, S., Elkin, M. E., & Zhu, X. (2020). Imbalanced learning for hospital readmission prediction using national readmission database. In *Proceedings of the 11th IEEE International Conference on Knowledge Graph (ICBK), Nanjing, China* (pp. 116–122).

Xiao, C., Ma, T., Dieng, A. B., Blei, D. M., & Wang, F. (2018). Readmission prediction via deep contextual embedding of clinical concepts. *PloS one, 13*, Article e0195024.

Xu, Z., Feng, Y., Li, Y., Srivastava, A., Adekkanattu, P., Ancker, J. S., et al. (2019). Predictive modeling of the risk of acute kidney injury in critical care: A systematic investigation of the class imbalance problem. AMIA Summits on Translational Science Proceedings (pp. 809–818).

Xue, J., & Titterington, D. M. (2008). Do unbalanced data have a negative effect on LDA? *Pattern Recognition, 41*, 1558–1571.

Yu, K., & Xie, X. (2020). Predicting hospital readmission: A joint ensemble-learning model. *IEEE Journal of Biomedical and Health Informatics, 24*, 447–456.

Zhang, J., Li, C., Lin, Y., Shao, Y., & Li, S. (2017). Computational drug repositioning using collaborative filtering via multi-source fusion. *Expert Systems with Applications, 84*, 281–289.

Zhang, J., Li, S., Jiang, M., & Tan, K. C. (2021). Learning from weakly labeled data based on manifold regularized sparse model. IEEE Transactions on. *Cybernetics*.

Zhang, J., Lin, Y., Jiang, M., Li, S., Tang, Y., & Tan, K. C. (2020). Multi-label feature selection via global relevance and redundancy optimization. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI, Yokohama, Japan* (pp. 2512–2518).

Zhang, J., Luo, Z., Li, C., Zhou, C., & Li, S. (2019). Manifold regularized discriminative feature selection for multi-label learning. *Pattern Recognition, 95*, 136–150.

Zheng, B., Zhang, J., Yoon, S. W., Lam, S. S., Khasawneh, M. T., & Poranki, S. (2015). Predictive modeling of hospital readmissions using metaheuristics and data mining. *Expert Systems with Applications, 42*, 7110–7120.