Artificial Intelligence in Medicine xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

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

A novel method for predicting kidney stone type using ensemble learning

Yassaman Kazemi, Seyed Abolghasem Mirroshandel*

Department of Computer Engineering, University of Guilan, Rasht, Iran

ARTICLE INFO

Article history:
Received 9 June 2017
Received in revised form 30 July 2017
Accepted 2 December 2017

Keywords:
Kidney disease
Data mining
Classification technique
Ensemble learning
Kidney stone

ABSTRACT

The high morbidity rate associated with kidney stone disease, which is a silent killer, is one of the main concerns in healthcare systems all over the world. Advanced data mining techniques such as classification can help in the early prediction of this disease and reduce its incidence and associated costs. The objective of the present study is to derive a model for the early detection of the type of kidney stone and the most influential parameters with the aim of providing a decision-support system. Information was collected from 936 patients with nephrolithiasis at the kidney center of the Razi Hospital in Rasht from 2012 through 2016. The prepared dataset included 42 features. Data pre-processing was the first step toward extracting the relevant features. The collected data was analyzed with Weka software, and various data mining models were used to prepare a predictive model. Various data mining algorithms such as the Bayesian model, different types of Decision Trees, Artificial Neural Networks, and Rule-based classifiers were used in these models. We also proposed four models based on ensemble learning to improve the accuracy of each learning algorithm. In addition, a novel technique for combining individual classifiers in ensemble learning was proposed. In this technique, for each individual classifier, a weight is assigned based on our proposed genetic algorithm based method. The generated knowledge was evaluated using a 10-fold cross-validation technique based on standard measures. However, the assessment of each feature for building a predictive model was another significant challenge. The predictive strength of each feature for creating a reproducible outcome was also investigated. Regarding the applied models, parameters such as sex, acid uric condition, calcium level, hypertension, diabetes, nausea and vomiting, flank pain, and urinary tract infection (UTI) were the most vital parameters for predicting the chance of nephrolithiasis. The final ensemble-based model (with an accuracy of 97.1%) was a robust one and could be safely applied to future studies to predict the chances of developing nephrolithiasis. This model provides a novel way to study stone disease by deciphering the complex interaction among different biological variables, thus helping in an early identification and reduction in diagnosis time.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Nowadays, knowledge gathering and data mining processes to obtain useful knowledge from data are considered a necessity. As a result, data mining is the most important tool for an efficient utilization of diverse sources of data [1]. The collection of data for various diseases is of considerable importance in medicine. Medical centers perform this task to obtain valuable information from their patients. The analysis of patients' data to unearth a trend of diseases is the prime goal of data collection. However, the enormous size of data and, consequently, its associated ambiguities cast a shadow on achieving reliable results. Data mining can eliminate this shortcoming and could be applied to obtain important parameters to predict diseases [2].

In other words, data mining can help predict the disease risk factor for patients by using their medical history and plays a key role in their healthcare. In fact, when the number of parameters in disease diagnosis increases, the diagnosis and prognosis become harder even for a qualified medical professional [2]. Hence, in the last few decades, disease diagnostics have been developed to help physicians reduce medical errors caused by inexperience, fatigue, stress, and heavy workload. Therefore, by using data mining algorithms, the required medical data can be analyzed by physicians within a shorter time and with greater precision and detail [3].

Kidney diseases, especaially nephrolithiasis (kidney stones), are widespread among people in all nations. The application of data mining tries to find a predictive model for the identification of nephrolithiasis patients and also the type of stones, thus helping

* Corresponding author at: Department of Computer Engineering, Faculty of Engineering, University of Guilan, Rasht, P.O. Box: 1841, Iran.
E-mail address: mirroshandel@guilan.ac.ir (S.A. Mirroshandel).

https://doi.org/10.1016/j.artmed.2017.12.001
0933-3657/© 2017 Elsevier B.V. All rights reserved.

Please cite this article in press as: Kazemi Y, Mirroshandel SA. A novel method for predicting kidney stone type using ensemble learning. Artif Intell Med (2017), https://doi.org/10.1016/j.artmed.2017.12.001

in early diagnosis, treatment, and a reduction in the effects of this disease [4].

Although various studies have been done in this field, we still lack a comprehensive predictive model that can identify all the essential features of kidney stones. Detection of the type of kidney stone needs different tests. This model can help physicians predict the type of existing kidney stones, based on the general characteristics of each patient.

Since the number of people suffering from kidney diseases is increasing, and data mining algorithms can help us predict these diseases, based on existing data, we decided to employ data mining techniques for prediction. To the best of our knowledge, there is no study on the prediction of the type of kidney stone using data mining algorithms. Therefore, the focus of this paper is to provide a dataset, which can be leveraged to predict the type of kidney stone before essential experiments are done on patients.

Thanks to innovative technologies, factors of illnesses are better measured and recorded. Moreover, using low-cost computer hardware and software technologies, high-quality data of patients is being collected and stored automatically. This body of valuable data makes researchers employ them to predict different diseases. Accordingly, the prediction of kidney diseases using data mining techniques has also become a challenging topic on which some studies have been done in recent years. These studies were performed to predict different kidney diseases involving kidney stone, kidney transplantation, kidney failure, dialysis, and hemodialysis.

In a study by Kumar and Abhishek on the diagnosis the kidney stone disease, a comparison of three neural network algorithms such as learning vector quantization (LVQ), multilayer perceptron (MLP), and radial basis function (RBF) was done to ascertain their level of accuracy, training dataset size, and the time consumed to build a model. Among the studied algorithm, MLP, with 92% accuracy, was selected as the best model that helped in the early detection of kidney stones in patients and in reducing the diagnosis time [5].

In another study, a model to detect kidney stones was presented with the help of an artificial neural network (ANN) and a discriminant analysis (DA) by using genetic polymorphisms and environmental factors such as water, milk consumption, and outdoor activities. It has been revealed that there are no obvious differences in the success of the models when only genetic factors are considered. But when environmental and genetic factors are considered together, the ANN model, with 89% accuracy, is more successful than DA with 75% [6].

Furthermore, Adl et al. evaluated the risk of morbidity and recurrence of kidney stone in patients having nephrolithiasis by using three data mining approaches such as the mining association rules, clustering, and classification algorithms such as support vector machines (SVM) and k-nearest neighbors (KNN). It was shown that SVM produced better results than KNN [7].

To the best of our knowledge, no study has been done to predict different types of kidney stones. To fill this gap, various data mining techniques have been employed in this study to predict the type of kidney stone, based on the patients' medical information records. A large number of input features used in our study may have improved the predictive power of our paper.

2. Methods

2.1. Patients

The research in this paper is based on information about patients suffering from kidney stone disease, collected from 2012 to 2016. In this study, the medical information of patients, who visited Razi Hospital, was analyzed after obtaining informed consent from the patients. This study was approved by the ethics committee of the Guilan University of Medical Sciences. An analysis of the collected data, led to the following observations:

Table 1
Distribution of continuous features in our dataset.

Name	Min	Max	Average	Standard deviation
Age	4	91	54.686	14.39
HB	1.1	30	11.743	2.10
CR	0.06	11.6	1.381	1.10
BUN	6	160	22.3	18.57
K	0.8	14	4.325	1.01
NA	10.2	152	136.001	12.54

- The age of the patients ranged from 4 to 91 years, and the age of most people suffering from kidney stone was in the range of 20–50 years. It should be noted that the number of male patients was more than females, indicating that the risk of kidney stones among men was higher than in women.
- Flank pain is the most common symptom of kidney stones. Nausea and dizziness are the other vital signs of this disease.
- Urinary stones, according to their location, can be divided into three major categories including kidney, ureter, and bladder stones. Notably, more than half of the stones are found in the kidney.
- As mentioned in the previous section, the target variables in the type of kidney stones can be divided as calcium (70–80%), uric acid (15–20%), staghorn (10%), and cystine (1%). The most common type is calcium stone.
- An evaluation of the frequency of different types of kidney stone revealed that uric acid stones were observed more in men than in women. Calcium and staghorn stone were highly observed in women.

2.2. Dataset characteristics

Information pertaining to 936 patients undergoing treatment for kidney stone at the Razi Hospital, in Rasht, was collected between 2012 and 2016. The collected dataset included 42 features. Each feature and its possible values are described as follows:

1. Age
2.HB: This feature shows blood hemoglobin level.
3.CR: This feature shows creatinine level.
4. BUN: This feature illustrates the level of urea in the blood.
5. K: This feature shows Potassium level.
6. NA: This feature shows Sodium level.
The distribution of continuous features (1–6) is shown in Table 1.
7. Birthplace: This feature shows where patients were born. More details are available in Appendix A.
8. Sex: This binary feature indicates whether the patient is M (Male) or Female (F)
9. Marital status: This binary feature indicates whether the patient is M (Married) or Single(S)
10. CA: This feature shows calcium level in blood which can have High (H), Normal (N), and Low (L) values.
11. BP: This feature shows blood pressure level which can have High (H), Normal (N), and Low (L) values.
12. Uric Acid: This feature shows uric acid level which can have High (H), Normal (N), and Low (L) values.
13. DM: This binary feature indicates whether the patient has diabetes (value 1) or not (value 0).
14. Fever: This binary feature indicates whether the patient has the fever (value 1) or not (value 0).
15. Nausea: This binary feature indicates whether the patient has nausea (value 1) or not (value 0).

Please cite this article in press as: Kazemi Y, Mirroshandel SA. A novel method for predicting kidney stone type using ensemble learning. Artif Intell Med (2017), https://doi.org/10.1016/j.artmed.2017.12.001

Download English Version:

https://daneshyari.com/en/article/6853367

Download Persian Version:

https://daneshyari.com/article/6853367

Daneshyari.com