ISSN: 2321-7782 (Online)
e-ISJN: A4372-3114
Impact Factor: 7.327

Volume 5, Issue 11, November 2017

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Multi-label Disease Diagnosis and Classification using Map Reduce

Ashwini Pawar[1]
Department of Computer Engineering
D. Y. Patil college of Engineering
India

Prof. Dhanashree Kulkarni[2]
Department of Computer Engineering
D. Y. Patil college of Engineering
India

Abstract: ***In Intensive Care Units (ICU) in the modern medical information scheme can keep the record of patient events in relational databases each second. Information mining from these enormous volumes of medical information is beneficial to equally caregivers as well as patients. Specified a set of electronic patient records, a scheme that efficiently gives the disease labels can enable medical database management as well as benefit other researchers, e.g. pathologists. Since, as data increasing day by day, thus to process on data is difficult. In this paper, a framework is proposed to achieve that goal by introducing Hadoop map reducing. Medical chart and note data of a patient are used to extract distinctive features. To encode patient features, a Bagof-Words encoding method is applied for both chart and note data. This paper also proposes model that takes into account both global information and local correlations between diseases. Associated diseases are considered by a graph structure that is embedded in proposed sparsity-based structure. The proposed algorithm captures the disease relevance when labeling disease codes rather than making individual decision with respect to a specific disease. In addition for evaluation purpose Naive Bays and Adaboost classifier are used for disease classification.***

Keywords: ***ICD code labeling, multi-label learning, sparsitybased regularization, disease correlation embedding, map Reduce.***

I. INTRODUCTION

Perfect understanding of a patient's disease state and the trajectory is serious in a clinical setup. Recent electronic healthcare records comprise a continuously increasing huge amount of records, and the capability to spontaneously identify the aspects that influence patient results viewpoint to significantly improve the effectiveness and quality of precaution. Doctors or physicians regularly need to retrieve related medical records for a patient in ICU for making better conclusions. The simplest approach is to input a collection of disease codes by means of diagnosis from the patient, into a system that can offer related cases rendering to the codes. The maximum famous and extensively used disease code system is the International Statistical Classification of Diseases and Related Health Problems (generally abbreviated as ICD) suggested and sometimes brush up by the World Health Organization (WHO). The newest version is ICD-10 is useful with native clinical changes in various regions. The objective of ICD is to deliver a unique hierarchical categorization system that is intended to map health circumstances to diverse categories. In the United States (US), the ICD9 has been persistently useful in numerous areas where disease classification is essential. For example, for every patient in ICU will be related to ICD9 list codes in the medical health records drives for example disease tracking, medical record information management or pathology. By examining the reverted historical information, caregivers are expected to suggest better cures to the patient. Therefore, complete as well as correct disease classification is very important. The ICD codes assignment to patients in ICU is usually prepared by caregivers in a hospital (for example nurses, physicians, and radiologists). This task might happen during or after admittance to ICU. In the earlier situation,

ICD codes are independently labeled by several caregivers during a patients stay in ICU as an end result of unlike work shifts time of a patients stay is typically greatly longer than the work time shift of the medical staff in a hospital. Therefore various caregivers are liable to create judgments rendering to the current situations. It is more necessary to assign a patients disease label by taking the complete patient record into the description. Once the assignment is accompanied afterward admission to ICU, the ICD codes are assigned by a specialized doctor who examines as well as reviews completely the records of a patient. Yet, it is still difficult for any physician expert to remember the connections of diseases when labeling a list of disease codes. But sometimes this process leads to missing code otherwise incorrect code classification. In fact, round about diseases is very much correlated. Correlations among diseases can increase the multi-label classification results.

The main focus in this study is to give disease labels to medical records of the patient. Instead of expecting the death risk of an ICU patient, the projected work can be observed as a multi-label prediction issue. The death risk prediction is a difficult binary classification in which the label designates the chances of survival. The problem of multi-label classification has continuously been an undefended but interesting problem in the machine learning as well as data mining communities. In presented model, the great attention is given to equally the medical chart also note data of patients. Medical chart records are similarly termed structured data since their structure is usually fixed. In the ICU, around famous health condition measurement scores are determined manually by staff in the ICU, rendering to the patients health situation. In contrast, medical chart records are raw copies mined from

the monitoring devices attached to a patient. The chart data hence return the physiological situations of a patient at a lower level. The patients note data has no structure since it is resulting from textual evidence. Thus, it is usually called as free-text note data. The benefits of these forms of data are that they are expressive and instructive because they are brief or determined by specialists. Though, medical note records are actually difficult to manage by various existing machine learning algorithms since no one of the structures in the notes can be openly recognized as patterns. Thus, medical notes are somewhat noisy, as well as their quality is often degraded by spelling mistake or abbreviations. Furthermore, the contents of medical notes are not permanently constant with the metrics.

Section II gives the essential literature survey. Section III addresses existing system. Section IV introduces the proposed architecture overview. Section V describes expectations predictable results. Section VI accomplishes the paper.

II. Review of Literature

In the literature review, topical methods over the disease diagnosis techniques are discussed.

M. Ghassemi et al. [1] observed the use of variable model called as Latent Dirichlet Allocation to convert free-textual hospital notes into significant features as well as the predictive control of these features for patient mortality risk.

A. E. Johnson et al. [2] tried to decrease the number of physiologic constraints collected in the Acute Physiology, Age, and Chronic Health Evaluation IV scheme deprived of losing predictive accurateness via a machine-learning method identified as particle swarm optimization.

O. Frunza [3] defines a ML-based approach for constructing an application that is capable of recognizing as well as distributing healthcare information. This approach extracts sentences from medical papers which are published and mention diseases with the treatments and identifies semantic associations that exist among diseases and treatments.

Y. Park and J. Ghosh [4] presents two types of decision tree ensembles for class imbalanced problems, widely using properties of (_)-divergence. They first present splitting standard based on (_)-divergence to generalize numerous well-known splitting conditions for example those used in C4.5 classifier and CART classifier. Other ensemble usages the equal alpha trees as base classifiers then use a lift-aware stopping criterion throughout tree growth.

P. Ordonez et al. [5] offered two multivariate time series illustrations to classify physiological information of different sizes. The illustrations might be applied to some collection of multivariate time series information that observes the health of an individual. These illustrations are (1) Multivariate Bagof-Patterns and (2) Stacked Bags-of-Patterns which inspired by the bag-of-words model. These illustrations progress on their univariate counterpart by spending multiple time series as well as examining the data in a multivariate manner.

F. Wang et al. [6] suggest a Nonnegative Matrix Factorization (NMF) framework via a convolutional method for open-ended chronological pattern discovery above large gatherings of clinical histories. This method is called as One-Sided Convolutional NMF (OSC-NMF).

J. Read et al. [7] presented a chaining technique for multilabel classification based on the binary relevance technique, which claimed has numerous benefits over extra sophisticated current methods, especially in terms of time expenses. By passing label correspondence data together with a chain of classifiers, their technique counteracts the drawbacks of the binary technique while preserving acceptable computational convolution.

G. Tsoumakas et al. [8] suggests a technique in which the initial set of labels are break into a number of insignificant random subsets, named label sets also employing LP on the way to train a corresponding classifier. The projected technique is named RAkEL (RAndom k labELsets), where k is a parameter that identifies the size of the subsets.

Y. Yang et al. [9] suggest an algorithm for multi-task feature selection then apply it to multimedia (for instance, video and image) study. Rather than evaluating the significance of every feature independently, their presented algorithm selects features in a batch mode, to consider the feature connection. The algorithm constructs upon the guess that dissimilar connected tasks have collective structures. Multiple feature assortment functions of dissimilar tasks are concurrently learned in a joint framework, which allows algorithm to develop the collective knowledge of multiple tasks as additional data to facilitate decision making.

X. Chang et al. [10] suggest a framework for semisupervised feature selection by extracting correlations between multiple tasks as well as apply it to various multimedia applications. This algorithm influences shared information from multiple associated tasks, therefore the performance of feature selection gets increase.

X. Zhu et al. [11] conduct a multiview multilabel (MVML) learning and a hierarchical feature selection for multiview image classification, through embedding a projected a blockrow regularizer into the MVML framework.

A. Perotte et al. [12] suggest evaluation metrics that reflect the distances between gold-standard as well as predicted codes and their locations in the ICD9 tree.

D. Zufferey et al. [13] suggest to procedure a hierarchy-based SVM model on MIMIC II dataset to manner classification of automated diagnosis code.

Zufferey et al. [14] compare dissimilar multi-label algorithms of classification for chronic disease classification as well as highlight the hierarchy-based SVM model to achieved superior capacity than other approaches when accuracy is very significant.

X. Kong et al. [15] apply heterogeneous data networks on a dataset named bioinformatic for two multi-label classification tasks (1) gene-disease association prediction and (2) drugtarget binding prediction by developing relationships among different kinds of entities.

O. Frunza et al. [16] defines a ML-based method for constructing an application that is efficient of classifying and disseminating healthcare data. It mines and identifies sentences from published medical papers that mention diseases and treatments.

F. Wang et al. [17] propose a Nonnegative Matrix Factorization (NMF) based system utilizing a convolutional approach for open-ended transient example disclosure over vast accumulations of clinical records call this strategy One-Sided Convolutional NMF (OSC-NMF). This structure can mine basic and additionally singular shift-invariant transient examples from heterogeneous events over various patient gatherings, and handle sparsity and also versatility issues well. Besides, an occasion network based representation is utilized that can encode quantitatively all key transient ideas including order, simultaneousness, and synchronicity.

S. Wang et al. [18] propose an algorithm for automatic video annotation by incorporating semi- supervised learning as well as shared structure investigation into a joint system for human activity acknowledgment. They apply calculation on both engineered and reasonable video datasets, including KTH (Kungl Tekniska Hgskolan Royal Institute of Technology in Stockholm), CareMedia dataset, Youtube activity and its amplified form, UCF50.

Y. Yang et al. [19] display another structure for multimedia content investigation and recovery which comprises of two autonomous algorithms: (1) the semi- supervised algorithm called ranking with Local Regression and Global Alignment (LRGA) to take in a strong Laplacian matrix for information ranking, (2) a semi- supervised long-term Relevance Feedback (RF) algorithm to refine the media information representation.

Z. Ma et al. [20] propose a novel feature selection method and apply it to automatic image annotation. There are two appealing properties of our method. First, it can jointly select the most relevant features from all the data points by using a sparsity-based model. Second, it can uncover the shared subspace of original features, which is beneficial for multi-label learning. To solve the objective function of our method, we propose an efficient iterative algorithm. Extensive experiments are performed on large image databases that are collected from the web.

Z. Mama et al. [20] propose a feature determination technique and apply it to automatic image annotation. There are two appealing properties of given strategy: (1) it can mutually choose the most pertinent features from every one of the information focuses by utilizing a sparsity-based model, (2) it can reveal the mutual subspace of unique components, which is useful for multi-name learning. To resolve the target capacity of projected strategy, efficient iterative algorithm is proposed.

J. Perused et al. [22] demonstrate that twofold relevancebased techniques have much to offer, particularly regarding adaptability to extensive datasets. They represent this with a chaining technique that can display label connections while keeping up adequate computational complexity.

G. Tsoumakas et al. [23] proposes breaking the underlying set of labels into various little arbitrary subsets, called label sets and utilizing LP to prepare a comparing classifier. The label sets can be either disjoint or covering relying upon which of two methodologies is utilized to develop them. The proposed strategy is called RAkEL (RAndom k labELsets), where k is a parameter that determines the measure of the subsets.

M. Saeed et al. [24] report the foundation of the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) explore database that is striking for four variables: it is openly and uninhibitedly accessible to other research associations upon demand; it envelops an diverse population of intensive care unit (ICU) patients; it contains high transient determination information, including lab results about, electronic clinical documentation, and bedside screen numeric patterns and waveforms, (for example, the electrocardiogram); and it has been deidentified in a Health Insurance Portability and Accountability Act-agreeable way.

P. L. Whetzel et al. [25] examine NCBO that has created BioPortal, a web-based interface that gives access to a library of biomedical ontologies and phrasings by means of the NCBO Web administrations. BioPortal empowers group support in the assessment and advancement of philosophy substance by giving features to including mappings between terms, to add remarks connected to particular ontology terms and to give ontology surveys.

M.- L. Zhang and L. Wu [26] propose system to multimark learning by utilizing label particular features, where a straightforward yet successful algorithm named LIFT is introduced. Quickly, LIFT develops highlight particular to each label by leading clustering analysis on its positive and negative occurrences and after that performs training and testing by querying the grouping results.

III. EXISTING APPROACH

A. Existing System Overview

Medicinal notes are very noisy, and their quality is regularly defiled by incorrect spellings or abbreviations. Additionally, the substances of medicinal notes are not generally reliable with the measurements. For instance, extraordinary caregivers take notes in various metrics when recording a parameter. Some want to utilize English units while others utilize the American framework (e.g. patient's temperature in Celsius versus Fahrenheit). In this way, compared with structured information, it is hard to remove precise and steady features from notes. It is consequently troublesome for medicinal notes to be used by machine learning algorithms.

B. Drawbacks of Existing Approach

Disadvantages of existing system are illustrated as:

- The multi-label classification issue has continuously been an open but challenging problem in the machine learning as well as data mining communities.
- Medical notes are very noisy, and their quality is frequently undermined by incorrect spellings or shortened forms. In addition, the substance of medicinal notes is not generally predictable with the metrics. In this way, contrasted with structured information, it is hard to remove exact and predictable components from notes.

IV. PROPOSED ARCHITECTURE

To address the previously mentioned issues, this paper proposes a system that will assign disease labels consequently while all the while considering relationships among diseases. In this, first medical information is extracted from two unique perspectives, organized and unstructured. Structured information can define patients' raw health situations from medical gadgets at a lower level, while unstructured information comprise of more semantic data at a more higher level which has turned out to be useful for characterizing features of patients for some expectation assignments. A BoW model is utilized to change over features of various lengths into a unique representation for every patient. Likewise, comparability examination can be led by supervised learning algorithms.

To above and beyond, an algorithm to classify disease labels with the assistance of the basic relationships among diseases is displayed. This paper proposes a framework that will assign disease labels automatically while simultaneously considering correlations between diseases. To step further, an algorithm is proposed to classify disease labels with the help of the underlying correlations between diseases. A large number of patient records are applied on this database in the evaluation. Therefore, Hadoop Map reduce is used for big data evaluation which uses parallel process that reduces time complexity. Label of disease is predicted using Nave bays and Adaboost algorithm.

A. Proposed Architecture Diagram

Figure 1 shows the proposed approach for diases diagnosis of testing data with the help of training data.

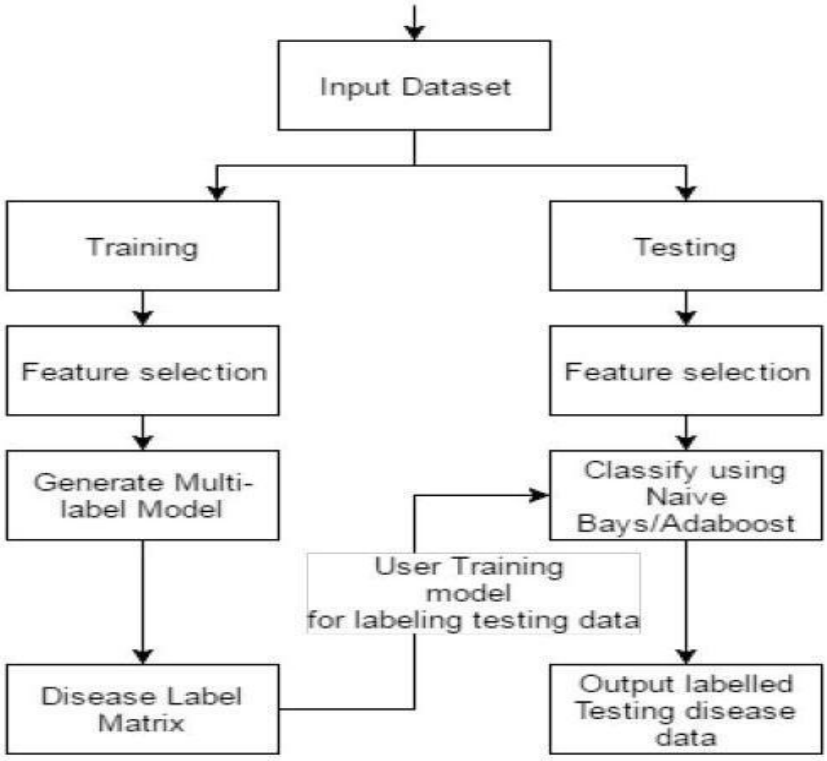

Fig. 1. Architecture Diagram of Proposed System

V. PROPOSED SYSTEM SETUP

A. Input

$x = [x_1; :::::; x_n]\ 2\ R^{(d+1)\ n}$ training dataset, where, n is the number of training samples.

$Y = [y_1; :::::; y_n]^T\ 2\ R^{n\ c}$ class indicator matrix. Where, c is the number of classes.

If x_i belongs to the j^{th} class, y_{ij} is 1, else $y_{ij} = 0$,

i 2 f1; :::; ng and j 2 1; :::; cg, where, c is number of classes.

1) Design objective function as follow:

$$\min_{w_i} \sum_{i=1}^{n} \sum_{j=1}^{P_{ij}} \log(1 + \exp(-y_{ij} w_i^T x_j)) + \sum_{i=1}^{c} \sum_{j=1}^{P} a\ Tr([w_i; w_j]^T D^{ij} [w_i; w_j])$$

Where, D^{ij} is a diagonal matrix with the d-th diagonal element.

Where, regularization parameter is 0

w_i and w_j are i^{th} and j^{th} column of coefficient matrix.

2) Apply Nave Bayes and Adaboost classifier on data for classification of patient record testing dataset.

B. Algorithm to solve the problem of objective function

Input: Data $x\ 2\ R^{(d+1)\ n}$, papramenters and k, Label corelation matrix $A\ 2\ R^{c\ c}$ which reflects the relationships between two arbitrary classes (diseases). Proess:

1) Randomly initialize coefficient matrix W

2) repeat following steps from 3 to 5

3) for each i and j, calculate Diagonal matrix Dij , where d^{th} diagonal element is

$$\frac{1}{zk[w_i;w_j]^d k_2}$$

4) for each i, calculate the diagonal matrix Q^i by $Q^i = 2\ ^P{}_j\ a_{ij}D^{ij}$

5) For each i, update w_i using equation $w_i^{t+1} = w_i^t + f5_{wi}L(w_i) + 5_{w\,i}\ (w_i)g$ where, is learning rate which is greater than 0, t is step index, $L(w_i)$ and (w_i) are differentiable w.r.t. w_i.

VI. Analysis and Results

A. Dataset

For evaluation, Diabetes dataset with 10,000 records is used to compare results. Dataset includes over 50 features representing patient and hospital outcomes. The data contains such attributes as patient number, race, gender, age, admission type, and time in hospital, medical specialty of admitting physician, and number of lab test performed, HbA1c test result, diagnosis, number of medication, diabetic medications, and number of outpatient, inpatient, and emergency visits in the year before the hospitalization, etc.

B. Expected results

To evaluate the performance, Multi-label annotation model is compare with Naive bays and Adaboost classifier. The expected results are evaluated according to classification outcome and time complexity using map reduce parallel processing.

Figure 2 shoes the accuracy requires for data when size of data is changed by comparing classifier. This comparison is performed to analyses better dataset for classification. From evaluation, expected nave bays classifier gives better classification results as compare to Multi-label annotation model and Adaboost. table 1 shows the redings of accuracies computed for algorithms.

TABLE I ACCURACY COMPARISION BETWEEN ALGORITHMS

Dataset Records	MLAM	Nave Bays	Adaboost
2000	94	94	92
4000	88	90	91
6000	83	84	81
8000	94	95	94
10000	78	84	79

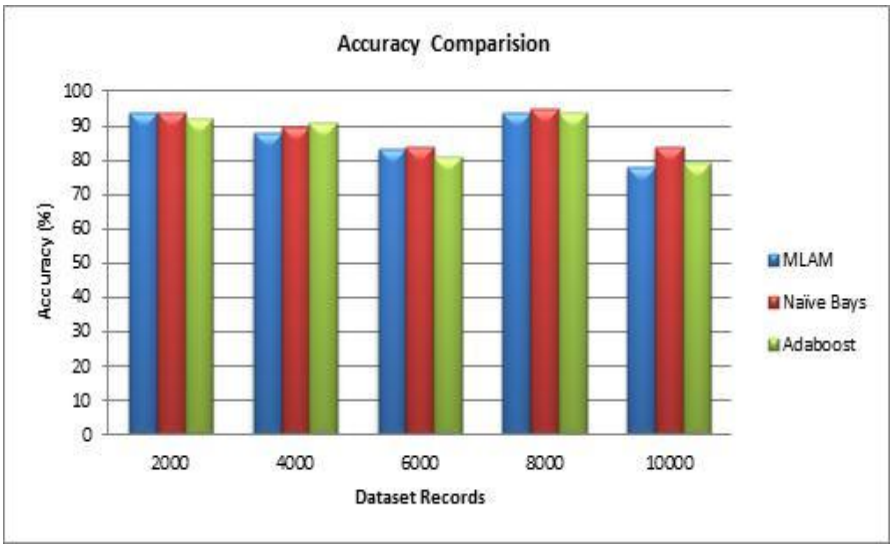

Fig. 2. Accuracy comparision graph

Time evaluation is analyse by comparing time require for each algorithm using map reduce and without map reduce. The expected results shoes that time require for algorithm processing using map reduce concept is very less as compare to processing without map reduce when data more than 5000 records. Table 2 shows the records for comparative time require to process algorithm. And figure 3 shows the graph for time requires without map reduce and with map reduce technique.

TABLE II TIME REQUIRE WITHOUT MAP REDUCE AND WITH MAP REDUCE

Algorithms	With Map Re duce	Without Map Reduce
MLAM	2048	975
Nave Bays	2107	894
Adaboost	2104	985

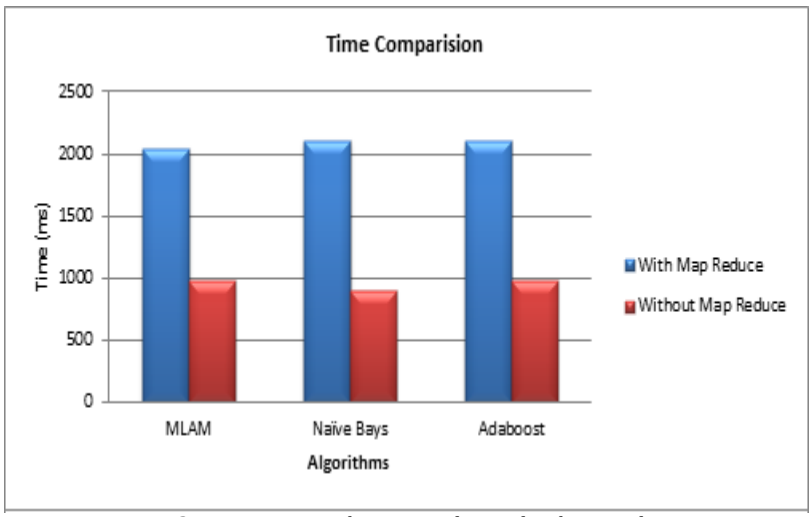

Fig. 3. Time require without map reduce and with map reduce

VII. CONCLUSION

This paper concentrates on disease labels assignment to patients' medicinal records. A system is proposed to accomplish that objective by presenting Hadoop map reducing. Medicinal note and charts information of a patient are utilized to remove unmistakable elements. To encode understanding elements, a Bag-of-Words encoding strategy is applied for both note and graph information. This paper likewise proposes model that considers both model data and local relationships among diseases. Related diseases are considered by a chart structure that is inserted in proposed sparsity-based structure. The proposed algorithm catches the disease pertinence while labeling disease codes as instead of settling on individual decision concerning a particular disease. In addition for evaluation purpose Naive Bays and Adaboost classifier are used for disease classification. Results are evaluated on the basis of time and accuracy.

ACKNOWLEDGEMENT

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance.

References

1. M. Ghassemi, T. Naumann, F. Doshi-Velez, N. Brimmer, R. Joshi, A. Rumshisky, and P. Szolovits, Unfolding physiological state: Mortality modelling in intensive care units, in ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2014, pp. 7584.
2. A. E. Johnson, A. A. Kramer, and G. D. Clifford, A new severity of illness scale using a subset of acute physiology and chronic health evaluation data elements shows comparable predictive accuracy, Critical Care Medicine, vol. 41, no. 7, pp. 17111718, 2013.
3. O. Frunza, D. Inkpen, and T. Tran, A machine learning approach for identifying disease-treatment relations in short texts, IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 6, pp. 801814, 2011.
4. Y. Park and J. Ghosh, Ensembles of -trees for imbalanced classification problems, IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, pp. 131143, 2014.
5. P. Ordonez, T. Armstrong, T. Oates, and J. Fackler, Using modified multivariate bag-of-words models to classify physiological data, in IEEE International Conference on Data Mining Workshop, Dec 2011, pp. 534539.

6. F. Wang, N. Lee, J. Hu, J. Sun, and S. Ebadollahi, Towards heterogeneous temporal clinical event pattern discovery: A convolutional approach, in ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2012, pp. 453461.
7. J. Read, B. Pfahringer, G. Holmes, and E. Frank, Classifier chains for multi-label classification, Machine learning, vol. 85, no. 3, pp. 333359, 2011.
8. G. Tsoumakas, I. Katakis, and L. Vlahavas, Random k-labelsets for multilabel classification, IEEE Transactions on Knowledge and Dat Engineering, vol. 23, no. 7, pp. 10791089, July 2011.
9. Y. Yang, Z. Ma, A. G. Hauptmann, and N. Sebe, Feature selection for multimedia analysis by sharing information among multiple tasks, IEEE Transactions on Multimedia, vol. 15, no. 3, pp. 661669, 2013.
10. X. Chang, H. Shen, S. Wang, J. Liu, and X. Li, Semi-supervised feature analysis for multimedia annotation by mining label correlation, in Advances in Knowledge Discovery and Data Mining -18th Pacific-Asia Conference, PAKDD 2014, Tainan, Taiwan, May 13-16, 2014. Proceedings, Part II, 2014, pp. 7485.
11. X. Zhu, X. Li, and S. Zhang, Block-row sparse multiview multilabel learning for image classification, IEEE Transactions on Cbernetics, vol. 46, no. 2, pp. 450461, 2016.
12. A. Perotte, R. Pivovarov, K. Natarajan, N. Weiskopf, F. Wood, and N. Elhadad, Diagnosis code assignment: models and evaluation metrics, Journal of the American Medical Informatics Association, vol. 21, no. 2, pp. 231237, 2014.
13. D. Zufferey, T. Hofer, J. Hennebert, M. Schumacher, R. Ingold, and S. Bromuri, Performance comparison of multi-label learning algorithms on clinical data for chronic diseases, Computers in biology and medicine, vol. 65, pp. 3443, 2015.
14. J. C. Ferrao, F. Janela, M. D. Oliveira, and H. M. Martins, Using structured ehr data and svm to support icd-9-cm coding, in Healthcare Informatics (ICHI), 2013 IEEE International Conference on. IEEE, 2013, pp. 511516.
15. X. Kong, B. Cao, and P. S. Yu, Multi-label classification by mining label and instance correlations from heterogeneous information networks, in ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2013, pp. 614622.
16. O. Frunza, D. Inkpen, and T. Tran, A machine learning approach for identifying disease-treatment relations in short texts, IEEE Transaction on Knowledge and Data Engineering, vol. 23, no. 6, pp. 801814, 2011.
17. F. Wang, N. Lee, J. Hu, J. Sun, and S. Ebadollahi, Towards heterogeneous temporal clinical event pattern discovery: A convolutional approach, in ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2012, pp. 453461.
18. S. Wang, Y. Yang, Z. Ma, X. Li, C. Pang, and A. G. Hauptmann, Action recognition by exploring data distribution and feature correlation, in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012, pp. 13701377.
19. Y. Yang, F. Nie, D. Xu, J. Luo, Y. Zhuang, and Y. Pan, A multimedia retrieval framework based on semi-supervised ranking and relevance feedback, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 4, pp. 723742, 2012.
20. Z. Ma, F. Nie, Y. Yang, J. R. Uijlings, and N. Sebe, Web image annotation via subspace-sparsity collaborated feature selection, IEEE Transactions on Multimedia, vol. 14, no. 4, pp. 10211030, 2012.
21. J. Read, B. Pfahringer, G. Holmes, and E. Frank, Classifier chains for multi-label classification, Machine learning, vol. 85, no. 3, pp. 333359, 2011.
22. G. Tsoumakas, I. Katakis, and L. Vlahavas, Random k-labelsets for multilabel classification, IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 7, pp. 10791089, July 2011.
23. M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L.-W. Lehman, G. Moody, T. Heldt, T. H. Kyaw, B. Moody, and R. G. Mark, Multiparameter intelligent monitoring in intensive care ii (mimicii): A public-access intensive care unit database, Critical Care Medicine, vol. 39, no. 5, p. 952, 2011.
24. P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, and M. A. Musen, Bioportal: enhanced functionality via new web services from the national center for biomedical ontology to access and use ontologies in software applications, Nucleic acids research, vol.39, no. suppl 2, pp. W541W545, 2011.
25. M.-L. Zhang and L. Wu, Lift: Multi-label learning with labelspecific features, in International Joint Conference on Artificial Intelligence, 2011, pp. 16091614.

