Kaladhar et al., 1:12
http://dx.doi.org/10.4172/scientificreports.543

Open Access Scientific Reports

Research Article Open Access

Statistical and Data Mining Aspects on Kidney Stones: A Systematic Review and Meta-analysis

DSVGK Kaladhar, Krishna Apparao Rayavarapu* and Varahalarao Vadlapudi

Department of Biochemistry/Bioinformatics, GIS, GITAM University, Visakhapatnam-530045, India

Abstract

To understand the statistical analysis of the kidney stones patients and to apply machine learning techniques to asses correctly classified data. Kidney stone formation is most common due to change of climatic factors and human generations. Ancient culture is losing in India and new industrialization and food habits are producing adverse effects in humans. A survey has been conducted with statistical analysis and data mining on kidney stones. A systematic review and meta-analysis has also been produced in the present work. The present studies predicted good accuracy with C4.5, Classification tree and Random forest (93%) followed by Support Vector Machines (SVM) (91.98%). Logistic and NNge has also shown good accuracy results with zero relative absolute error and 100% correctly classified results. Machine learning approaches may provide better results in the treatment of kidney stones. Diagnosis and treatment of kidney stones is one of the challenging task for researchers in the present decades due to anthropogenic climate change. In the present decades "Prevention is better than Cure", where the health conditions of humans can be increased. The present work provides statistical analysis and classification of the data from people living in Kaviti mandal, Ichapuram area of Andhra Pradesh of which 42 attributes has been selected for machine learning techniques, provided good results in the analysis and prediction of the kidney stone data. The attributes such as Marital status, Intake of Tea and milk, B group, Rh+, Travel by walking, Drinking tap water, Bathing with hot water, Taking rice as meals during Breakfast and dinner, eating non-veg are more prone in occurence of kidney stones. ROC and Calibration curves using Naive Bayes has also been constructed for predicting acuracy of the data.

Keywords: Statistical analysis; Kidney stones; Data mining

Introduction

Obesity and weight gain increase the risk of calcium-containing kidney stone formation due to increased urinary excretion of calcium, oxalate, and uric acid [1]. Information on risk factors for stones including age, race, education, body mass, hypertension, and diuretic. Consumption of milk, coffee, tea, soft drinks, and alcohol and vitamin C supplement use can be obtained by self administered questionnaire [2]. The increased urinary calcium excretion commonly detected in hypertension may be the pathogenesis link [3].

Components of metabolic syndrome, such as obesity, Alzheimer, hypertension, Parkinson, and diabetes, were associated with kidney stone disease and published work is not present on large-scale study to examine the association between metabolic syndrome and history of kidney stones [4]. After adjustment for possible confusing effects of age, sex, education, and baseline protein and fluid intake, the relative risk of a repeated stone, advice to follow a low animal protein, high fluid diet, high fiber has no benefit over advice to increase fluid intake alone [5].

Urinary stone disease is reportedly associated with genetic polymorphisms with the complexity of gene-gene and gene-environmental factor interactions. The development of artificial intelligence, data-mining tools can be used to derive more from patient data in predicting disease [6]. The Support Vector Machines (SVM), an instance of kernel methods for classification are cross-validated by training and testing on randomly selected train and test-set from partitions of the data, collected from patients affected by renal colic and reporting the average performance on the test sets [7].

Since ancient times, kidney stone formation in humans became an unsolved problem and a wide research in this area has to be conducted. Data mining and machine learning techniques can provide the diagnosis and treatment of kidney stones in future. An analysis of the metabolic variables indicated the significance of repeated dietary counselling, and specific dietary therapy by adjusted metabolic evaluation is more effective in preventing the formation of a second urinary stone [8]. The accumulation of metals such as Nickel (Ni) and Copper (Cu) in the organs and tissues have associated changes of the elementary ratios within the organism mostly associated with the kidneys [9].

Data mining offer a more quantitative approach to quality control with more uniform and user friendly for clinicians in reading the reports and reduce errors. Patient-centered care, patient partnership in decision making, and patient-doctor transparency are new keywords in medicine with greater sharing of information with patients, one of the major changes in the doctor-patient relationship [10].

The major advance in nephrolithiasis has been made from past three decades causing stone formation in specific molecular defects. The crystallization stone forming salts is either metabolic or environmental in origin [11]. Obesity and weight gain increase the risk of kidney stone formation may be grater in women than in men [12].

The formation of kidney stones can be closely reported to dietary habits and so importance of elevated Calcium (Ca) and Potassium (K) intake the prevention of kidney stone disease. The interaction between urinary Calcium, Oxalate, Citrate for stone formation in children using one R, J4.8 classifieds has been reported by the Danko et al. [13] using data mining software.

***Corresponding author:** Krishna Apparao Rayavarapu, Department of Biochemistry/Bioinformatics, GITAM University, Visakhapatnam–530045, India, Tel: +919542031726; E-mail: Kr.rayavarapu@gmail.com

Received March 24, 2012; **Published** November 03, 2012

Citation: Kaladhar DSVGK, Krishna Apparao R, Varahalarao V (2012) Statistical and Data Mining Aspects on Kidney Stones: A Systematic Review and Meta-analysis. 1:543 doi:10.4172/scientificreports.543

A meta-analysis combines results of a number of studies that deal with a set of related research hypotheses. A meta-analysis may be conducted on several clinical trials of a medical treatment which refer to statistical methods of combining evidence. In the present experimentation, we had analyzed a set of parameters related to kidney stone formation collected from patients in Kaviti, Srikakulam District, Andhra Pradesh, India.

Materials and Methods

Collection of data

The data has been collected from Kaviti, near to the Ichapuram and the boarder of Andhra and Orissa, India. The analysis has been done by collecting 10,000 datasets that contains 5000 positive instances and 5000 negative instances.

Attributes

There are 42 attributes collected for statistical analysis and classification of the data (class, gender, age, number of family members, marital status, children's, food habits, fruits, vegetables, smoking, drinking, tea, coffee, milk, curd rice, sleeping time, wakeup time, bodyweight, height, blood group, Rh factor, any family background with kidney stones, any family background with diabetes, background with Parkinson or paralysis, travel source, drinking water, florosis, treatment source, any tensions from relatives, any tensions in field/ office, type of water using for bathing, breakfast, dinner, pains, type of oil using, food type, taking cool drinks, habit of gutka, vegetarian or non-vegetarian, habit of panmasala, habit of killi, spit off killi/ panmasala/ gutka).

There are some missing values in the dataset at blood group and food type as some of the people are unable to answer for that questioner. The data is minimized to hundred due to reduction of redundancy while running the algorithm. The data has been analyzed using machine learning tools such as Distribution and Attribute statistics, followed by classification algorithms. Evaluation has been done using Receiver Operating Characteristic (ROC) plot and calibration plots.

Results

Region studies

Kidney stones disease has increased day by day in the Srikakulam District, Andhra Pradesh State, due to modernization and change of culture. In this district one of the area, Kaviti is nearby Ichapuram area, border of Andhra Pradesh and Orissa. This Kaviti mandal also called as Vudyanam Kaviti. It contains 121 villages and 21 panchayats. This Kaviti mandal has two parts - one is rocky place and second one is near sea shore. The people in rocky place cultivated coconut trees. In these two places many researchers conducted and identified that people living near sea shore known as Teera prantham or Coastal region did not have any kidney problems and remaining peoples living in rocky places have many kidney and renal problems.

People living in rocky place takes wine heavily and they also take steroids, antibiotics and pain killers for the purpose of body pains. People from Kaviti are taking irregular drug treatment. These people maintenance is very poor sanitation, irregular food intake, low quality of food and oil, and Illiteracy.

In this place we observed water with 48 elements, but no reason

S.No	Attribute	Sample Type					
1	Gender	57% Males		43% Females			
2	Marital Status	89% Married		11% Unmarried			
3	Food Habits	51% Fruits		49% Vegetables			
4	Smoking and Drinking	16% Smoking 84% No Smoking		32% Drinking 68% No Drinking			
5	Tea and Coffee	86% Tea taken 14% None		27% Coffee taken 73% None			
6	Milk and Curd rice	80% Milk Taken 20% None		47% Curd rice Taken 53% None			
7	Blood Grouping	8% A Group 14% AB Group		55% B group 23% O group		95% Rh +ve 5% Rh +ve	
8	Family Background	3% Previous 97% None		1% Diabetes 99% None		1% Parkinson 99% None	
9	Travel Source	6% Bus 2% Car		9% Cycle 4% Motorcycle		79% Walk	
10	Drinking Water	76% Tap Water 18% Bore water 6% Pure water		5% fluorosis 95% None			
11	Treatment Source	100% Allopathy					
12	Tensions	5% Fields 95% None		4% Relatives 96% None			
13	Daily Timings	Bathing	89% Hot Water 11% Cold water	Breakfast	33% Tiffin 48% Meals	Dinner	100% Meals 0% None
14	Pain Regions	27% Spinal Cord 13% Kidney		7% Stomach 1% Heart		0% Leg pains 55% None	
15	Oil Type	39% Low Quality 50% Pure		11% Refined Oil			
16	Fast food and Cool drinks	40% Fast Food 60% None		46% Cool Drinks 54% None			
17	Habit of taking Pan	21% Gutka 23% Panmasala		47% Killi 9% Inside			
18	Non vegetarian	2% Veg		98% Non- Veg			

Table 1: Data analysis of some attributes collected from Kidney stone patients.

for kidney stones and kidney infections. In this area, serum-creatinine levels are 1.25 mg/dl and urea levels observed as 45 mg/dl, observed in kidney patients.

In this area coconut trees are yellow and orange coloured and have fungal infection. These people take gutka, killi, panmasala and low quality cool drink. Previously people used to drink well water but presently the wells are closed and taking tap and bore water known as "Rakshitha manchineeti padhakam" issued by the government of Andhra Pradesh, India.

In this area we have consulted many people suffering with kidney stones and they expressed their views on this disease based on questionnaire. Mainly problems related to spinal cord, kidneys and teeth have been reported. Some people are having yellowish teeth, because people from Kaviti use some gutka, panmasala and water containing high florin content (Tables 1 and 2).

Class category and gender

The collected dataset contains 50% positive and 50% negative instances in the occurrence of kidney stones. The dataset presented more occurrence in males (57%) compared with females (43%).

Age, Members in family and number of children

The mean age for the occurance of arteriosclerosis is 42.44 ± 12.79. 25% of the occurances are observed below 32 years of age groups and seventy five percent of the occurences are observed below 53.50 years. The age groups collected are from 12 to 70 years. Hence the kidney stone disease can occur at smaller age groups also and hence we reports thet this disease is not related to aging diseases in the present decades. The disease occurrence is only due to deficiency or mutation in the genes due to metabolic disorders.

The data related to family members ranged from 1 to 9 members/family. Most of the kidney stone patients are having average number of 4.48 ± 1.33 with 25% occurences below 4 members and 75% occurences below 5 members. The number of children with the family members ranged from 0 to 5, with the mean of 2.13 ± 0.93 with 25% patients below 2 children and 75% patients below 3 children.

Marital status

Most of the dataset are lying towards the married people (89%). This provides the information that both married and unmarried humans can be presented with kidney stones. Based on data analysis, we predicted that mostly married people will be positive towards the occurrence of kidney stones.

Food habits

The present results provides the information of food habits of the kidney stone patients. Data related to intake of foods such as fruits, vegitables, grains, animal products etc has been collected on their food habits. Most of the people in those regions are taking rice and potato followed by brinjal, buttermilk and leafy vegetables like spinach with tomato. Most of the people are taking all types of fruits (51%) but mostly taking Banana, apple and orange. Most of the people are taking all types of vegetables (48%) but taking mostly spinach and tomato. Hence these vegetables may lead to kidney stones, if taken in excess.

Smoking and drinking habits

The present data has also provides the information regarding smoking and drinking habits of the people. The information provides less number of people with these habits. Nearly 16% of the data has smoking habit and 32% of the people has drinking habit. Hence smoking may not effect kidneys but intake of alcohol may produce adverse effects on kedneys.

Tea and coffee

The present data also provides the best analysis on intake of tea and coffee. Most of the people take tea (86%) and nearly 27% of the datasets provided of taking coffee. Hence the intake of tea may show adverse effect in occurence of kidney stones. The intake of coffee show low level of occurence of kidney stones.

Milk and curd rice

Most of the people are intrested in taking milk (80%) but decreased in taking curd rice (47%). provided the information regarding negative interest in taking curd rice. The intake of curd rice (fresh) may provide the decrease of kidney stone formation due to the increase of activity of probiotic bacteria in the human intestine.

Wake up habits

Fifty percent of the people are used to wakeup at 5.00 am in the morning. The present dataset has provided the information that 75% of the people wake up at 5.30 am with an average wakeup time of 5.12 ± 0.22 am leads for occurence of kidney stones.

Weight and height

Based on the questionair regarding the weight and height of the kidneystone dataset, average weigh and height are more in occurence of kidney stones. The dataset contains the range of body weight from 30 kg to 85 kg. 75% of the attribute is with a weight below 61 percent. An average weight of the dataset lies at 56.14 ± 9.35. The height of the people are from 140 cm to 170 cm and the dataset lies an average height at 161.09 ± 7.06.

Blood grouping

Most of the people are predicted with B blood group, followed by O group. Most of the people with Rh factor positive shows the occurence of kidney stones. Bood groups with A negative may show less intensity in occurance of Kidney stone.

Family back ground

Further investigation is also been conducted with the family history. Nearly 3% of the data has been predicted with family history with kidney stone, 1% with Diabetes 2 Mellitis and 1% with parkinson or paralysis diseases.

Travel source

Most of the people suffering with kidney stones from this region is choosen walking as a travel method. Next to walking, Cycling is using

S.No	Attribute	Mean	Median	Max	Min
1	Age	42.44±12.79	43.00	70.00	12.00
2	Family Members	4.48±1.33	4.00	9.00	1.00
3	Number of Childs	2.13±0.93	2.00	5.00	0.00
4	Wake-Up Time	5.12±0.22	5.00	6.00	4.30
5	Weight	56.14±9.35	55.50	85.00	20.00
6	Height	161.09±7.06	162.00	170.00	140.00

Table 2: Measuring central tendency on some attributes leading to kidney stones.

as a travel source. Two percent of the people suffering with this disease prefered to take bus as traveling source.

Drinking water

The present investigation also provided the quality of drinking water in this area. Previously 100% of the people used to drink pure water from wells. Due to increase of industrialisation and decrease in farm land, water quality is changing from healthy to diseased forms. Wells are closed and water is being supplied through taps. Hence 76% of the data is shown with drinking tap water. 5% of the data is predicted with florosis.

Treatment source

All the people has prefered allopatic treatments rather than ayurvedic or Homeopathy.

Tensions

There are nearly 4-5% tensions in fields and relatives. This can lead to decrease in health conditions. About 95% of the dataset predicted not having tensions from relatives and field work.

Daily timings

Daily work timings has also been investigated in the present study. Most of the people prefered to bath with hot water (89%) rather than cold water (11%). People used to take curd rice in the ancient times in this region. Due to change of culture, the intake of tiffins has been increased (33%) and also meals (48%).

Pain regions

Fifty two percent of the data is provided with absence of pains due to adding of negative datasets. Most of the kidney stone patients has complications mostly with spinalcord (27%), followed by kidney pain (13%), stomach pain (7%) and heart pain (1%). Hence there may be interrelationship of kidney stone disease with spine, stomach and heart related diseases.

Oil type

The investigation has also provided the usage of oil types. Thirty percent of the people are using low quality oils based on the investigation of the survey.

Fast food and cool drinks

The intake of fast foods and cool drinks has also been increased in this regions. Nearly fourty percent of the people are taking fast foods. Fourty percent of the people also taking cool drinks in this regions.

Habit of taking pan

Taking hot, sour and sweet pans has also been investigated in the occurence of kidney stones. The intake of this items is getting as a habbit to the people from the past decades. 21% are using Gutka, 23% using pan masala and 40% of the people take killi are effecting with kidney stones. Killi and pan masala contains calcium carbonate on piper betle. 24% are not spitting these items and taking inside as food items. Hence the utilty may cause adverse effects with other food items but have good effect in certain situations.

Non-vegetarian

Non-vegetarian food habit is also have relationship with kidney stone formation, most of the people take calcium carbonate present in bones. 98% of the people like to eat non-veg, leads to produce kidney stones in humans.

Machine Learning

Machine learning approaches has also been analysed to know the accuracy of the datasets (Table 3). Orange version 2.0b predicted

S.No	Method	CA	Sens	Spec	AUC
1	NaiveBayes	0.79	0.96	0.75	0.97
2	Majority(TestLearners)	0.76	0	1	0.42
3	kNN	0.7377	0.297	0.8816	0.7138
4	Classification tree	0.9352	0.8333	0.9737	0.8254
5	C4.5	0.9352	0.8333	0.9737	0.8243
6	SVM	0.9198	0.75	0.9737	0.9079
7	Random Forest	0.9352	0.8333	0.9737	0.9435

Table 3: Classification of Kidney stone dataset using Orange v 2.0b (at cross-validation of 27 folds).

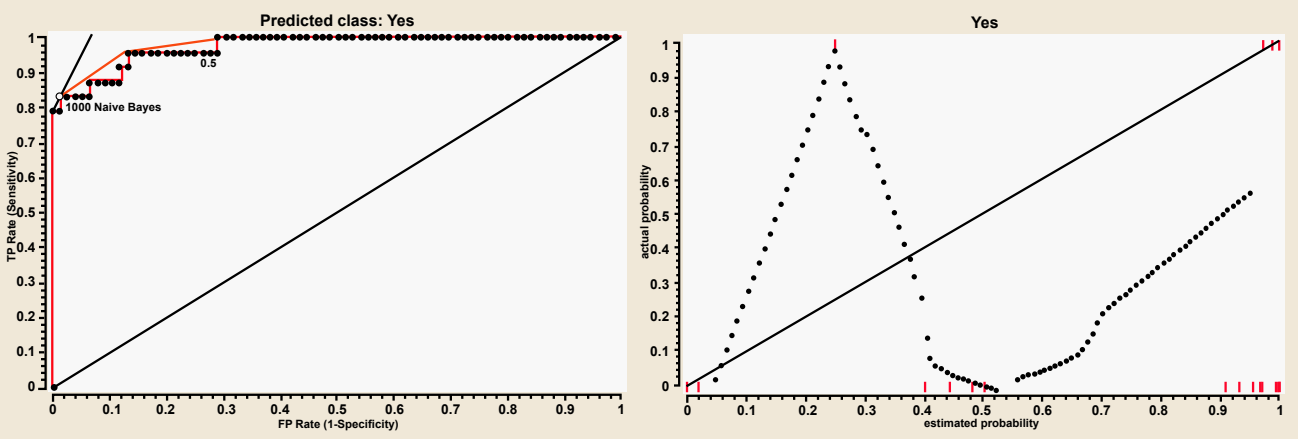

Figure 1: ROC and Calibration plot for Naïve Bayes.

S.No	Method	CA	Kappa Statistic	Relative absolute error (%)	Precision
1	NaiveBayes	0.99	0.98	2.5696	0.99
2	Logistic	1.00	0	0	1
3	NNge	1.00	0	0	1
4	Bagging	0.99	0.98	3.3341	0.99
5	J48	0.97	0.94	5.9152	0.972
6	LMT	1.0	1	23.8213	1
7	Random Forest	0.98	0.96	5.69	0.981

Table 4: Classification of kidney stone dataset using weka v3.6.6 (at cross-validation of 27 folds).

good accuracy with C4.5, Classification tree and random forest (93%) followed by SVM (91.98 %).

ROC and Calibration plot for Naive Bayes

Figure 1 provided the ROC and Calibration plot for Naive Bayes of the kidney stone disease dataset. The plots predicted 79% truly classified data of a collected dataset.

Table 4 shows the classification of datasets using various clssification algorithms at 27 fold cross-validation with weka v3.6.6. Logistic and NNge has shown good accuracy results with zero relative absolute error and 100% correctly classified results.

Discussion

Data mining is the process of analyzing and predicting data from many different dimensions and summarize the relationships identified using analytical tools into useful information that can be used to increase accuracy of the data [14].

Data mining techniques can be well applied using various algorithms like CART, Simple Logistic, BayesNet, Random forest, and j48, LMT, Naïve Bayesian, Apriori and simple KMeans. Random forest method using training dataset outperforms the remaining classification methods.

Acknowledgements

The author would like to thank GITAM Institute of Science, GITAM University for providing lab facility and access to e-journals to carry out this research.

References

1. Taylor EN, Stampfer MJ, Curhan GC (2005) Obesity, Weight Gain, and the Risk of Kidney Stones. JAMA 293: 455-462.
2. Michael SJ, Ralph JC, McClellan W, Harland A, Michael T (1996) Relation between Geographic Variability in Kidney Stones Prevalence and Risk Factors for Stones. Am J Epidemiol 143: 487-495.
3. Cappuccio FP, Strazzullo P, Mancini M (1990) Kidney stones and hypertension: population based study of an independent clinical association. BMJ 300: 1234-1236.
4. West B, Luke A, Durazo-Arvizu RA, Cao G, Shoham D, et al. (2008) Metabolic Syndrome and Self-Reported History of Kidney Stones. The National Health and Nutrition Examination Survey (NHANES III) 1988- 1994. Am J Kidney Dis 51: 741-747.
5. Robert AH, Bruce E, Bette C, Charles PQ, Debra D, et al. (1996) Randomized Controlled Trial of a Low Animal Protein, High Fiber Diet in the Prevention of Recurrent Calcium Oxalate Kidney Stones. Am J Epidemiol 144: 25-33.
6. Chiang D, Chiang HC, Chen WC, Tsai FJ (2003) Prediction of stone disease by discriminant analysis and artificial neural networks in genetic polymorphisms: a new method. BJU Int 91: 661-666.
7. Moro FD, Abate A, Lanckriet GR, Arandjelovic G, Gasparella P, et al. (2006) A novel approach for accurate prediction of spontaneous passage of ureteral stones: Support vector machines. Kidney Int 69: 157–160.
8. Kocvara R, Plasgura P, Petrík A, Louzenský G, Bartonícková K, et al. (1999) A prospective study of nonmedical prophylaxis after a first kidney stone. BJU Int 84: 393-398.
9. Moiseenkoa TI, Kudryavtseva LP (2001) Trace metal accumulation and fish pathologies in areas affected by mining and metallurgical enterprises in the Kola Region, Russia. Environ Pollut 114: 285-297.
10. Ferris M (2009) The Radiology Report of the Future. Radiology 251: 313-316.
11. Pak CY (2003) Kidney stones. Lancet 351: 1797-1801.
12. Milosević D, Batinić D, Konjevoda P, Blau N, Stambuk N, et al. (2003) Analysis of Calcium, Oxalate, and Citrate Interaction in Idiopathic Calcium Urolithiasis in Children. J Chem Inf Comput Sci 43: 1844-1847.
13. Kaladhar DSVGK, Chandana B, Bharath KP (2011) Predicting Cancer Survivability Using Classification Algorithms. International Journal of Research and Reviews in Computer Science 2: 340-343.
14. Kaladhar DSVGK, UmaDevi T, Lakshmi PV, Harikrishna RR, SriTeja RK, et al. (2012) Analysis of *E.coli* Promoter Regions Using Classification, Association and Clustering Algorithms. Advances in Intelligent and Soft Computing 132: 169-177.