Heliyon 10 (2024) e36121

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

The comparative experimental study of rehabilitation program decision for spinal cord injury based on electronic medical records

Botao Qie [a,c], Xin Guo [a,b,c,*], Wei Chen [d], Suiran Yu [e], Zhengtao Wang [e]

[a] School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin, 300401, China
[b] Qinhuangdao Research Institute, National Research Center for Rehabilitation Technical Aids, Qinhuangdao, 066004, China
[c] Control Engineering Technology Innovation Center of Hebei Province, Hebei University of Technology, Tianjin, 300401, China
[d] National Research Center for Rehabilitation Technical Aids, Beijing, 100176, China
[e] School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

ARTICLE INFO

Keywords:
Electronic medical record
MLSMOTE
Multi-label classification
Rehabilitation program decision
Spinal cord injury

ABSTRACT

Objective: Electronic medical records (EMRs) contain patients' medical and health information. The Utilization of EMRs for assisted diagnosis is of significant importance for the rehabilitation of spinal cord injury (SCI) patients. Therefore, this study proposes a decision-making model for rehabilitation programs of SCI patients based on EMRs.
Methods: First, an Electronic Medical Records (EMR) dataset comprising 1252 Spinal Cord Injury (SCI) patients was constructed, and data preprocessing was completed. Second, the Random Forest (RF) feature extraction algorithm was utilized to select case features with high contribution levels. Then, to address the imbalance issue in EMRs, a multi-label learning framework based on the improved MLSMOTE was adopted. Finally, seven multi-label classification models were employed to predict patients' physical therapy (PT) prescriptions.
Results: The proposed improved MLSMOTE multi-label learning framework can solve the problem of class imbalance. Compared with the other six models, the CC model has improved significantly in many metrics. Its hamming loss and ranking loss were 0.1388 and 0.2467, and precision, recall, and F1-score were 83.33 %, 81.20 %, and 79.82 % respectively.
Conclusions: The improved MLSMOTE multi-label learning framework proposed in this study can make full use of the information in EMRs and effectively improve the decision-making accuracy of rehabilitation treatment programs.

1. Introduction

Spinal Cord Injury (SCI) refers to the damage inflicted upon the spinal cord due to external force, trauma, or disease. It is a complex medical condition in clinical practice. SCI are typically classified into complete spinal cord injuries and incomplete spinal cord injuries, both of which result in varying degrees of impairment in the patient's motor functions [1]. SCI not only causes serious harm to the physical health of patients, but also has a huge economic impact on families and society [2,3]. Patients with SCI are at increased risk of cardiometabolic diseases, osteoporosis, and neuralgia, mostly due to reduced physical activity [4]. Regardless of the duration of the disease and the degree of injury, the restoration of motor ability is a priority for patients with spinal cord injuries. Studies have shown

* Corresponding author. School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin, 300401, China.
E-mail addresses: qbthebut@163.com (B. Qie), gxhebut@aliyun.com (X. Guo), 444146239@qq.com (W. Chen), sryu@sjtu.edu.cn (S. Yu), wangzhengtao@sjtu.edu.cn (Z. Wang).

https://doi.org/10.1016/j.heliyon.2024.e36121
Received 29 March 2024; Received in revised form 20 June 2024; Accepted 9 August 2024
Available online 13 August 2024

Abbreviation list:

ADL	Activity of Daily Living
ADL_ADM	ADL on Admission
BR	Binary Relevance
CC	Classifier Chains
CLR	Calibrated Label Ranking
EMRs	Electronic medical records
KNN	K-Nearest Neighbor
LP	Label Powerset
MLC	Multi-label Classification
ML-KNN	Multi-Label K-Nearest Neighbor
PT	Physical Therapy
RAkEL	Random k-Labelsets
Rank-SVM	Ranking Support Vector Machine
SCI	Spinal Cord Injury

that reasonable rehabilitation training can effectively help SCI patients recover their physical functions [5]. Physical therapy (PT), as a kind of rehabilitation training, can improve patients' daily functional abilities, increase cardiometabolic health [6]. The main training content includes mat exercise, walking training, crutch use training, etc. It is one of the important means to promote the rehabilitation of SCI patients [7].

Electronic medical records (EMRs) are detailed records of patients' health information maintained by healthcare professionals [8]. Compared to traditional paper-based medical records, the advantages of EMRs lie in the ability to quickly create and share patients' medical information. How to use these resources to assist doctors in clinical decision-making and improve the level of medical care is an urgent problem to be solved. Michael Shapiro et al. attempted to use natural language processing technology to analyze the text information of EMRs and successfully detected COVID-19 symptoms [9]. Yung-Ta Kao et al. developed a machine learning-based atrial fibrillation risk prediction model for elderly patients using the clinical EMRs database [10].

Multi-label classification (MLC) is a machine learning classification method, which means that a sample may have multiple labels at the same time. For example, the complications of a disease often have multiple symptoms. Multi-label classification has received great attention in text classification, image annotation, biological gene detection and so on. In recent years, multi-label classification has been very popular in disease diagnosis and clinical treatment [11–13]. In 2020, Hiba Chougrad et al. built a multi-label breast lesion diagnosis model using convolutional neural network technology based on mammography images [14]. In 2021, Ankush Jamthikar and others used six different multi-label classification methods to establish a predictive model for phenotypes of cardiovascular disease [15]. In actual clinical treatment, the rehabilitation prescription of SCI patients may contain a variety of training methods, which belongs to a typical multi-label classification problem.

At present, due to the lower incidence rate of SCI compared to other diseases, it possesses a particular and complex nature. Therefore, obtaining large-scale datasets for SCI is relatively challenging. Additionally, there is limited research in the country on the use of electronic medical records for diagnostic assistance. Traditional offline diagnoses undoubtedly hinder diagnostic efficiency. Therefore, there is an urgent need to conduct research on intelligent prescription decision models for spinal SCI patients. Simultaneously, medical datasets often encounter data imbalance, with significant disparities among different label categories.

In view of the above problems, this paper establishes the EMRs data set of 1252 SCI patients. Further, we propose a multi-label classification framework based on the improved MLSMOTE, which utilizes multiple MLC models for SCI EMRs for rehabilitation prescription decision-making. It is a preliminary attempt to assist the diagnosis based on the EMRs of Chinese SCI patients.

2. Material

2.1. Data acquisition

This research was conducted in accordance with the principles of the Declaration of Helsinki, using data from a tertiary hospital in Beijing and approved by the hospital ethics committee. The basic medical records of SCI patients in the hospital's rehabilitation department over the past three years were extracted. Through thorough discussion with a specialist in rehabilitation medicine, the final criteria for including cases are determined to be: (1) patients with traumatic spinal cord injury confirmed by CT, (2) injury sites were cervical, thoracic, lumbar, and caudal vertebra, (3) the lowest normal sensory plane was located below C1, (4) age 18–65 years, and (5) signed the informed consent. A total of 1252 SCI patients' medical records were screened, of which 1017 (81.23 %) were males with a mean age of (34.09 $\pm$ 13.69) years and 235 (18.77 %) were females with a mean age of (27.53 $\pm$ 18.33) years. Each report includes the patient's case information and prescription information. The case information mainly included the patients' sex, age, height, weight, Activity of Daily Living on admission (ADL_ADM), ASIA damage index, Damage reason, Damage position, Damage segment, Damage degree, and fracture situation, etc. Among them, Activity of Daily Living (ADL) refers to assessing whether a person can complete basic activities of daily living, such as dressing, eating, bathing, going to the toilet, walking, etc. The ADL scores mentioned in this research were evaluated by the modified Barthel index scale. The prescription information was the six PT rehabilitation training methods frequently used by the patients in the medical records. To obtain more visual data results, the paper-based medical records collected from the hospital's medical records department will be imported into the backend database through a self-developed data collection system. This will establish a dataset of EMRs for spinal cord injury (SCI) patients. The data collection interface is shown in Fig. 1. To ensure patient privacy and data security, the patient 's name, address and other personal information are not included.

2.2. Data preprocessing

Formally, the data in EMRs can be divided into numerical and text data, which cannot be directly used by the machine learning models. And the dimensional standards between the attributes in the medical records are not uniform, and the data structures are inconsistent. In this case, training the data is easy to affect the classification effect of the classifier. In addition, the medical record data collected manually has problems such as non-standard data and missing important fields. To obtain effective data for the model used, it is necessary to preprocess the medical record data. Fig. 2 is the specific process of data preprocessing.

Firstly, the missing data in the medical records of SCI patients are filled, and the missing data are shown in Fig. 3. It can be seen that the characteristics of some patients contain fewer missing values. For continuous indicators, the mean is used to fill the missing values, and for discrete indicators, the mode is used to fill the missing values. To facilitate machine learning model training and improve computational efficiency, it is necessary to convert the textual data in the medical records of SCI patients into numerical data. In this research, sex, injury degree, fracture condition, etc. are encoded by 0–1, and the injury reason and injury position were encoded by One-hot encoding. For multi-dimensional features, such as injury segment, Orderal encoding was used to categorical features are converted to categorical values. Some of the feature coding rules and descriptive statistics were shown in Table 1.

Then the features are normalized, the purpose is to unify the dimension standard, the formula is as follows [16].

$$x^{*} = \frac{x - Min}{Max - Min} \tag{1}$$

Finally, use the Random Forest (RF) feature selection method to filter the input features of the model. The RF feature selection method is an embedded approach that identifies and filters key features in high-dimensional data by evaluating the importance scores of features across multiple decision trees. This effectively enhances the model's interpretability and predictive accuracy. The results of the feature importance scores are shown in Fig. 4. Select features with an importance score higher than 0.1 as the input features for the model, which are AISA damage index, ADL_ADM, age, height, weight, and damage segment.

The rehabilitation training program was determined by the rehabilitation physician according to the patient 's own physical characteristics and health status, including six different PT training methods. These treatment methods were selected as the research target objects. The frequency of each PT is different, and the frequency statistics are shown in Fig. 5. The number of samples in each category of the data set is too different. Training the model on the unbalanced data set will make the prediction results of the machine learning algorithm more inclined to the majority of samples, so that the performance of the trained model is poor. If all small class samples are predicted as large class samples, the recall rate of the model will be reduced, which is obviously not suitable for medical

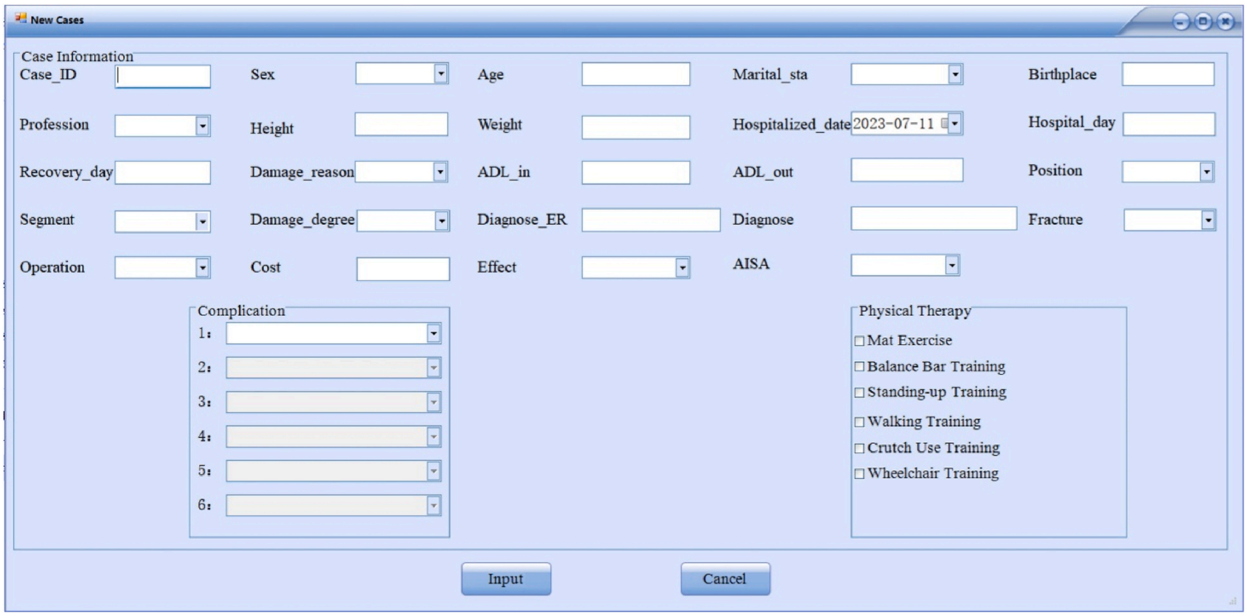

Fig. 1. Collection interface of medical records.

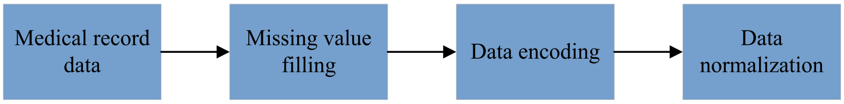

Fig. 2. Process of data preprocessing.

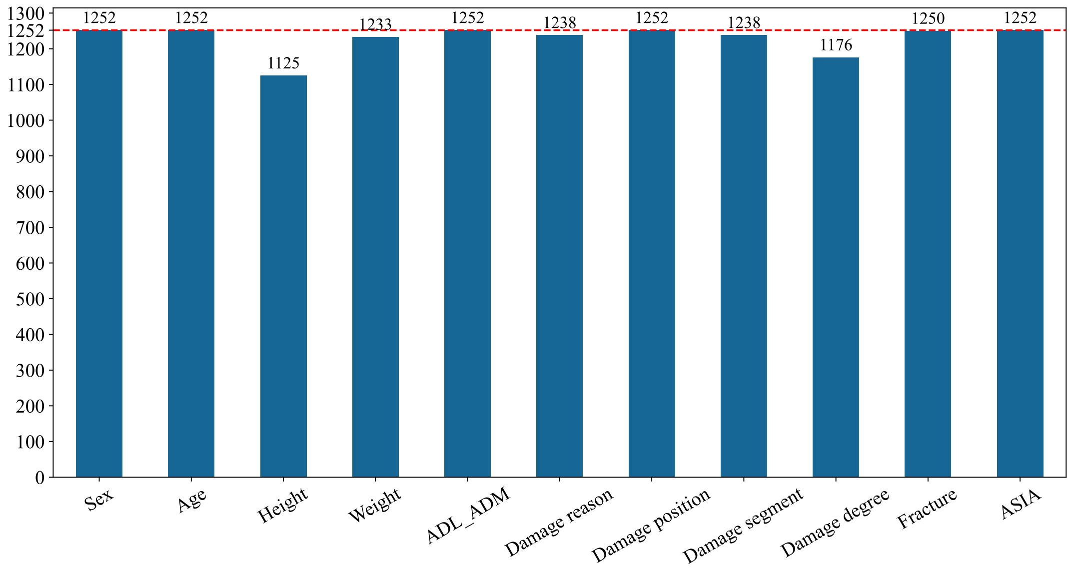

Fig. 3. Data missing situation.

Table 1
Partial feature encoding rules and statistics.

Patient characteristics	Value content	Coding rules	Proportion
Sex	Male	1	81.23 %
	Female	0	18.77 %
Damage degree	Complete injury	1	45.93 %
	Incomplete injury	0	54.07 %
Fracture situation	Yes	1	71.72 %
	No	0	28.28 %
ASIA	A	(1,0,0,0)	45.93 %
	B	(0,1,0,0)	12.54 %
	C	(0,0,1,0)	16.61 %
	D	(0,0,0,1)	24.92 %
Damage position	Cervical vertebra	(1,0,0,0)	42.25 %
	Thoracic vertebra	(0,1,0,0)	48.16 %
	Lumbar vertebra	(0,0,1,0)	7.10 %
	Caudal vertebra	(0,0,0,1)	2.49 %
Damage segment	C4	3	19.89 %
	C5	4	11.18 %
	T10	17	9.74 %
	T11	18	9.66 %
	T12	19	5.11 %
	Others	…	44.42 %
Damage reason	Traffic accident	(1,0,0,0,0,0)	34.19 %
	Heavy object injury	(0,1,0,0,0,0)	9.34 %
	High fall	(0,0,1,0,0,0)	12.86 %
	Industrial injury	(0,0,0,1,0,0)	16.45 %
	Disease	(0,0,0,0,1,0)	11.58 %
	Others	(0,0,0,0,0,1)	15.58 %

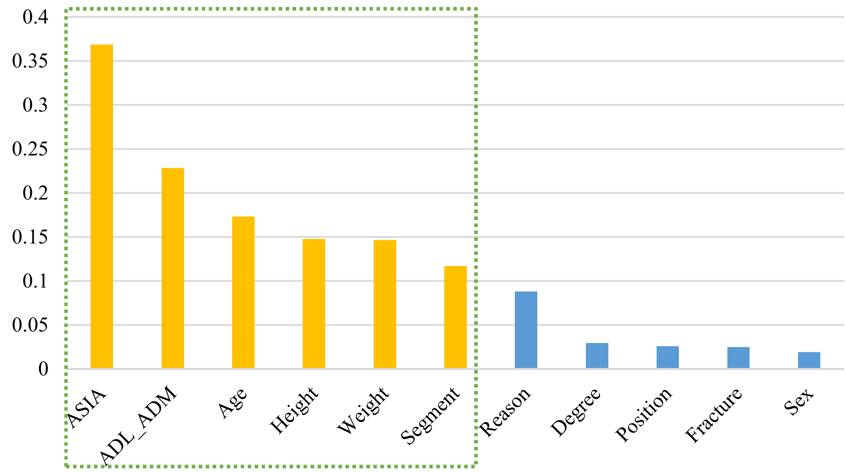

Fig. 4. Importance score of case characteristics.

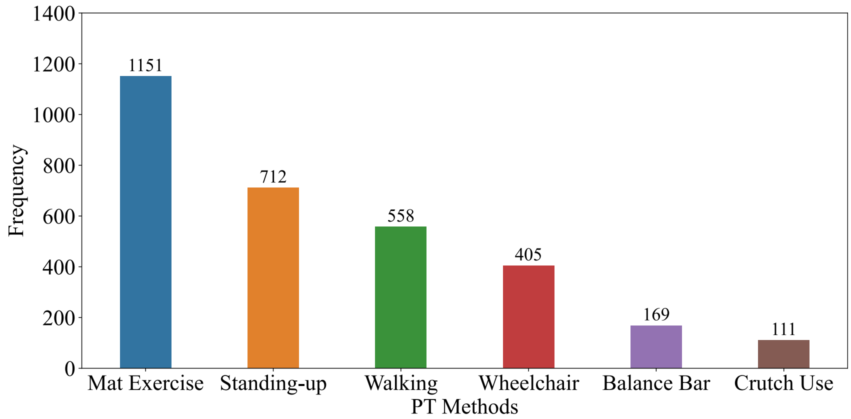

Fig. 5. Six PT rehabilitation methods and their frequency.

data sets. Therefore, before using machine learning algorithms to predict, data enhancement is needed to alleviate the impact of category imbalance.

3. Methods

3.1. Correlation analysis

Mutual information is a metric used to measure the correlation between two variables. The mutual information possesses the property of symmetry, where a higher value indicates a stronger correlation between the two variables. Using the SCI electronic medical records dataset to calculate mutual information, we aim to analyze the statistical relationships between different PT methods, as illustrated in Fig. 6. To enable comparisons across different data distributions and further reduce bias, we adopt the calculation method of Normalized Mutual Information (NMI). The computation approach is as follows.

$$NMI(A,B) = \frac{2 \times (H(A) - H(A|B))}{H(A) + H(B)} \tag{2}$$

From Fig. 6, it can be observed that the NMI between cane usage training and walking training is relatively high, with a correlation coefficient of 0.11, indicating a strong dependency between the predictions of these two PT methods on each other. This prompted the introduction of a multi-label classification model, taking into account the correlation between PT labels for further prediction.

3.2. MLSMOTE

MLSMOTE algorithm is an oversampling technique for multi-label proposed by Charte et al. [17] As an extension of the SMOTE algorithm, the basic idea of the MLSMOTE algorithm is to create additional synthetic data samples by interpolation to increase the

Fig. 6. Correlation coefficients of the six PT methods.

number of samples in a small number of categories in the data set. It specifically deals with the class imbalance problem in multi-label classification problems. Firstly, the algorithm calculates the imbalance ratio of each label and the average imbalance ratio of all labels to determine the sample data to be generated.

The formula for calculating label imbalance ratio is as follows:

$$IRPL(j) = \frac{\operatorname{argmax}_{j'=Y_1}^{Y_q}\left(\sum_{i=1}^{m} h(j', Y_i)\right)}{\sum_{i=1}^{q} h(j, Y_i)}, h(j, Y_i) = \begin{cases} 1, j \in Y_i \\ 0, j \notin Y_i \end{cases} \tag{3}$$

Where, Y is a set of labels with q as the number of labels, and MIR is defined as the average value of all labels $IRPL(j)$, calculated as follows.

$$MIR = \frac{1}{q}\sum_{j=Y_1}^{Y_q}(IRPL(j)) \tag{4}$$

MLSMOTE considers labels that satisfy condition $IRPL(j) > MIR$ as minority labels, and for each identified minority label, it finds all sample instances that contain that label.

Secondly, for each minority class sample, the algorithm finds its k nearest neighbors in the feature space. Finally, the voting mechanism is used to generate the corresponding label set for each newly synthesized sample [18]. Therefore, MLSMOTE algorithm not only considers the similarity of feature space, but also considers the similarity of label space. This makes its performance superior to other resampling techniques.

3.3. MLTL

The Multi-Label Tomek Links (MLTL) algorithm, proposed by Pereira in 2020, is an undersampling technique used for addressing imbalanced datasets in multi-label classification problems [19]. Unlike MLSMOTE, MLTL considers labels that satisfy condition $IRPL(j) < MIR$ as majority labels, based on the criteria for majority labels. It calculates the difference between two sets of labels using

Hamming distance. By identifying Tomek Links in the multi-label dataset, and for each pair of Tomek Links formed by the label sets, it removes instances that do not contain minority labels, thereby reducing the overlap between labels. This, in turn, enhances the classifier's ability to recognize minority classes. It is worth noting that, in the case where a Tomek Link pair does not contain any minority labels, MLTL opts to remove this Tomek Link pair from the dataset.

3.4. Our proposed method

Since MLSMOTE generates synthetic samples using the k-nearest neighbors of minority class samples, if the nearest neighbors contain outliers or noise, then the generated synthetic samples may also carry this noise, thereby affecting the model's performance. MLTL focuses on the distribution of noise and can effectively eliminate noisy samples. Therefore, this paper proposes an improved MLSMOTE algorithm, which is divided into two steps: first, perform MLSMOTE oversampling, followed by MLTL undersampling. Additionally, to thoroughly identify imbalanced labels, this paper introduces another metric for measuring category imbalance. This metric reflects the degree of imbalance within each label by calculating the number of positive and negative samples for each label, as shown in the following formula.

$$IMR_j = \frac{\max\left(\left|D_j^+\right|, \left|D_j^-\right|\right)}{\min\left(\left|D_j^+\right|, \left|D_j^-\right|\right)}, \begin{cases} D_j^+ = \left\{x_j \middle| \left(x_j, y_j\right) \in D, y_{ij} = 1\right\} \\ D_j^- = \left\{x_j \middle| \left(x_j, y_j\right) \in D, y_{ij} = 0\right\} \end{cases} \tag{5}$$

AIMR is the average value of all labels IMR_j. The larger the value, the more imbalanced the label set is. Its expression is as follows.

$$AIMR = \frac{1}{q}\sum_{j=1}^{q} IMR_j \tag{6}$$

This study proposes a multi-label classification framework based on an improved MLSMOTE, as shown in Fig. 7. When resampling the SCI dataset using the improved MLSMOTE algorithm, labels that simultaneously satisfy conditions $IRPL(j) > MIR$ and $IMR_j > AIMR$ are considered as minority labels, while those that meet conditions $IRPL(j) < MIR$ and $IMR_j < AIMR$ are treated as majority labels. To verify the performance of the improved MLSMOTE algorithm, this paper compares it with commonly used MLSMOTE and MLTL algorithms. The imbalance degree of the resampled SCI dataset is measured using indicators *MaxIRPL*, *MIR*, *MaxIMR*, and *AMIR*, where *MaxIRPL* and *MaxIMR* represent the maximum values among all labels. The sampling results are shown in Table 2.

As shown in Table 2, the four indicators of the improved MLSMOTE algorithm are all lower than those of other resampling algorithms, significantly improving the imbalance of the EMRs dataset of SCI.

4. Multi-label classification model

In the multi-label classification task, let $X = \mathbb{R}^d$ be a data sample space consisting of feature vector $X_1, X_2, \ldots, X_d$, $L = \{l_1, l_2, \ldots, l_q\}$ is the set of category labels, which contains a total of q label vectors. Let $D = \{(X_i, Y_i) | 1 \le i \le m\}$ denote the training dataset consisting of m samples. $T = \left\{\left(X_j', Y_j'\right) | 1 \le j \le n\right\}$ denotes the test dataset consisting of n samples. Let the feature vector of a training dataset sample X_i be represented as $X_i = \{x_1, x_2, \ldots, x_d\} \in X$, d is the number of characteristics of the sample, the label vector output of one instance X_i is represented as $Y_i = \{y_1, y_2, \ldots, y_q\} \in L$. Similarly, $Y_i' = \{y_1', y_2', \ldots, y_q'\} \in L$ is the label vector corresponding to the test dataset sample $X_i' = \{x_1', x_2', \ldots, x_d'\} \in X$, $\widehat{Y}_i = \{\widehat{y}_1, \widehat{y}_2, \ldots, \widehat{y}_q\}$ is the predicted label vector, if $l_j \in \widehat{Y}_i$, then $\widehat{y}_j = 1$, otherwise $\widehat{y}_j = 0$, where $j = 1, 2, \ldots q$. The task of MLC is to learn a classifier $h : X \rightarrow \{0, 1\}^q$ for the training dataset D such that the set of predicted labels $\widehat{Y}_x$ of the unknown

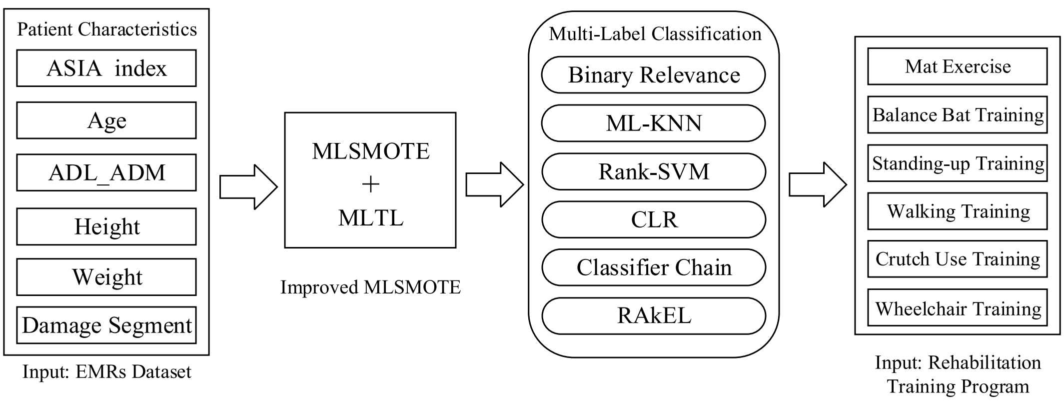

Fig. 7. A multi-label classification framework based on improved MLSMOTE.

Table 2
Results of Data sampling.

Algorithms	MaxIRPL	MIR	MaxIMR	AIMR
Raw data	12.8023	4.6844	13.5581	5.5193
MLSMOTE	7.1333	3.0602	7.4784	3.6935
MLTL	10.8333	4.2066	11.6060	4.8862
Improved MLSMOTE	5.7486	2.8708	6.6770	3.3940

samples $x \in X'$ is as identical as possible to the set of true labels Y'_x [20].

In general, based on the different ways and degrees of utilizing the relevance of each label, existing multi-label classification algorithms can be divided into three categories: first-order strategies, second-order strategies, and advanced strategies [21]. The first-order strategy does not consider the relationship between any labels and regards each label as an independent classification task. Common first-order strategy algorithms are Binary Relevance (BR) and Multi-Label K-Nearest Neighbor (ML-KNN). The second-order strategy considers the relationship between label pairs, which attempts to capture the dual relationship and interaction between labels. Ranking Support Vector Machine (Rank-SVM) and Calibrated Label Ranking (CLR) are typical second-order strategy algorithms. The higher-order strategy considers the relationship between three or more labels, and it attempts to simulate the complex dependencies between each label and all other labels. The representative algorithms are Classifier Chain (CC), Label Powerset (LP) and Random k-Label sets (RAkEL).

4.1. Binary Relevance

The basic idea of BR is to decompose the multi-label problem into multiple single-label sub-problems. In the training process, it considers the binary classification problem of each label, and learns a classifier independently for each label. In the prediction stage, the prediction results of each classifier are combined into a label set. The set includes all labels that are predicted to be positive by the corresponding binary classifier. BR algorithm is the cornerstone of many advanced multi-label algorithms, and it is also the simplest and most direct method [22,23].

4.2. Multi-label K-Nearest neighbor

This algorithm is an extension of the traditional K-Nearest Neighbor (KNN) for multi-label environment. For an unknown sample x, MLKNN first identifies the k most similar samples in the training dataset and represents the set of k nearest neighbors of sample x as $N(x)$. For each label category y_j, MLKNN counts the number of samples in $N(x)$ that contain that label as C_i. It uses H_i^1 to represent the event of sample x having label y_j, and H_j^0 to represent the event of sample x not having label y_j. The calculation formulas for the prior probability $P\left(H_j^1\right)$、 $P\left(H_j^0\right)$ and posterior probability $P\left(H_j^1|C_i\right)$、 $P\left(H_j^1|C_i\right)$ for each label are as follows. The algorithm uses Bayesian rules combined with prior and posterior probabilities to calculate the probability that the new sample has the label and determines whether the unknown sample contains a label according to the maximum posterior probability criterion[24].

$$P\left(H_j|C_i\right) = \frac{P\left(H_j\right) * P\left(C_i|H_j\right)}{P\left(C_i\right)} \tag{7}$$

4.3. Ranking Support Vector Machine

Rank-SVM algorithm is an improvement of SVM classification algorithm [25]. This method uses the maximum margin strategy to construct a series of linear classifiers with the goal of reducing the ranking loss value. In addition, by using the kernel method, Rank-SVM can properly handle the case of nonlinear classification. Specifically, the goal of Rank-SVM is to identify a hyperplane, placing samples with higher rankings on one side of the hyperplane, and placing samples with lower rankings on the other side. This algorithm learns a linear classifier $W = \left\{\left(w_j, b_j\right)|1 \le j \le q\right\}$ for each label, where w_j and b_j represent the weight and bias of the *j-th* label. Given a multi-label training set D, the Rank-SVM algorithm finds the weights w_j and bias term b_j such that for every pair of samples $\left(x_i, x_j\right)$ with different rankings, the classification margin accurately reflects the relative ranking of these two samples. The classification interval of instance x_i is defined as follows.

$$\min_{\left(x_i, y_i\right) \in \mathbb{D}} \frac{\langle \omega_j - \omega_k, \mathrm{x}_i \rangle + b_j - b_k}{\left\|\omega_j - \omega_k\right\|} \tag{8}$$

4.4. Calibrated Label Ranking

The CLR algorithm transforms the multi-label classification problem into a label sorting problem, in which the label sorting is implemented by pairwise comparison. The CLR algorithm introduces a virtual label to help distinguish between related labels and unrelated labels. For the label set containing q labels, the CLR algorithm creates all possible label pairs, and uses a binary classifier to distinguish each pair of labels. In the training phase, the CLR algorithm selects samples that contain only one label in the current label pair from the original data set and uses the voting method to integrate each binary classifier. The virtual label plays a role in predicting unknown samples, which acts as a dividing line: all the labels before the virtual label are considered to be related to the sample, and the labels after the virtual label are considered to be irrelevant [26].

4.5. Classifier Chain

CC is one of the most common algorithms for exploiting label relevance [27]. It transforms the multi-label classification problem into q single-label binary classifiers $h = (h_1, h_2, \dots, h_q)$, q classifiers randomly generate a sequential chain $\{h_1 \rightarrow h_2 \rightarrow \dots \rightarrow h_q\}$, in the prediction phase, the first classifier is trained using the feature attributes of the sample as input and proceeds down the specified chain after training is complete, and the attribute space of each subsequent classifier h_j consists of the original features and the labels predicted by the first *j-1* classifiers in the chain until all base classifiers have completed training. Using $h_i(x')$ to represent the prediction result of each classifier, and x_i to represent the original features of a sample, when constructing the *j-th* binary classifier, the input vector of the model is represented as follows:

$$x_i' = \{x_i, [h_1(x_i'), h_2(x_i'), \dots, h_{j-1}(x_i')] | j \le q\} \tag{9}$$

4.6. Label Powerset

The LP algorithm transforms the multi-label problem into a multi-classification problem. It encodes different label sets and introduces a new category. In other words, it treats each unique combination of labels as a new category. During the prediction phase, the model predicts a new category, which is actually a combination of labels. Finally, this label combination is the prediction result of the model for the input instance. However, when the label set is too large, it will lead to high LP complexity and affect the classification accuracy [28].

4.7. Random k-Labelsets

The RAkEL algorithm is an improved multi-label classification model based on the LP method [29]. It uses an ensemble learning framework, and each weak classifier is acted by LP. The RAkEL algorithm first randomly selects k label subsets from the original data set and creates a corresponding data set. Then, each weak classifier is used to train the new data set, and these classifiers are merged by voting mechanism. For an unknown sample, the RAkEL algorithm uses k trained classifiers to predict the label of the sample. This will generate a vote count, according to the vote count to determine the sort of labels, in order to complete the multi-label prediction.

5. Results

In this research, the training set and test set are divided according to the ratio of 80:20. In the training set, the improved MLSMOTE algorithm is used to resample the EMRs dataset of SCI patients to mitigate the impact of imbalanced data. The model inputs are the patient's ASIA damage index, ADL_ADM, age, height, weight, and damage segment. The model output is the patient's PT prescription.

5.1. Evaluation metrics

According to the research needs, the following evaluation metrics are selected to measure the performance of the multi-label classification model [30].

Hamming Loss: This metric represents the proportion of incorrectly predicted labels to the total number of labels in the test set. The smaller the value, the better the performance. The calculation formula is as follows.

$$H - Loss = 1 - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{q}\sum_{j=1}^{q} I\left(y_j'^{(i)}, \hat{y}_j^{(i)}\right), where\ I(a,b) = \begin{cases} 1 & \text{if a} = \text{b} \\ 0 & \text{if a} \neq \text{b} \end{cases} \tag{10}$$

Ranking Loss: This metric represents the probability of irrelevant labels being ranked before relevant labels in the predicted label sequence.

$$R - Loss = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|Y_i'||\overline{Y}_i|}\left|\left\{\left(y_i, y_j\right) \middle| f(x_i, y_i) \le f\left(x_i, y_j\right), y_j \in \overline{Y}_i, y_i \in Y_i'\right\}\right| \tag{11}$$

Where $\overline{Y}_i$ is the complement of the label set Y_i' and $f(\bullet)$ is the prediction function.

Accuracy: The index calculates the average accuracy of all samples.

$$Accuracy = \frac{1}{n}\sum_{i=1}^{n}\frac{|y'^{(i)} \cap \hat{y}^{(i)}|}{|y'^{(i)} \cup \hat{y}^{(i)}|} \tag{12}$$

Precision: This metric represents the proportion of correctly predicted instances for each label out of the total instances predicted for that label in the test set.

$$Precision = \frac{1}{n}\sum_{i=1}^{n}\frac{|y'^{(i)} \cap \widehat{y}^{(i)}|}{|\widehat{y}^{(i)}|} \tag{13}$$

Recall: This metric represents the number of correct predictions for each label as a percentage of the total number of instances of that label in the test set. In the MLC task, Precision and Recall represents the average values of all labels.

$$Recall = \frac{1}{n}\sum_{i=1}^{n}\frac{|y'^{(i)} \cap \widehat{y}^{(i)}|}{|y'^{(i)}|} \tag{14}$$

F1-score: This metric is the harmonic mean of Precision and Recall. Its values should be as large as possible.

$$F1 - score = \frac{1}{n}\sum_{i=1}^{n}\frac{2|y'^{(i)} \cap \widehat{y}^{(i)}|}{|y'^{(i)}| + |\widehat{y}^{(i)}|} \tag{15}$$

F1-measure-macro: This metric is the macro-averaged F1-measure. TP_j, TN_j, FP_j, and FN_j represent True Positive, True Negative, False Positive, and False Negative, respectively, for the *j-th* label.

$$\text{F1-measure-macro} = \frac{1}{q}\sum_{j=1}^{q}\frac{2TP_j}{2TP_j + FP_j + FN_j} \tag{16}$$

5.2. Comparison test of algorithms

The first part of the experiment involves oversampling the SCI dataset using the improved MLSMOTE algorithm. The suitability of the improved MLSMOTE algorithm is then validated using the seven multi-label classification models. The F1-measure index is selected to evaluate the imbalance of the whole sample set before and after sampling. In the ML-KNN algorithm, the k-nearest neighbor parameter is set to 5, and in the RAkEL algorithm, the size of the label subset is set to 3. The test results are shown in Figs. 8 and 9.

After using oversampling algorithm, except for ML-KNN and Rank-SVM algorithm, other algorithms have been significantly improved. Since the ML-KNN algorithm combines the KNN method and the Bayesian inference method, the use of Bayesian posterior rules can largely overcome the imbalance problem. Therefore, the ML-KNN algorithm is not sensitive to unbalanced data sets, and its improvement effect is limited. Similarly, SVM focuses only on the samples near the decision boundary, so it is not very sensitive to data imbalance issues. Therefore, its improvement effect is also limited.

The second part of the experiment involves redividing the dataset obtained after MLSMOTE oversampling into training and testing sets and designing a control experiment. The experimental results of the seven multi-label models are shown in Table 3.

It can be seen from Table 3 that most of the high-order algorithms are superior to the first-order algorithm and the second-order algorithm in terms of the six indicators. For ML-KNN and Rank-SVM algorithms, after the sample is balanced, the model will pay more attention to the minority class, which changes the decision boundary and affects the performance of the algorithm. Among the three advanced algorithms, the CC algorithm produced the best performance on both F1-measure-macro and F1-measure-micro metrics. The reason is that the CC algorithm passes each label among single classifiers, training the multi-label learning model through a chain structure. It treats the associations between labels as attribute relationships. The prediction results of the previous classifier in the chain are used as feature attributes for the current classifier, reducing the impact of label noise on the overall model.

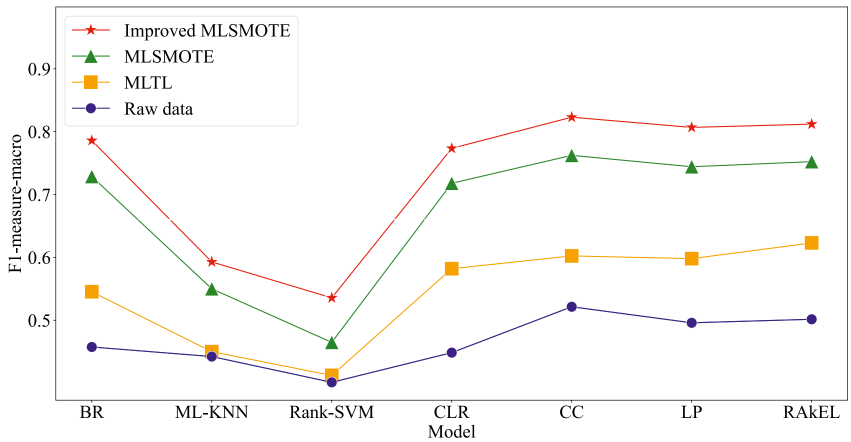

Fig. 8. Comparison of F1-measure-macro metrics for data balance experiment results.

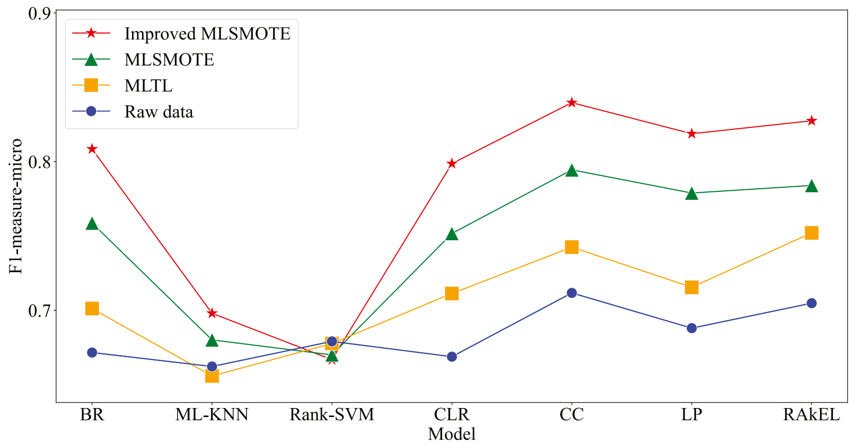

Fig. 9. Comparison of F1-measure-micro metrics for data balance experiment results.

Table 3
Results of seven multi-label classification Models.

Model	H-Loss	R-Loss	Accuracy	Precision	Recall	F1-score
BR	0.1502	0.2688	0.6991	0.8080	0.8115	0.7787
ML-KNN	0.1859	0.3338	0.6055	0.7829	0.7597	0.7310
Rank-SVM	0.1741	0.3294	0.6133	0.8085	0.7231	0.7425
CLR	0.1545	0.2783	0.6846	0.8128	0.7878	0.7837
CC	0.1388	0.2467	0.7281	0.8333	0.8120	0.7982
LP	0.1478	0.2638	0.7114	0.8197	0.8044	0.7871
RAkEL	0.1451	0.2505	0.7221	0.8294	0.8105	0.7925

6. Discussion

The high rate of disability associated with spinal cord injury continues to pose a threat to people's life and health [31]. Due to damage to the spinal cord structure and function, this results in impairments in movement, sensation, and autonomic nervous function below the level of injury. Physical therapy employs a diverse range of movement modalities to aid in the rehabilitation of SCI patients [32]. This includes enhancing muscular strength, reducing muscle spasms, improving motor control, enhancing synaptic plasticity of nerves, and improving cardiovascular function in patients. It is one of the crucial means in clinical practice to promote functional recovery in spinal cord injury patients. The selection of exercise modalities is typically based on the specific condition of the patient and rehabilitation goals. Rehabilitation professionals need to conduct a thorough assessment of each patient and develop a rehabilitation plan, leading to issues of low diagnostic efficiency and subjectivity. This significantly increases the workload for rehabilitation therapists. Meanwhile, traditional offline rehabilitation diagnostics are often a standardized approach, making it challenging to fully meet the individualized rehabilitation needs of each patient. This is because of a lack of sufficient objective data, making it challenging for even experienced rehabilitation therapists to have a comprehensive understanding of the patient's physical and functional status.

With the application of big data and artificial intelligence technologies in the medical field, electronic medical records are increasingly favored by healthcare professionals. EMRs provide detailed records of a patient's health history, physical characteristics, and clinical manifestations. By mining the detailed information within, analyzing patient data can assist rehabilitation professionals in better understanding the clinical patterns of SCI, grasping personalized characteristics of rehabilitation treatment, and maximizing the promotion of patients' functional recovery. This enhances the decision-making efficiency of rehabilitation professionals. Therefore, this study has constructed an MLSMOTE-MLC learning framework, successfully achieving intelligent decision-making for personalized rehabilitation training programs for SCI patients.

Firstly, following institutional ethical approval, we extracted 1252 EMRs from the rehabilitation department of the hospital, encompassing patient case information and rehabilitation training prescriptions. Based on this data, the MLSMOTE multi-label resampling algorithm was employed to balance the PT label categories. Finally, various MLC algorithms were used to successfully classify different PT prescriptions for patients. To develop a more efficient and unified decision-making method for rehabilitation training programs. We use six performance evaluation metrics to evaluate the prediction results of several MLC models.

6.1. Novelty and potential value

To our knowledge, this is the first study in China to try to establish an EMRs data set of spinal cord injury and apply the MLC model to the field of prescription diagnosis of SCI patients. Comparative experiments show that it is very effective to use high-order strategies for prediction by using the correlation between labels. This research can provide data samples for clinical research and help doctors and patients gain a deeper understanding of the etiology and pathogenesis of spinal cord injury. By analyzing the information in the EMRs, medical researchers can provide more personalized rehabilitation programs for patients and improve the quality of hospital services. The use of multi-label classification technology can help rehabilitators to establish a more comprehensive rehabilitation program decision-making system. Therefore, this study is highly innovative in the use of EMRs and the prediction of rehabilitation programs.

6.2. Limitations and future directions

In this research, machine learning technology was used to construct a rehabilitation training decision model for spinal cord injury, which effectively improved the efficiency of clinical decision-making, but it also had some limitations. When the number of samples in the training data set is too small, the model will be overly dependent on a small amount of training data and perform poorly on the test set of new data. In other words, the model falls into an overfitting state and reduces the generalization ability of the model. In addition, this study's approach aims to develop a decision model for SCI rehabilitation training programs. The next focus will be applying this model to diagnose and treat other medical datasets.

7. Conclusions

Based on the analysis of SCI electronic medical records, this research regards rehabilitation prescription decision-making as a multi-label learning task, proposes an improved MLSMOTE multi-label learning framework, and uses different MLC models for comparative verification. It effectively addresses the issue of imbalanced sample categories when training models using medical electronic health record data. The experimental results show that the performance of the CC model is higher than that of the other six models. Its Precision is 83.33 %, Recall is 81.20 %, and F1-score is 79.82 %. This research is necessary because it not only provides a better model for rehabilitation prescription decision-making for SCI patients but also offers new insights for achieving online intelligent medical prescription decision-making. It is expected that this research can achieve high medical application value.

Ethics approval code for ethics committee

E20230294I.

Funding

The research work is supported by National Key R&D Program of China (2019YFB1312500).

Data availability statement

Since the data pertains to patient privacy, it will not be made public.

CRediT authorship contribution statement

Botao Qie: Writing – original draft, Visualization, Validation, Methodology, Data curation. **Xin Guo:** Writing – review & editing, Funding acquisition, Conceptualization. **Wei Chen:** Writing – review & editing. **Suiran Yu:** Writing – review & editing. **Zhengtao Wang:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgments

Thanks to the clinical work of the rehabilitation doctors in the Beijing Hospital.

References

[1] X. Liu, Y. Zhang, Y. Wang, et al., Inflammatory response to spinal cord injury and its treatment, J World Neurosurgery 155 (2021) 19–31, https://doi.org/10.1016/j.wneu.2021.07.148.

[2] J. Johns, K. Krogh, G.M. Rodriguez, et al., Management of neurogenic bowel dysfunction in adults after spinal cord injury: clinical practice guideline for health care providers, J Topics in spinal cord injury rehabilitation 27 (2021) 75–151, https://doi.org/10.46292/sci2702-75.

[3] J.M. Wecht, N.Y. Harel, J. Guest, et al., Cardiovascular autonomic dysfunction in spinal cord injury: epidemiology,Diagnosis,and management, J Seminars in neurology 40 (2020) 550–559, https://doi.org/10.1055/s-0040-1713885.

[4] M. Garnier-Villarreal, D. Pinto, C. Mummidisetty, et al., Predicting duration of outpatient physical therapy episodes for individuals with spinal cord injury based on locomotor training strategy, J Archives of physical medicine and rehabilitation 103 (2022) 665–675, https://doi.org/10.1016/j.apmr.2021.07.815.

[5] Y. Yuan, X. Yu, Therapeutic effects of rehabilitation training methods on spinal cord injury: a meta-analysis, J. Lancet 394 (2019) 27, https://doi.org/10.1016/S0140-6736(19)32363-3.
[6] D.I. Kim, H. Lee, B.S. Lee, et al., Effects of a 6-week indoor hand-bike exercise program on health and fitness levels in people with spinal cord injury: a randomized controlled trial study, J Archives of Physical Medicine & Rehabilitation 96 (2015) 2033–2040, https://doi.org/10.1016/j.apmr.2015.07.010.
[7] S. Rigot, L. Worobey, M.L. Boninger, Gait training in acute spinal cord injury rehabilitation–utilization and outcomes among non-ambulatory individuals: findings from the SCI rehab project, J Archives of Physical Medicine & Rehabilitation 99 (2018) 1591–1598, https://doi.org/10.1016/j.apmr.2018.01.031.
[8] R. El-Yafouri, L. Klieb, V. Sabatier, The impact of office-related metrics on meeting physician expectations from Electronic Medical Record systems, J Healthcare Analytics 4 (2023) 100208, https://doi.org/10.1016/j.health.2023.100208.
[9] M. Shapiro, R. Landau, S. Shay, et al., Early detection of COVID-19 outbreaks using textual analysis of electronic medical records, J Journal of Clinical Virology 155 (2022) 105251, https://doi.org/10.1016/j.jcv.2022.105251.
[10] Y.T. Kao, C.Y. Huang, Y.A. Fang, et al., Machine Learning-Based Prediction of Atrial Fibrillation Risk Using Electronic Medical Records in Older Aged Patients *J the American Journal Of Cardiology*, vol. 198, 2023, pp. 56–63, https://doi.org/10.1016/j.amjcard.2023.03.035.
[11] Y. Ma, X. Liu, L. Zhao, et al., Hybrid embedding-based text representation for hierarchical multi-label text classification, J Expert Systems with Applications 187 (2021) 115905, https://doi.org/10.1016/j.eswa.2021.115905.
[12] S. Coulibaly, B. Kamsu-Foguem, D. Kamissoko, et al., Deep Convolution Neural Network sharing for the multi-label images classification, J Machine Learning with Applications 10 (2022) 100422, https://doi.org/10.1016/j.mlwa.2022.100422.
[13] J.F. Samah, T. Aditya, Exploiting MEDLINE for gene molecular function prediction via NMF based multi-label classification, J Journal of Biomedical Informatics 86 (2018) 160–166, https://doi.org/10.1016/j.jbi.2018.08.009.
[14] C. Hiba, Z. Hamid, A. Omar, Multi-label transfer learning for the early diagnosis of breast cancer, J Neurocomputing 392 (2020) 168–180, https://doi.org/10.1016/j.neucom.2019.01.112.
[15] A. Jamthikar, D. Gupta, A.M. Johri, et al., A machine learning framework for risk prediction of multi-label cardiovascular events based on focused carotid plaque B-Mode ultrasound: a Canadian study, J Computers in Biology and Medicine 140 (2022) 105102, https://doi.org/10.1016/j.compbiomed.2021.105102.
[16] Md Ri, A. Arif, S.R. Mohammad, et al., Bangla text normalization for text-to-speech synthesizer using machine learning algorithms, J Journal of King Saud University - Computer and Information Sciences 36 (1) (2024) 101807, https://doi.org/10.1016/j.jksuci.2023.101807.
[17] F. Charte, A.J. Rivera, M.J. del Jesus, et al., MLSMOTE: approaching imbalanced multilabel learning through synthetic instance generation, J Knowledge-Based Systems 89 (2015) 385–397, https://doi.org/10.1016/j.knosys.2015.07.019.
[18] B. Liu, K. Blekas, G. Tsoumakas, Multi-label sampling based on local label imbalance, J Pattern Recognition 122 (2022) 108294, https://doi.org/10.1016/j.patcog.2021.108294.
[19] Rodolfo M. Pereira, M.G.C. Yandre, Carlos N. SJ. Mltl, A multi-label approach for the Tomek Link undersampling algorithm, J Neurocomputing 383 (2020) 95–105, https://doi.org/10.1016/j.neucom.2019.11.076.
[20] N.K. Mishra, P.K. Singh, Linear ordering problem based classifier chain using genetic algorithm for multi-label classification, J Applied Soft Computing 117 (2022) 108395, https://doi.org/10.1016/j.asoc.2021.108395.
[21] S. Moral-García, J.G. Castellano, C.J. Mantas, et al., A new label ordering method in Classifier Chains based on imprecise probabilities, J Neurocomputing 487 (2022) 34–45, https://doi.org/10.1016/j.neucom.2022.02.048.
[22] M. Jethanandani, A. Sharma, T. Perumal, et al., Multi-label classification based ensemble learning for human activity recognition in smart home, J Internet of Things 12 (2020) 100324, https://doi.org/10.1016/j.iot.2020.100324.
[23] J. Huang, G. Li, S. Wang, et al., Multi-label classification by exploiting local positive and negative pairwise label correlation, J Neurocomputing 257 (2017) 164–174, https://doi.org/10.1016/j.neucom.2016.12.073.
[24] A.Z. Huang, R. Xu, Y. Chen, et al., Research on multi-label user classification of social media based on ML-KNN algorithm, J Technological Forecasting and Social Change 188 (2023) 122271, https://doi.org/10.1016/j.techfore.2022.122271.
[25] C. Hou, F. Nie, C.S. Zhang, et al., Multiple rank multi-linear SVM for matrix data classification, J Pattern Recognition 47 (2014) 454–469, https://doi.org/10.1016/j.patcog.2013.07.002.
[26] S. Moral-García, C.J. Mantas, J.G. Castellano, J. Abellán, et al., Using credal C4.5 for calibrated label ranking in multi-label classification, J International Journal of Approximate Reasoning 147 (2022) 60–77, https://doi.org/10.1016/j.ijar.2022.05.005.
[27] J. Read, B. Pfahringer, G. Holmes, et al., Classifier chains for multi-label classification, J Machine Learning 85 (2011) 333–359, https://doi.org/10.1007/978-3-642-04174-7_17.
[28] J. Shan, C. Hou, H. Tao, et al., Co-learning binary classifiers for LP-based multi-label classification, J Cognitive Systems Research 55 (2019) 146–152, https://doi.org/10.1016/j.cogsys.2019.01.009.
[29] R. Wang, S. Kwong, X. Wang, et al., Active k-labelsets ensemble for multi-label classification, J Pattern Recognition. 109 (2021) 107583, https://doi.org/10.1016/j.patcog.2020.107583.
[30] P. Teisseyre, Classifier chains for positive unlabelled multi-label learning, J Knowledge-Based Systems 213 (2021) 106709, https://doi.org/10.1016/j.knosys.2020.106709.
[31] J.S. Lou, M.R. Jin, C.H. Zhou, et al., Ezrin inhibition alleviates oxidative stress and pyroptosis via regulating TRPML1-calcineurin axis mediated enhancement of autophagy in spinal cord injury, J. Free Radic. Biol. Med. 212 (2024) 133–148, https://doi.org/10.1016/j.freeradbiomed.2023.12.020.
[32] P. Marc, B. Rachel, B.M. Désirée, et al., A proposal for a universal physical therapy diagnostic concept, J Brazilian Journal of Physical Therapy 27 (6) (2023) 100560, https://doi.org/10.1016/j.bjpt.2023.100560.