Chapter 59
Diagnosis Rule Extraction from Patient Data for Chronic Kidney Disease Using Machine Learning

Alexander Arman Serpen
Southview High School, Sylvania, USA

ABSTRACT

This research study employed a machine learning algorithm on actual patient data to extract decision making rules that can be used to diagnose chronic kidney disease. The patient data set entails a number of health-related attributes or indicators and contains 250 patients positive for chronic kidney disease. The C4.5 decision tree algorithm was applied to the patient data to formulate a set of diagnosis rules for chronic kidney disease. The C4.5 algorithm utilizing 3-fold cross validation achieved 98.25% prediction accuracy and thus correctly classified 393 instances and incorrectly classified 7 instances for a total patient count of 400. The extracted rule set highlighted the need to monitor serum creatinine levels in patients as the primary indicator for the presence of disease. Secondary indicators were pedal edema, hemoglobin, diabetes mellitus and specific gravity. The set of rules provides a preliminary screening tool towards conclusive diagnosis of the chronic kidney disease by nephrologists following timely referral by the primary care providers or decision-making algorithms.

INTRODUCTION

Chronic kidney disease (CKD), also known as chronic renal disease, is a medical condition in which kidney function is lost over a long period ranging from months to years (NKF, 2002). There is no specific set of detectable symptoms for the disease but possible symptoms may include feeling generally unwell or having a reduced appetite. Therefore, chronic kidney disease is most often detected in individuals who are at high risk through advanced screening processes that determine health attributes associated with chronic kidney disease. Such attributes include high blood pressure, diabetes, or having a blood type strongly associated with the presence of chronic kidney disease. The presence of this disease can

also be identified from a blood test for creatinine, a breakdown product of muscle metabolism. Chronic kidney disease progresses through multiple stages, with each having its own unique health complications. What differentiates chronic kidney disease from its counterpart, acute kidney disease, is that the reduction in kidney function develops and strengthens over the course of at least three months in CKD. Acute kidney disease occurs rapidly over the course of a few hours or days and is easily reversible (Mayo Clinic, n.d.). The attributes or indicators associated with CKD are the subject of this study as they were used to assess the presence of disease (NKF, 2002).

Often chronic kidney disease is discovered in patients at a later stage in which kidney transplants are necessary and mortality from cardiovascular disease or other related conditions is highly likely (Naghavi et al., 2015). It is important to detect the disease in individuals at an early stage so that treatments that delay the progression of chronic kidney disease can be applied. Additionally, if an underlying cause of chronic kidney disease is discovered, such as obstructive nephropathy, then that specific cause can be treated to slow the progression of CKD as well. In doing so, the overall mortality risk from this disease can be limited (Naghavi et al, 2015).

Artificial intelligence (AI), specifically machine learning (ML), can be used for improving the process of screening for, assessing, and evaluating the risk in individuals for chronic kidney disease. Artificial intelligence in general and machine learning algorithms in specific have been previously used by other researchers to develop risk assessment and evaluation models for kidney diseases (Baby & Vital, 2015). Many health factors are involved in predicting the onset or presence of chronic kidney disease, as it is a complex reasoning process to conclude with a definitive diagnosis. Among many different health conditions and factors that can be tested for in humans, levels of urea and creatinine, two byproducts of glomerular filtration directly related to kidney function, are highly significant (Zhang et al., 2008). These two factors, along with several other important ones, have been used for assisting in detection of chronic kidney disease in national surveys. For example, a study by Coresh et al. (2003) analyzed 15,625 adults over age 20 for kidney function, kidney damage, and stages of CKD using five factors including urea and creatinine levels. Although they did not predict the risk of CKD, they were successfully able to quantify the prevalence of the disease with analysis of the same factors. In 2015, another study (Vijayarani et al., 2015) used data mining algorithms to test six attributes closely associated with CKD in 584 different instances. Attributes tested were age, gender, urea, creatinine, and glomerular filtration rate. Patients were then classified into predefined classes or groups to successfully determine their risk factors. Several other studies report successful results when a host of machine learning classification algorithms are applied to patient data to identify those patients with high risk of chronic kidney disease (Ramya et al., 2016; Dubey, 2015; Kumar, 2016; Sinha et al., 2015; Rubini et al., 2015). These studies create classifier models and do not offer the insight gained from analysis of the data set on which the classifier models were induced. In other words, once these classifier models are induced from the dataset they are used as black-box decision making tools. The data however can also be mined to extract knowledge that can be readily interpreted by humans to identify the biomarkers for CKD.

Predicting CKD in later stages of life is a difficult process involving many factors. Physicians and other health professionals must evaluate manually individual attributes when considering a patient's risk in developing CKD through a painstakingly detailed process and often may leave out important ones or skip entire parts of the overall process due to its high cost for time and effort. The aim of conducting this study was to analyze and extract knowledge from a comprehensive set of human test data that involves up to twenty-five health factors, tests, and attributes to benefit non-specialist physicians in predicting the developmental risk for chronic kidney disease and assisting with its diagnosis. Machine learning

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/diagnosis-rule-extraction-from-patient-data-for-chronic-kidney-disease-using-machine-learning/252076

Related Content

A Machine Learning Approach for Predicting Bank Customer Behavior in the Banking Industry
Siu Cheung Ho, Kin Chun Wong, Yuen Kwan Yauand Chi Kwan Yip (2019). *Machine Learning and Cognitive Science Applications in Cyber Security (pp. 57-83).*
www.irma-international.org/chapter/a-machine-learning-approach-for-predicting-bank-customer-behavior-in-the-banking-industry/227576

Spatial Modeling and Geovisualization of Rental Prices for Real Estate portals
Harald Schernthanner, Hartmut Asche, Julia Gonschorekand Lasse Scheele (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 962-977).*
www.irma-international.org/chapter/spatial-modeling-and-geovisualization-of-rental-prices-for-real-estate-portals/252065

Inferring Intent and Action from Gaze in Naturalistic Behavior: A Review
Kristian Lukander, Miika Toivanenand Kai Puolamäki (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 1464-1482).*
www.irma-international.org/chapter/inferring-intent-and-action-from-gaze-in-naturalistic-behavior/252092

Implicit Social Cognition and Language Attitudes Research
Andrew J. Pantos (2015). *Exploring Implicit Cognition: Learning, Memory, and Social Cognitive Processes (pp. 104-117).*
www.irma-international.org/chapter/implicit-social-cognition-and-language-attitudes-research/120855

Comparison of Several Acoustic Modeling Techniques for Speech Emotion Recognition
Imen Trabelsiand Med Salim Bouhlel (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 283-293).*
www.irma-international.org/chapter/comparison-of-several-acoustic-modeling-techniques-for-speech-emotion-recognition/252030