Computer Methods and Programs in Biomedicine 159 (2018) 87–101

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Decision based on big data research for non-small cell lung cancer in medical artificial system in developing country

Jia Wu[a,c,d,*], Yanlin Tan[b,**], Zhigang Chen[c,d], Ming Zhao[c,d]

[a] School of information science and engineering, Central South University, Changsha 410083, China
[b] PET-CT Center, The Second Xiangya Hospital of Central South University, Changsha 410083, China
[c] "Mobile Health" Ministry of Education-China Mobile Joint Laboratory, Changsha 410083, China
[d] School of Software, Central South University, Changsha 410083, China

ARTICLE INFO

Article history:
Received 29 June 2017
Revised 22 January 2018
Accepted 9 March 2018

Keywords:
NSCLC
PET–CT
Medical decision making
Big data
Image
Diagnostic parameters

ABSTRACT

Non-small cell lung cancer (NSCLC) is a high risk cancer and is usually scanned by PET–CT for testing, predicting and then give the treatment methods. However, in the actual hospital system, at least 640 images must be generated for each patient through PET–CT scanning. Especially in developing countries, a huge number of patients in NSCLC are attended by doctors. Artificial system can predict and make decision rapidly. According to explore and research artificial medical system, the selection of artificial observations also can result in low work efficiency for doctors. In this study, data information of 2,789,675 patients in three hospitals in China are collected, compiled, and used as the research basis; these data are obtained through image acquisition and diagnostic parameter machine decision-making method on the basis of the machine diagnosis and medical system design model of adjuvant therapy. By combining image and diagnostic parameters, the machine decision diagnosis auxiliary algorithm is established. Experimental result shows that the accuracy has reached 77% in NSCLC.

1. Introduction

In developing countries, lung cancer has become the first leading cause of death in these years, which has experienced a dramatic increase in the cigarette smoking rate during the past 2 decades [1–3]. Non-small cell lung cancer (NSCLC) accounts for 85% of lung cancer and the 5-year survival rate is only 15%. Approximately 70% of patients with lung cancer commonly present with locally advanced or distant metastasis at the time of diagnosis, which was a difficult condition to manage due to the lack of effective treatments [4]. However, if a patient can be diagnosed in the early stage, the 5-year survival rate can be raised to 80% [5]. So the early diagnosis of this type of lung cancer has important significance for the prognosis of the patient [16–18].

In NSCLC, conventional clinical staging is most often performed with computed tomography (CT) of the thorax and upper abdomen. However, CT imaging has limited sensitivity for distal metastatic disease and is frequently unable to discriminate between malignant and benign lymph nodes. As a noninvasive and useful inspection method, 18F-FDG PET/CT is commonly referred for evaluation of primary neoplastic lesions and exploration of any possible metastasis. It has greater sensitivity for the detection of metabolically active malignant disease and can lead to changes in initial staging and treatment plans for NSCLC [5–7].

In many developing countries, doctors, hospitals and governments must face two serious problems.

1.1. Limit medical recourses and a great number of population

In developing countries, people's lives may not be protected because the medical technology may be underdeveloped. At the same time, the population is large, limit medical recourses can cover some parts of them. China is one of the developing countries in the world. There is a statistics data in 2016 from China's ministry of health. The population had reached 1.45 billion. Over 5600 people may share one doctor, and a doctor may treat 72 patients one day. Another statistics data is that a hospital in a big city has treatment over 1 million people a year; a better hospital must be received 3.5 million treatments per year. From 2011 to 2016, patients in lung cancer have reached over 83 million. Not all the patients can be cared by doctors and hospitals.

According to the data came from China's ministry of health, limited medical resources are mainly distributed in large cities and developed areas. The population of these cities and hospitals is

* Corresponding author at: School of information science and engineering, Central South University, Changsha 410083, China.
** Corresponding author.
E-mail addresses: jiawu5110@163.com (J. Wu), eric.tyanlin@gmail.com (Y. Tan), czg@mail.csu.edu.cn (Z. Chen), meanzhao@mail.csu.edu.cn (M. Zhao).

https://doi.org/10.1016/j.cmpb.2018.03.004
0169-2607/

300 million. Only 7% of the population receives 80% of medical resources, and 93% only receives 20% less of the medical resources.

If artificial medical system can be adopted by hospitals, more patients can be treated in developing countries.

1.2. Limit artificial medical technology and complicated diagnosis process

In the actual hospital system, at least 640 images must be generated for each patient through PET–CT scanning. Selecting all the images in an artificial way can be inefficient for doctors. If doctors must distinguish each image, only one or two patients can be looked over a day. In effect, each doctor has to diagnose over 40 patients per day in many hospitals.

At the same time, the 640 images also require at least two gigabytes of storage, and some of these images can be adopted by doctors' diagnosis, thereby resulting in a waste of resources for the medical system.

An effective machine scanning image analysis can be conducted to select valuable images of patients as the research foundation. Especially, primary prediction and decision based on lung cancer may decrease workload by doctors and severs.

On this basis, intelligent diagnosis methods can be established using the diagnosis of computer system parameters and decision-making results as a basis treatment. As a result, the doctors can use the machine auxiliary diagnosis results as a reference to ensure accurate diagnosis and effective treatment, improve the doctors' working efficiency effectively, and reduce misdiagnosis rate.

In this study, data information of 2,789,675 patients in three hospitals in China are collected, compiled, and used as the research basis; these data are obtained through image acquisition and diagnostic parameter machine decision-making method on the basis of the machine diagnosis and medical system design model of adjuvant therapy. Combining image and diagnosis parameters is an effective approach for doctors to solve patients' diagnosis in large data environment.

The contributions of the study are summarized as follows:

(1) A clustering range calculation model based on the image recognition method is established.
(2) Prediction and decision algorithm is established by the concentrated correlation parameter of non-small cell lung cancer (NSCLC).
(3) The recommended model of treatment is designed through decision analysis in big data research.
(4) According to deep learning research in big data, we may improve accuracy by prediction and decision in medical system.

2. Related works

Artificial medical decision making system have become hot research directions in the field of medical treatment. Many research methods are widely applied in the medical field.

Literature [7] designed eXiTCDSS medical decision support system. This system uses a case-based reasoning engine to retrieve similar cases. In eXiTCDSS, cases are stored in a Comma-Separated Value (CSV) format. A case consists of multiple attributes; each property is represented by a column in the CSV. The property type includes the Boolean, text type, and type. The weight of each type is per-allocated to case similarity calculation. In this system, the cases in every attribute are associated with the elements in the clinical diagnosis and treatment process; therefore, eXiTCDSS is mainly used in medical decisions to support workflow.

Cheng [17] assumes that cases, such as production function and similar case retrieval methods, are successful based on case reasoning method integrated into the key to Hospital Information System (HIS). In his study, case data structures are defined and modified by the doctor. Case data were extracted from the patient's electronic medical records in order to realize the reuse of medical experience. When a new patient is enrolled into the system, the system uses the weighted *K* nearest neighbor algorithm to retrieve the most similar cases. Cases benefit from the production function, which enhances the flexibility of knowledge extraction; however, the workload of doctors in the maintenance of the case library is certainly increased.

The diagnosis of the disease is usually through the patient's judgment of symptoms. Thus, in addition to the case form, case reasoning system is another focus for research to improve work efficiency by doctors.

Literature [8] compares the cases based on the inductive and deductive reasoning characteristics and put forward a combination of advantages from both systems to support the diagnosis and treatment process. To increase the basis for case reasoning method, Literature [9] recommended treatment availability and used the explanation in text form of the relationship between the patient and the explained recommended precedent. Literature [10] finds that based on case reasoning and other methods, such as BP (Brief Introduction of Back Propagation) neural network, the combination of models has a better performance in liver disease diagnosis.

Literature [11] introduced the time series data of breathing patterns based on case reasoning to improve diagnosis decision making. By integrating first the system in HIS(Hospital Information System) of knowledge, the discovered model defines a series of breathing patterns related to the diagnosis and calculates the new breathing pattern of the patient and the similarity system classification model in advance to obtain the final diagnosis.

Literature [12] proposed a method based on text similarity and on the use of Word Net literature [13]. This paper proposes a method based on the dictionary similarity calculation method of similarities between entities in different ontologies. In addition, a kind of algorithm, which is based on the rules of ontology matching algorithm, is the core idea that uses the association rules of discovery algorithm and finds hidden relevance in ontology. In literature [14], the inclusion relation existing in the real world was concluded to be far greater than equivalence relation; thus, the discovery of the hierarchical relationships between things is important. Therefore, it puts forward a hybrid, extensible, and asymmetric matching algorithm. Through association rule mining, this algorithm can determine the level of the relationship between entities.

In the literature [15], the author discussed the difference between open and closed world; this paper further proposed a horn rule mining method based on the open world assumption, which was used to realize the heterogeneous knowledge-based identity matching. However, this method of gaining confidence in association rules is often inaccurate, which leads to the emergence of a large number of false connections; thus, its practical application was not given attention.

In this paper, we hope to establish an artificial medical system to solve prediction and decision based on big data research in NSCLC

3. System design

3.1. The master model of prediction and decision tree

We model the diagnostic decision flow using six sequential stages: parameter selection, module reconstitution and data matching, pre-processing of the collected data for machine learning models (MLMs) pre-trained using MLAs, decision making through MLMs, obtaining disease signatures, and responding according to the decisions. Diagnosis of disease *i* is done through its