A Soft Computing Approach to Kidney Diseases Evaluation

José Neves 1 · M. Rosário Martins 2 · João Vilhena 2 · João Neves 3 · Sabino Gomes 1 · António Abela 1 · José Machado 1 · Henrique Vicente 4

Received: 30 April 2015 / Accepted: 4 August 2015 / Published online: 27 August 2015
© Springer Science+Business Media New York 2015

Abstract Kidney renal failure means that one’s kidney have unexpectedly stopped functioning, i.e., once chronic disease is exposed, the presence or degree of kidney dysfunction and its progression must be assessed, and the underlying syndrome has to be diagnosed. Although the patient’s history and physical examination may denote good practice, some key information has to be obtained from valuation of the glomerular filtration rate, and the analysis of serum biomarkers. Indeed, chronic kidney sickness depicts anomalous kidney function and/or its makeup, i.e., there is evidence that treatment may avoid or delay its progression, either by reducing and prevent the development of some associated complications, namely hypertension, obesity, diabetes mellitus, and cardiovascular complications. Acute kidney injury appears abruptly, with a rapid deterioration of the renal function, but is often reversible if it is recognized early and treated promptly. In both situations, i.e., acute kidney injury and chronic kidney disease, an early intervention can significantly improve the prognosis.

The assessment of these pathologies is therefore mandatory, although it is hard to do it with traditional methodologies and existing tools for problem solving. Hence, in this work, we will focus on the development of a hybrid decision support system, in terms of its knowledge representation and reasoning procedures based on Logic Programming, that will allow one to consider incomplete, unknown, and even contradictory information, complemented with an approach to computing centered on Artificial Neural Networks, in order to weigh the Degree-of-Confidence that one has on such a happening.

The present study involved 558 patients with an age average of 51.7 years and the chronic kidney disease was observed in 175 cases. The dataset comprise twenty four variables, grouped into five main categories. The proposed model showed a good performance in the diagnosis of chronic kidney disease, since the sensitivity and the specificity exhibited values range between 93.1 and 94.9 and 91.9–94.2 %, respectively.

1 Algoritmi, Universidade do Minho, 4710-057 Braga, Portugal
2 Departamento de Química, ICAAM, Escola de Ciências e Tecnologia, Universidade de Évora, 7000-671 Évora, Portugal
3 Drs. Nicolas & Asp, Dubai, United Arab Emirates
4 Departamento de Química, Centro de Química de Évora, Escola de Ciências e Tecnologia, Universidade de Évora, 7000-671 Évora, Portugal
Keywords Kidney disease · Soft computing · Artificial neural networks · Incomplete information · Logic programming · Knowledge representation and reasoning

Introduction

Chronic Kidney Diseases (CKDs) or Chronic Renal Failure (CKF) describes abnormal kidney function and/or structure. CKD is a progressive disease with high mortality rate, which threatens to become a major public health problem, i.e., when symptoms are severe they can be treated only by dialysis and transplantation (end-stage of CKD) [1, 2]. It is estimated that in Portugal more than 800,000 patients experience CKD. Every year are recorded over 2200 new cases of CKF (there are at present 15,000 patients in this condition) [3].

CKD was usually caused by a long-term malady, often associated with Hypertension, Obesity, Diabetes Mellitus (DM) and/or CardioVascular Diseases (CVD), which slowly damages the organ and reduces their function over time [4]. Some studies reporting also prevalence in the female patients comparatively with males with the same age, still it may progress faster in males [1, 4]. CKD is common, frequently unrecognised and often exists together with Acute Kidney Injury (AKI), that appears abruptly, with a rapid deterioration of the renal function, resulting in inability to maintain fluid, electrolyte and acid–base balance, including glomerulonephritis, infective, obstructive and reflux nephropathies or untreated urinary outflow tract obstruction [4–6].

Epidemiologic evidence suggests that CKD is a risk factor for AKI, due to the prevalence of CKD in patients who have episodes of AKI [7]. However, the high burden of comorbidities such as age, diabetes, peripheral vascular, cardiovascular, and liver disease accompanying CKD, and the difficulties of defining AKI in the setting of CKD, make these observations difficult to interpret [7, 8]. Other factors should be also considered for the risk of CKD, as a family history of CKD, hereditary kidney disorder (e.g., polycystic kidney), neoplasms or myelomas or multisystem syndromes with potential kidney involvement (e.g., Systemic Lupus Erythematosus (SLE)) [4, 9].

Kidney failure is traditionally regarded as the most serious outcome of chronic kidney disease and symptoms are usually caused by complications of reduced kidney function. When symptoms are severe they can be treated only by dialysis or transplantation, corresponding to the end-stage of renal disease [6]. CKD prevalence was determined based on persistent presence of kidney damage (e.g., albuminuria) or the decreased kidney function (e.g., low Glomerular Filtration Rate (GFR), for at least during a period of 3 months) [10]. Earlier stages of CKD can be detected through routine laboratory measurements, as albuminuria, an early and sensitive marker of kidney damage [10]. The GFR values are considered as an index of the kidney function, accepted as the best measure of overall kidney behavior, either in health or disease [10]. Additionally, serum creatinine or serum urea levels should be monitored as kidney function [1, 4]. Nevertheless, it is important to note that CKD can progress silently for a long time, i.e., prevention and diagnosis at an early stage is imperative to stop the progression of CKD to a more advanced stage, where dialysis or transplant is the only substitute available to treatment. Cystatin C is an alternative filtration marker for estimating GFR (or estimated GFR), i.e., the use of the combined balance between Creatinine and Cystatin C to determined eGFR may be useful as a confirmatory test for chronic kidney disease [11]. Then again, accurate assessment of GFR is essential for interpreting the symptoms, signs, and laboratory abnormalities that may indicate the disease, or to set drug dosing and evaluating the prognosis. A reduction in GFR to less than 60 ml per minute per 1.73 m² (i.e., the body surface area) for 3 months or longer is a diagnostic criterion for chronic kidney dysfunction, and is associated with an increased risk of adverse outcomes, including death [11].

Risk factors of CKD included obesity, lipid diseases and anemia. Obesity may increase glomerular size and glomerular function abnormalities, which is often accompanied by rapid loss of renal function [12]. Additionally, obese patients (classification based on the Body Mass Index (BMI)) usually develop proteinuria and chronic renal failure after unilateral nephrectomy [13]. Anemia is a common complication in patients with CKD and its treatment has been based on a large body of evidence, suggesting that patients with the lowest hemoglobin values have worse outcomes than those with higher hemoglobin ones [14]. Otherwise, cigarette smoking is associated with accelerated progression of renal disease in patients with diabetic and non-diabetic nephropathy, along with an increased risk of cardiovascular difficulties [15].

Chronic kidney diseases constitute a public health problem, considering the increased prevalence of the kidney failure and early stages of chronic kidney disease [16]. The inappropriate use of drugs in patients with renal disease may be harmful and may have deleterious effects. Moreover, the decreased renal function is a well-known source for variability in drug response [17]. In the most recent years several decision support systems have been applied to various specific problems of renal disease, such as the prescriptions of pharmaceutical drugs in patients with reduced renal function [18, 19], the dosage adjustment for patients with renal disease [20]. Shemeikka et al. [19] developed a clinical decision support system to support prescriptions of pharmaceutical drugs in patients with reduced renal function. These authors state that the system has potential to improve kidney-related drug prescribing. Under this setting, Terrel et al. [20] developed a decision support system for medication dosing based on the estimation of patient’s creatinine clearance level. This system
contributed to reduce significantly the excessive dosage of target medications. Wei et al. [21] used data mining techniques to investigate renal disease and to analyze the differences among various administrative areas in order to further draw up a disease distribution map. Authors reported that the major disease forms for residents under the age of 60 were hypertension, hyperglycemia and hyperlipidemia. Di Noia et al. [22] presented a software tool based on artificial neural networks to classify patients’ health status potentially leading to end-stage kidney disease that could help physicians to increase the quality of medical decisions. These authors used ten predictors (i.e., in addition to gender, age at last follow-up, age at renal biopsy, age of onset, type of onset, histological grade, the input variables include serum creatinine, severity of chronic kidney disease, proteinuria, and hypertensive state, concerning the time of renal biopsy). The network was trained, tested and validated using a dataset composed by 587 instances and have the topology 10–6–1. The model accuracy, sensitivity and positive predictive value were 91.37, 70.76 and 75.15 % respectively. Akgundogdu et al. [23] developed a decision support system in order to diagnose renal failure. These authors used seven parameters (i.e., urea, creatinine, uric acid, potassium, calcium, phosphorus, and patient’s age). The sample comprised 112 patients, 64 with a positive diagnosis and the remaining with a negative one. It was used an Adaptive Neuro-Fuzzy Inference (ANFI) system with five layers and three rules, being the rules determined by the subtractive clustering method. Support Vector Machine (SVM) and Artificial Neural Networks (ANNs) were also used, aiming to evaluate the performance of the ANFI model.

The main drawbacks of approaches referred to above are related with the fact that they are not subject to a formal proof, and do not handle incomplete data, information or knowledge. Indeed, the early diagnosis of CKD predisposing should be correlated with many variables and requires a multidisciplinary approach. Consequently, it is difficult to assess the CKD predisposing since it needs to consider different conditions with intricate relations among them, where the available data may be incomplete, contradictory and/or unknown. In order to overcome these drawbacks, the present work reports the founding of a computational framework that uses knowledge representation and reasoning techniques to set the structure of the information and the associate inference mechanisms. We will centre on a Logic Programming (LP) method to knowledge representation and reasoning [24, 25], that is object of a formal proof, and look at a soft computing approach to knowledge processing based on ANNs [26].

Knowledge Representation and Reasoning

On the one hand, it is consensual that in some decisions that one may take, some information is not always exact in the sense that we handle estimated values, probabilistic measures, or degrees of uncertainty. On the other hand, knowledge and belief are generally incomplete, contradictory, or even error sensitive, being desirable to use formal tools to deal with the problems that arise from the use of partial, contradictory, ambiguous, imperfect, nebulous, or missing information [24, 25, 27]. Some general models have been presented where uncertainty is associated to the application of Probability Theory [28], Fuzzy Set Theory [29], Similarities [30, 31]. Other approaches for knowledge representation and reasoning have been proposed using the Logic Programming (LP) paradigm, namely in the area of Model Theory [32–34] and Proof Theory [24, 25].

In this work it is followed the proof theoretical approach in terms of an extension to the LP language to knowledge representation and reasoning. An Extended Logic Program (ELP) is a finite set of clauses in the form:

\[
\{ \neg p_1, p_1, \ldots, p_n, \neg q_1, \ldots, \neg q_m \} \cup \{ \text{scoring value} \} \]

where “?” is a domain atom denoting falsity, the \( p_1, q_j \) and \( p \) are classical ground literals, i.e., either positive atoms or atoms preceded by the classical negation sign \( \neg \) [24]. Under this formalism, every program is associated with a set of abducibles [33, 34] given here in the form of exceptions to the extensions of the predicates that make the program. The term \( \text{scoring value} \) stands for the relative weight of the extension of a specific predicate with respect to the extensions of the peers ones that make the overall program.

In order to evaluate the knowledge that stems from a logic program an evaluation of the Quality-of-Information (QoI) was set in dynamic environments aiming at decision-making purposes [35, 36]. The objective is to build a quantification process of QoI and an assessment of the argument values of a given predicate with relation to their domains (here understood as Degree-of-Confidence (DoC), which stands for one’s belief that its unknown values fits into the arguments ranges, taking into account their domains). The QoI with respect to the extension of a predicate \( i \) will be given by a truth-value in the interval [0, 1].

The universe of discourse can be engendered according to the information given in the logic programs that endorse the information about the problem under consideration, according to productions of the type:

\[
\text{extensions of predicate}_i = \bigcup_{1 \leq j \leq m} \text{clause}_j(x_1, \ldots, x_n) : \text{QoI}_i : \text{DoC}_i
\]

where \( \cup \) and \( m \) stand, respectively, for set union and the cardinality of the extension of \( \text{predicate}_i \). On the other hand,
$DoC_i$ denotes one’s confidence on the attribute’s values of a particular term of the extension of $\text{predicate}_i$, whose evaluation is given in [37].

### A Case Study

As a case study, consider a database given in terms of the extensions of the relations (or tables) depicted in Fig. 1, which stand for a situation where one has to manage information about kidney diseases detection. The tables include features obtained by both objective and subjective methods, i.e., the physicians will fill the tables that link to the **Chronic Kidney Disease** one while executing the health check. The clinics may populate some issues, others may be perceived by additional exams. Under this scenario some incomplete and/or unknown data is also available. For instance, in the **Patients Information** table, the GFR value in case 1 is unknown, while the **Primary Risk Factors** ranges in the interval [0, 1].

In the **Family Story** column of **Patients’ Information** table, in **Associated Kidney Diseases** and in **Primary and Secondary Risk Factors** tables 0 (zero) and 1 (one) denote, respectively, **no** and **yes**. In **Renal Biomarker** table 0 (zero), 1 (one) and 2 (two) denote, respectively, **normal**, **high** and **very high** values, while in the **Gender** column of **Chronic Kidney Diseases** table 0 (zero) and 1 (one) stand, respectively, for **female** and **male**. The **BMI** column in **Chronic Kidney Diseases** table is populated with 0 (zero), 1 (one) or 2 (two) according to patient’ Body Mass Index (BMI), evaluated using the equation $\text{BMI} = \text{Body Mass}/\text{Height}^2$. Thus, 0 (zero) denotes BMI$\leq$25; 1 (one) stands for a BMI ranging in interval [25, 30[$]; and 2 (two) denotes a BMI$\geq$30. The GFR column in **Patients Information** table presents the value of the glomerular filtration rate in cm$^3$/min/1.73 m$^2$ of body surface area.

The values presented in the **RB**, **AKD**, **PRF** and **SRF** columns of **Chronic Kidney Disease** table are the sum of the correspondent tables, ranging between [0, 4], [0, 4], [0, 2] and [0, 7], respectively.

Now, applying the rewritten algorithm presented in [37], to all the tables that make the **Extension of the Relational Model for Chronic Kidney Disease Diagnosis** (Fig. 1), excluding of such a process the **Chronic Kidney Disease** one, and looking to the DoCs values obtained in this manner, it is possible to set the arguments of the predicate referred to below, that also denotes the objective function with respect to the problem under analyze.

$$\text{ckd} : \text{Age, Gender, BodyMassIndex, FamilyStory, GFR, RenalBiomarkers, AssociatedKidneyDiseases, PrimaryRiskFactors, SecondaryRiskFactors} \rightarrow \{0, 1\}$$

where 0 (zero) and 1 (one) denote, respectively, the truth values **false** and **true**. Indeed, the arguments of this predicate where set by a process of sensibility analysis, where the arguments chosen where those that present the higher $DoC_i$, values, i.e., the ones that have a greater influence on the output of the objective function referred to above. Their terms also make the training and test sets of the Artificial Neural Network (ANN) given in Fig. 2.

---

![Fig. 1 An extension of the relational model for chronic kidney disease diagnosis](image-url)
Now, let us consider the first patient present in Fig. 1, that presents the feature vector \((Age=59, Gender=1, BMI=1, FS=0, GFR=\perp, RB=1, AKD=0, PRF=0, SRF=0)\), to which it is applied the rewritten algorithm presented in [37]. One may have:

\begin{verbatim}
Begin,
    The predicate’s extensions that make the Universe-of-Discourse are set according Chronic Kidney Disease table of Fig. 1--
    {ckd(Age, Gender, BMI, FS, GFR, RB, AKD, PRF, SRF)}
    not ckd(Age, Gender, BMI, FS, GFR, RB, AKD, PRF, SRF)
    ckd(59, 1, 1, 0, \perp, 1, 0, [0, 1], 0) :: 1 :: DoC
    ckd\{[59,59],[1,1],[0,0],[25,98],[1,1],[0,0],[1,0],[0,0]\}
    attribute’s values ranges :: 1 :: DoC
    [32,92][0,1][0,2][0,1][25,98][0,4][0,4][0,2][0,7]
    attribute’s domains
    ckd() :: 1
    The attribute’s boundaries are set to the interval \([0,1]\) --
    {ckd(Age, Gender, BMI, FS, GFR, RB, AKD, PRF, SRF)}
    not ckd(Age, Gender, BMI, FS, GFR, RB, AKD, PRF, SRF)
    ckd\{[0.45,0.45],[1,1],[0.5,0.5],[0,0],[0,1],[0,25,0.25],[0,0],[0,0.05],[0,0]\}
    attribute’s values ranges once normalized
    :: 1 :: DoC
    [0,1][0,1][0,1][0,1][0,1][0,1][0,1][0,1][0,1]
    attribute’s domains once normalized
    ) :: 1
    The DoC’s values are evaluated --
    {ckd(Age, Gender, BMI, FS, GFR, RB, AKD, PRF, SRF)}
    not ckd(Age, Gender, BMI, FS, GFR, RB, AKD, PRF, SRF)
    ckd\{1,1,1,1,0,1,1,0.87,1\} :: 1 :: 0.87
    ([0.45,0.45],[1,1],[0.5,0.5],[0,0],[0,1],[0,25,0.25],[0,0],[0,0.05],[0,0])
    attribute’s values ranges once normalized
    :: 1 :: DoC
    [0,1][0,1][0,1][0,1][0,1][0,1][0,1][0,1][0,1]
    attribute’s domains once normalized
    ) :: 1
    End.
\end{verbatim}

where its terms make the training and test sets of the Artificial Neural Network (ANN) given in Fig. 2.

\section*{Artificial Neural Networks}

The previously presented model of chronic kidney disease works well and demonstrate how all the information comes together to form a diagnosis. In this section, it is set a soft computing approach to model the universe of discourse of any patient suffering from chronic kidney disease, based on ANNs, which are used to structure data and capture complex relationships between inputs and outputs [38–40].

One of the main contributions of this work is related with the ability to deal with incomplete data/information. Besides to a classifier that enables the early detection of CKD predisposing, the approach intends also obtain the DoC associated to this diagnosis. Thus, is necessary apply an algorithm that allows more than one output variable. The choice fell on ANNs due to their dynamics characteristics like adaptability, robustness and flexibility. ANNs simulate the structure of the human brain, being populated by multiple layers of neurons, with a
valuable set of activation functions. As an example, let us consider the case listed above, where one may have a situation in which the diagnosis of chronic kidney disease is needed. In Fig. 2 it is shown how the normalized values of the interval boundaries and their DoCs and QoIs values (i.e., the tuple \((\text{minimum}, \text{maximum}, \text{DoC}, \text{QoI})\) work as inputs to the ANN. Exemplifying with the arguments Age, GFR and PRF of the ckd predicate, one may have \((0.45, 0.45, 1, 1); (0, 1, 0, 1); (0, 0.5, 0.87, 1)\). The output depicts a chronic kidney disease diagnostic, plus the confidence that one has on such a happening.

In this study 558 cases were considered with an age average of 51.7 years, ranging from 32 to 92 years old. CKD was observed in 175 cases, i.e., in 31.4% of the cases. The dataset holds information about the factors considered critical in the prediction of chronic kidney disease. Twenty four variables were selected allowing one to have a multivariable dataset with 558 records. These variables were grouped into five main categories, i.e., Patients’ Information, Renal Biomarkers, Associated Kidney Diseases and Primary and Secondary Risk Factors (Fig. 1). Thus, the number of variables used as input of the ANN model was reduced to nine (Table 1), i.e., the predicate’s arguments were workout according to a process of sensibility analysis, based on their DoCs values. A technique used to determine how different values of an independent variable will impact a particular dependent variable under a given set of assumptions.

Table 1  Variables description

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Data type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Patient’s age;</td>
<td>Numeric</td>
</tr>
<tr>
<td>Gender</td>
<td>Patient’s gender;</td>
<td>Nominal</td>
</tr>
<tr>
<td>BMI</td>
<td>Patient’s body mass index;</td>
<td>Nominal</td>
</tr>
<tr>
<td>Family story</td>
<td>Presence of the disease in any relative</td>
<td>Nominal</td>
</tr>
<tr>
<td>GFR</td>
<td>Presents the value of the glomerular filtration rate in cm(^3)/min/1.73 m(^2) of body surface area;</td>
<td>Numeric</td>
</tr>
<tr>
<td>Renal biomarkers</td>
<td>Includes information related with renal biomarkers, like serum creatinine, serum urea, serum cystatin C or albuminuria;</td>
<td>Nominal</td>
</tr>
<tr>
<td>Associated kidney diseases</td>
<td>Includes the issues related with associated kidney diseases, like the presence of glomerulonephritis, nephropathies, urinary outflow obstruction or polycystic kidneys;</td>
<td>Nominal</td>
</tr>
<tr>
<td>Primary risk factors</td>
<td>Presents information related with primary risk factors like hypertension or diabetes mellitus; and</td>
<td>Nominal</td>
</tr>
<tr>
<td>Secondary risk factors</td>
<td>Presents information related with secondary risk factors like dyslipidaemias, cardiovascular diseases, systemic lupus erythematosus, anaemia, neoplasms/myelomas, long term use of anti-inflammatory drugs or tobacco habits.</td>
<td>Nominal</td>
</tr>
</tbody>
</table>

To ensure statistical significance of the attained results, 25 (twenty five) experiments were applied in all tests. In each simulation, the available data was randomly divided into two mutually exclusive partitions, i.e., the training set with two thirds of the available data, used during the modeling phase, and the test set with the remaining cases, used after training in order to evaluate the model performance and to validate it. The back propagation algorithm was applied in the learning process of the ANN. The activation function used in the pre-processing layer was the identity one. In the other layers was used the sigmoid activation function.

A common tool to evaluate the results presented by the classification models is the coincidence matrix, a matrix of size \(L \times L\), where \(L\) denotes the number of possible classes. This matrix is created by matching the predicted and target values. \(L\) was set to 2 (two) in the present case. Table 2 presents the coincidence matrix (the values denote the average of the 25 experiments). A perusal of Table 2 shows that the model accuracy was 94.4% for the training set (355 correctly classified in 376) and 92.3% for test set (168 correctly classified in 182).

Based on coincidence matrix it is possible to compute sensitivity, specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) of the classifier:

\[
sensitivity = \frac{TP}{TP + FN}
\]

\[
specificity = \frac{TN}{TN + FP}
\]

\[
PPV = \frac{TP}{TP + FP}
\]

\[
NPV = \frac{TN}{TN + FN}
\]

where \(TP, FN, TN\) and \(FP\) stand, respectively, for true positive, false negative, true negative and false positive. Briefly, sensitivity and specificity are statistical measures of the performance of a binary classifier, while sensitivity measures the proportion of true positives that are correctly identified as such. Specificity measures the proportion of true negatives that are correctly identified. Moreover, it is necessary to know the probability of the classifier that give the correct diagnosis. Thus, it is also calculated both \(PPV\) and \(NPV\), while \(PPV\) stands for the Proportion of cases with Positive Values that were correctly diagnosed, \(NPV\) denotes the Proportion of cases with Negative Values that were successfully labeled.

The corresponding sensitivity, specificity, \(PPV\) and \(NPV\) values are displayed in Table 3 for training and test sets. A

Table 2  The coincidence matrix for ANN model

<table>
<thead>
<tr>
<th>Target</th>
<th>Predictive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training set</td>
</tr>
<tr>
<td></td>
<td>True (1)</td>
</tr>
<tr>
<td>True (1)</td>
<td>111</td>
</tr>
<tr>
<td>False (0)</td>
<td>15</td>
</tr>
</tbody>
</table>
perusal of Table 3 shows that the sensitivity ranges from 93.1 to 94.9 %, while the specificity ranges from 91.9 to 94.2 %.

The sensitivity of the proposed model is higher than the reported in literature [22]. PPV ranges from 84.4 to 88.1 %, while NPV ranges from 96.6 to 97.6 %. For comparison, the reported PPV values ranges between 75.2 and 92.8 % [22, 23]. Moreover, the Receiver Operating Characteristic (ROC) curves for the training and test sets are shown in Fig. 3. The areas under ROC curves are higher than 0.9 for both cases, denoting that the model exhibits a good performance in early recognition of CKD.

Serum urea and creatinine are key biomarkers to assess the integrity of renal function, however, they have poor predictive ability for renal disease since they are quite variable with the patient’s characteristics, once urea is largely dependent on the individual diet and creatinine is mutable with the muscle mass. Early diagnosis on the basis of presence of albuminuria or reduced estimated glomerular filtration rate could permit early intervention to reduce the risks of cardiovascular events, kidney failure, and death that are associated with CKD. Nevertheless, the screening for the kidney disorder is most efficient when targeted at high-risk groups, including elderly people and those with concomitant illness or a family history of chronic kidney disease.

The present model allows to integrate the results of the renal biochemical markers with other factors such as, family story of renal disease, presence of associated kidney diseases and integrate them with the main kidney sickness risk factors, allowing to be assertive in the diagnosis of CKD. This model showed a high sensibility, enabling the diagnosis of kidney disease comparing with the patients that really presented this pathology as well classifying properly the absence of renal pathology (i.e., specificity), therefore it can be a major contribution to the early recognition and prevention of chronic kidney disease.

Table 3  Sensitivity, specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) for the ANN model

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>94.9</td>
<td>94.2</td>
<td>88.1</td>
<td>97.6</td>
</tr>
<tr>
<td>Test set</td>
<td>93.1</td>
<td>91.9</td>
<td>84.4</td>
<td>96.6</td>
</tr>
</tbody>
</table>

Conclusions and Future Work

Kidney renal failure is often reversible if it is recognized early and treated promptly. The incidence of AKI and CKD is reaching epidemic proportions. In both situations, early intervention can significantly improve the prognosis. Once the life-threatening conditions are addressed, the healthcare institution information systems should turn to gathering a complete patient history and performing a detailed examination of such records, like signs/symptoms, allergies, medications, past medical history, last oral intake and events preceding the call for help. However, to set an early diagnosis of the CKD predisposing is a hard and complex task, which needs to consider many different factors, where some of them are not represented by fully objective data. Being an area filled with incomplete and unknown data, information or knowledge it may be tackled by Artificial Intelligence based methodologies and techniques for problem solving. Indeed, this work presents the founding of a computational framework that uses powerful knowledge representation and reasoning techniques to set the structure of the information and the associate inference mechanisms. A method that brings a new approach that can revolutionize prediction tools in all its variants, making it more complete than the existing methodologies and tools available. The knowledge representation and reasoning techniques presented above are very versatile and capable of covering every possible instance by considering incomplete, contradictory, and even unknown data.
The model presented in this study showed a good performance in the detection of chronic kidney disease predisposing, since their sensitivity and specificity exhibited values higher than 90%. These findings were corroborated by the area under ROC curves (>0.9). The main contribution of this work relies on the fact that at the end, the extensions of the predicates that make the universe of discourse are given in terms of DoCs values that stand for one’s confidence that the predicates arguments values fit into their observable ranges, taking into account their domains. It also encapsulates in itself a new vision of Multi-value Logics, once a proof of a theorem in a conventional way, is evaluated to the interval [0, 1]. The vision of Multi-value Logics, once a proof of a theorem in a conventional way, is evaluated to the interval [0, 1]. The ANN were the selected due to their dynamics characteristics like adaptability, robustness and flexibility. Future work may recommend that the same problem must be approached using others computational formalisms like Genetic Programming [25], Case Based Reasoning [41] or Particle Swarm [42], just to name a few.

References


