An Artificial Neural Network Framework to Predict Patients with High Likelihood of Chronic Kidney Disease

Faisal A. Alkhaldi, *Student Member, IEEE*

Abstract—**Chronic kidney disease (CKD) is an important health and healthcare system problem. The ability to predict which patients will develop CKD is a difficult task due to the complex nonlinear relationships among related factors. Using artificial neural networks (ANN), applied to a population 17 through 90 years of age, we achieved 97% accuracy in classification, based on standard laboratory test and patient data. The technique was also helpful in determining which features of the data are most predictive; 75% of the features were sufficient to reach this high level of accuracy.**

I. Introduction and Literature Review

Chronic kidney disease (CKD) is a dangerous condition in which the kidneys are partially damaged and cannot filter blood normally. The prevalence of CKD is increasing in the US and around the world; 1 in every 10 American adults has some level of CKD [1]. In healthcare practices, patients who are diagnosed with CKD in the early stages are usually managed medically, allowing them to avoid unwanted and costly hemodialysis (HD) treatment. In this fashion, this study aims to help care providers and to empower individuals who may potentially develop CKD, so they may create attentive healthcare plans and make better health-related decisions.

Machine learning methods and data mining techniques have been utilized in studies related to kidney disease. We have identified two main characteristics of the majority of research in this area: 1) studies with specific focus on patients already diagnosed with either CKD or end-stage renal disease (ESRD), and 2) studies that discuss specific medical issues at a microscopic level. This study, however, addresses a macroscopic problem that is rarely considered. We developed a decision-support system in order to predict which individuals (clients) in a population have a high likelihood of developing CKD, to improve CKD management and outcomes.

Based on a comparative study we conducted, we chose the ANN technique to build a predictive model. Similar use of this technique has been reported in various application domains including healthcare, where it has been applied to highly nonlinear multivariable problems [2], [3].

II. Data, Data Preparation, Modeling and Results

The data, collected by Apollo Hospitals, India, are based on obtained clinical laboratory tests of participants. The data includes 400 participants, with 24 features each. The data are not balanced; however, as the majority of the samples are from patients with CKD, no data balancing technique is applied. Fig. 1 shows the process we followed to prepare the data for modeling. The proposed ANN framework recorded very encouraging results (see Fig. 2 and Fig. 3). For independent hold-out testing, the ANN achieves over 97% classification accuracy. Furthermore, through a trial and error process, we were able to identify 18 out of 24 significant features (75%) for the prediction. Future work will focus more on these 18 features.
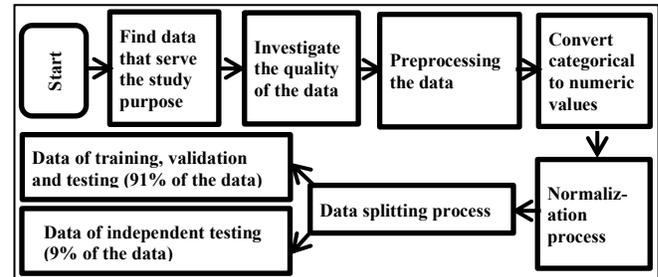

Figure 1. Block diagram of the chronological steps to process the data. Data preparation begins with finding the data that serves the study purpose, and ends with splitting the data into the training-testing-validation dataset and hold-out dataset.
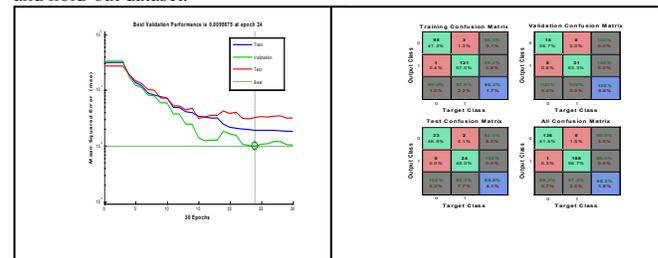

Figure 2 (left). Training, testing and validation performance are shown in terms of MSR; at epoch 24, the model performs the best in reducing the error while keeping the best level of generality of the model.

Figure 3 (right). Confusion matrices for training, validation, and testing performance. With 70% of the data assigned for training, and 15% of the data assigned for validation process, 100% accuracy is achieved; with another 15% assigned for testing stage, 95.9% accuracy of classification is achieved. The overall performance achieved 98.2% accurate classification. The data were randomly split.

References

[1] National Institute of Diabetes and Digestive and Kidney Diseases. ONLINE, [http://www.niddk.nih.gov]

[2] Nura Esfandiari, Mohammad Reza Babavalian, Amir-Masoud Eftekhari Moghadam, Vahid Kashani Tabar. "Knowledge discovery in medicine: Current issue and future trend," 2014, Elsevier Journal

[3] Zoran Bosni, Petar Vra, Milos D. Radovi, Goran Deved, Nenad D. Filipovi, and Igor Kononenko. "Mining data from hemodynamic simulations for generating prediction and explanation models," IEEE Transactions on Information Technology in Biomedicine, Vol. 16, No. 2, March 2012