Identifying Important Attributes for Early Detection of Chronic Kidney Disease

Anandanadarajah Nishanth and Tharmarajah Thiruvaran, *Member, IEEE*

***Abstract*—Individuals with chronic kidney disease (CKD) are often not aware that the medical tests they take for other purposes may contain useful information about CKD, and that this information is sometimes not used effectively to tackle the identification of the disease. Therefore, attributes of different medical tests are investigated to identify which attributes may contain useful information about CKD. A database with several attributes of healthy subjects and subjects with CKD are analyzed using different techniques. Common spatial pattern (CSP) filter and linear discriminant analysis are first used to identify the dominant attributes that could contribute in detecting CKD. Here, the CSP filter is applied to optimize a separation between CKD and non-CKD subjects. Then, classification methods are also used to identify the dominant attributes. These analyses suggest that hemoglobin, albumin, specific gravity, hypertension, and diabetes mellitus, together with serum creatinine, are the most important attributes in the early detection of CKD. Further, it suggests that in the absence of information on hypertension and diabetes mellitus, random blood glucose and blood pressure attributes may be used.**

***Index Terms*—Chronic kidney disease (CKD), common spatial pattern (CSP) filter, estimated glomerular filtration rate (eGFR) and serum creatinine, linear discriminant analysis (LDA).**

I. Introduction

CHRONIC kidney disease [CKD (also called chronic renal disease)] is a condition in which kidneys gradually lose their function. If the kidney does not function properly, this could cause waste and excess fluid accumulation in the body, affecting its functionality, and potentially leading to complications. The disease can progress to end-stage renal disease (complete kidney failure). This occurs when kidney function is worsened to a point where dialysis or kidney transplantation is required for survival. People with CKD also have an increased risk of developing cardiovascular diseases (CVD) [1]–[2]. Further, the CVD in the CKD population is different from those who are not affected by CKD (nonCKD) [1]; i.e., CVD is the leading cause of death in individuals who are on dialysis [1], as they are more vulnerable, but renal disease can also lead to other causes of death.

A person who is affected by early CKD may not feel unwell or notice any problems. So identifying CKD at the early stage is not easy without proper tests, specifically urine and blood tests. However, several attributes of medical tests taken for other purposes contain useful information for CKD. To effectively use these attributes, their relation and importance to CKD should be studied in detail. There have been many studies to find the risk factors of CKD. In [4], Johnson summarized the risk factors as obesity, hypertension, diabetes mellitus, cigarette smoking, established CVD, age greater than 60 years, aboriginal and Torres Strait Islander peoples, Maori and Pacific peoples, family history of stage 5 CKD or hereditary kidney disease in a first- or second-degree relative, and severe socioeconomic disadvantage.

In-order to reduce the chances of CKD leading to dialysis or kidney transplantation, early detection is important. Kidney imaging may be used to confirm the disease, although due to sheer numbers, it is not possible to test everyone, and only those who have high probability to have CKD may be recommended. In [5], it is suggested that if CKD is detected earlier, then even combined specialized nephrology nurses and primary care clinicians can provide special attention on screening, monitoring, and advising the patient to prevent or reduce the development of CKD.

In-order to detect CKD, glomerular filtration rate (GFR) was earlier calculated from Cockcroft–Gault formula (developed in 1973) to check the condition of the renal function [6]. This formula especially considers the widely used index, serum creatinine concentration, to measure the renal function [7]. However, serum creatinine concentration does not rely only on GFR but other factors as well [7]. Then, the equation is modified as estimated GFR (eGFR) with medication of diet in renal disease. This uses serum creatinine, age, ethnicity, gender, blood urea nitrogen, and albumin [7]. Furthermore, a CKD epidemiology collaboration equation was developed in 2009 to estimate GFR from serum creatinine, age, gender, and race. This equation was updated in 2012 based on a cross-sectional analysis of 13 studies [8]. The inclusion of age in [8] was motivated by the continuous population studies in [9]–[14]. The inclusion of gender in [8] was motivated by the meta-analysis in [14] and [15]. However, the analysis of the importance of age and gender based on eGFR using serum creatinine carries an implied assumption that serum

Manuscript received January 23, 2017; revised November 4, 2017; accepted December 9, 2017. Date of publication December 25, 2017; date of current version July 24, 2018. *(Corresponding author: Anandanadarajah Nishanth.)*

A. Nishanth was with the Department of Electrical and Electronic Engineering, Faculty of Engineering, University of Jaffna, Kilinochchi 44000, Sri Lanka. He is now with the Department of Computer Science, University of Louisiana at Lafayette, Lafayette, LA 70504 USA (e-mail: eng_nishanth@eng.jfn.ac.lk).

T. Thiruvaran was with the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW 2052, Australia. He is now with the Department of Electrical and Electronic Engineering, Faculty of Engineering, University of Jaffna, Kilinochchi 44000, Sri Lanka (e-mail: thiruvaran@eng.jfn.ac.lk).

Digital Object Identifier 10.1109/RBME.2017.2787480