Abstract

This paper focused on the development of a predictive model for the classification of the risk of kidney stones in Nigerian using data mining techniques based on historical information elicited about the risk of kidney stones among Nigerians. Following the identification of the risk factors of kidney stone from experienced endocrinologists, structured questionnaires were used to collect information about the risk factors and the associated risk of kidney stones from selected respondents.

The predictive model for the risk of kidney diseases was formulated using three (3) supervised machine learning algorithms (Decision Tree, Multi-layer perception and Genetic Algorithm) following the identification of relevant features. The predictive model was simulated using the Waikato Environment for Knowledge Analysis (WEKA) environment; and the model was validated using historical dataset of kidney stone risk via performance metrics: accuracy, true positive rate, precision and false positive rate.
The paper concluded that the multi-layer perceptron had the best performance overall using the 33 initially identified variables by the endocrinologists with an accuracy of 100%. The performance of the genetic programming and multi-layer perceptron algorithms used to formulate the predictive model for the risk of kidney stones using the 6 variables outperformed the model formulated using the 6 variables identified by the C4.5 decision trees. The variables identified by the C4.5 decision trees algorithm were: obese from childhood, eating late at night, BMI class, family history of hypertension, taking coffee and sweating daily. In conclusion, the multi-layer perceptron algorithm is best suitable for the development of a predictive model for the risk of kidney stones.

References

32. Kaladhar D, Krishna Apparao Rayavarapu K and Vadlapudi V 2012. Statistical and Data Mining Aspects on Kidney Stones: A Systematic Review and Meta-analysis Department of Biochemistry/Bioinformatics, GIS, GITAM University, Visakhapatnam-530045, India


Index Terms

Computer Science
Data Mining

Keywords

Kidney Stone Risk Factors, C4.5, Prediction, Classification, Decision Trees, Genetic Algorithms, Multilayer Perception