

An Intelligent Web-Based Decision Support System for Monitoring the Chronic Kidney Disease in Brazilian Communities

Álvaro Sobrinho (✉ alvaro.alvares@ufape.edu.br)

Federal University of the Agreste of Pernambuco

Andressa C. M. da S. Queiroz

Federal University of Alagoas

Gyovanne Bezerra Cavalcanti

Federal Rural University of the Semiárid

Josaias de Moura Silva

Federal Rural University of the Semiárid

Leandro Dias da Silva

Federal University of Alagoas

Evandro de Barros Costa

Federal University of Alagoas

Angelo Perkusich

Federal University of Campina Grande

Maria Eliete Pinheiro

Federal University of Alagoas

Research Article

Keywords: Primary Care, Clinical Document Architecture, Machine Learning, Early Diagnosis

Posted Date: December 31st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-125006/v2>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Chronic Kidney Disease (CKD) is a worldwide public health problem, usually diagnosed in the late stages of the disease, increasing public health costs and mortality rates. The late diagnosis is even more critical in developing countries due to the high levels of poverty, a large number of hard-to-reach locations, and sometimes lack/precarious primary care.

Methods: We designed and evaluated an intelligent web-based Decision Support System (DSS) using the J48 decision tree machine learning algorithm, knowledge-based system concepts, the clinical document architecture, Cohen's kappa statistic, and interviews with an experienced nephrologist.

Results: We provided a DSS methodology that guided the development of the system, that provides remote monitoring features, to assist patients, primary care physicians, and the government in identifying and monitoring the CKD in Brazilian communities. A CKD dataset enabled the training and evaluation of the J48 decision tree algorithm, while Cohen's kappa statistic guided the evaluation of the knowledge-based system by interviews with an experienced nephrologist.

Conclusion: The DSS facilitates the identification and monitoring of the CKD considering low-income populations in Brazil. In addition, the methodology and DSS can be reused in other developing countries with similar scenarios.

1. Introduction

The high prevalence and mortality rates of persons with chronic diseases, for example, the Chronic Kidney Disease (CKD) [1], are real-world public health problems. The World Health Organization (WHO) estimated that chronic diseases would cause 60 percent of the deaths reported in 2005, 80 percent in low-income and lower-middle-income countries, increasing to 66.7 percent in 2020 [2]. According to the WHO health statistics 2019 [3], people who live in low-income and lower-middle-income countries have a higher probability of dying prematurely from known chronic diseases such as Diabetes Mellitus (DM).

For the specific case of CKD, the early identification and monitoring of this disease and its risk factors reduce the CKD progression and prevent adverse events, such as sudden development of diabetic nephropathy. The present study considers the CKD identification and monitoring focusing on people living in Brazil, a continental-size developing country. Developing countries stand for low- and middle-income regions, while developed countries are high-income regions, such as the USA [4]. The population of developing countries suffers from increased mortality rates caused by chronic diseases, e.g., CKD, Arterial Hypertension (AH), and DM [5]. AH and DM are two of the most common CKD risk factors. People with type 1 or type 2 DM are at high risk of developing diabetic nephropathy [6], while severe AH cases may increase kidney damage. For example, about 10 percent of the adult Brazilian population is aware of having some kidney damage, while about 70 percent remains undiagnosed [7].

The CKD is characterized by permanent damage, reducing the kidneys' excretory function, easily measured using the glomerular filtration [8]. However, the diagnosis usually occurs during more advanced stages because it is asymptomatic, postponing the application of counter measure, decreasing people's quality of life, and possibly leading to lethal kidney damage. For example, in 2010, about 650-500 people per million of the Brazilian population faced dialysis and kidney transplantation [9]. This number has grown, warning governments about the relevance of the CKD early diagnosis. In 2016, according to the Brazilian chronic dialysis survey, the number of patients under dialysis was 122,825.00, increasing this number by 31,000.00 in the last five years [10]. In 2017, the prevalence and incidence rates of patients under dialysis were 610 and 194 per million population [11]. The incidence continued to be high in 2018 (133,464.00) [12].

The high prevalence and incidence of dialysis and kidney transplantation increase cost with public health. Therefore, the CKD has an expressive impact in the health economics perspective [13]. For instance, the Brazilian Ministry of Health reported that the transplantation and its procedures resulted in spending about 720 million reais in 2008 and 1.3 billion in 2015 [14]. The costs and the high rate of persons waiting for transplantation suggest the increased public spent on kidney diseases. Preventing CKD has a relevant role in reducing mortality rates and public health costs [15]. The CKD early diagnosis is even more challenging for people who live in remote and hard-to-reach settings because of either lack of or precarious primary care.

Organizations such as the Brazilian society of nephrology and the national kidney foundation have proposed tools to support the CKD diagnosis, assisting physicians in identifying kidney damage by estimating the glomerular filtration rate (GFR). The Cockcroft-Gault [16] and the modification of diet in renal disease (MDRD) [17] are classical equations to estimate the GFR. This type of tool assists the CKD diagnosis when physicians have adequate and simple access during clinical evaluations; however, this is not always the reality, mainly for developing countries. For example, Brazil, a continental-size country, has many problems related to computer-assisted healthcare compared to developed countries that maintain Electronic Health Records (EHRs) infrastructures. EHRs are composed of large amounts of data, enabling analysis empowered by machine learning techniques to assist physicians during clinical evaluations [18]. Unfortunately, some remote settings, e.g., the Amazon Jungle, are subject to precarious public health and the absence of EHR, sometimes even facing the lack of primary care physicians.

In this study, we present a decision support system (DSS) methodology and propose an intelligent web-based DSS to assist patients, primary care physicians, and the government in identifying and monitoring the CKD in Brazilian communities. Thus, the system named MultiCare may be integrated with existing government systems, e.g., the Brazilian SUS, to address increased mortality rates and public health costs. By providing continuous and remote monitoring, the DSS methodology addresses precarious public health, the absence of EHR, and the lack of primary care physicians. The study extends results of previous research articles [19,4], consisting of two main new contributions: (i) an intelligent DSS available for patients, physicians, and government to assist the CKD monitoring using machine learning and knowledge-based system concepts; and (ii) a methodology to design DSS for identifying and monitoring chronic diseases in Brazilian communities. These contributions represent state-of-the-art advances, considering that results presented in [19] do not consider machine

learning, knowledge-based system concepts (knowledge modeling of a Brazilian expert), and functionalities for physicians and government. Besides, the results presented in [4] only relate to a comparative analysis of classifiers.

Nowadays, considering the increasing need for social distancing required by epidemics such as the COVID-19, the remote identification and monitoring of chronic diseases are relevant to prevent increasing morbidity and mortality rates due to possible infections and lack of treatment for existing conditions. Healthcare systems designed for these purposes are powered by artificial intelligence techniques such as machine learning [20].

Decision support frameworks and systems have received the attention of researchers in the last years. For instance, Li et al. [21] present a utilization-based knowledge discovery framework to assist the assessment of healthcare access, meaning the presence of potential resources or actual use of healthcare services. The authors evaluate the framework by instantiating a DSS to analyze physician shortage areas. Hsu [22] describes a framework based on a ranking and feature selection algorithm to assist physicians' decision-making on the most relevant risk factors for cardiovascular diseases. The author also applies machine learning techniques to enable identifying the risk factors.

Walczak and Velanovich [23] developed an artificial neural network (ANN) system to assist physicians and patients' decision-making in selecting the optimal treatment of pancreatic cancer. The system determines the 7-month survival or mortality of patients based on a specific treatment decision. Topuz et al. [24] propose a decision support methodology guided by a Bayesian belief network algorithm to predict kidney transplantation's graft survival. The authors use a database with more than 31,000 U.S. patients and argue that the methodology can be reused to other datasets.

Wang et al. [25] evaluate a murine model, induced by intravenous Adriamycin injection, using optical coherence tomography (OCT) to assess the CKD progression by image of rat kidneys. The authors highlight that OCT images contain relevant data about kidney histopathology. Jahantigh, Malmir, and Avilaq [26] propose a fuzzy expert system to assist the medical diagnosis, focusing initially on kidney diseases. The system is guided by the experience of physicians to indicate disease profiles. Neves et al. [27] present a DSS to assist in identifying acute kidney injury and CKD using knowledge representation and reasoning procedures based on logic programming and ANN. Polat et al. [28] used the support vector machine technique and the two feature selection methods wrapper and filter to conduct the CKD identification in the early stages. The authors justify the computed-aided diagnosis based on high mortality rates of CKD. Finally, Arulanthu and Perumal [29] presented a DSS for CKD prediction (CKD or non-CKD) using a logistic regression model.

However, these CKD studies have some limitations. For example, no one considers the monitoring of CKD risk factors. Additionally, the solutions do not apply well-accepted standards to simplify the representation and sharing of evaluation results, such as the health level 7 (HL7) clinical document architecture (CDA) [30]. Other relevant topics to point out are the machine learning technique used to identify the disease and the costs of required examinations (predictors). Most of the studies use a large number of predictors and apply complex analysis, increasing costs, and making difficult the double-checking of DSS results by physicians. Indeed, this type of functionality is relevant because other clinical conditions influence the CKD, and the diagnosis is usually improved when physicians collaborate to conclude. Integrating functionalities focusing on patients and physicians is also not considered by the previous studies.

2. Materials And Methods

2.1 Data analysis

We primarily selected the dataset features based on medical guidelines, specifically, the KDIGO guideline [31], the national institute for health and care excellence guideline [32], and the KDOQI guideline [33]. Besides, we interviewed a set of Brazilian nephrologists to confirm the relevance of the features in the context of Brazil. The final set of CKD features focusing on Brazilian communities included AH, DM, creatinine, urea, albuminuria, age, gender, and GFR. Table 1 presents descriptions and types of features of the dataset.

Table 1. Feature description for the dataset.

Feature	Type	Description
AH	Integer	The subject presents arterial hypertension.
DM	Boolean	The subject presents diabetes mellitus. The number 0 represents absence, while 1 represents presence.
Crea	Real	Result of blood test used to assess kidney function.
Urea	Real	Result of blood test of a substance produced (mainly) by the liver.
Albu	Real	Result of a test to measure the amount of albumin in the urine.
Age	Integer	The age of the subject.
Gender	Integer	The number 0 represents male, while 1 represents female.
GFR	Real	Result of the glomerular filtration rate of the subject.

In a previous study [19], we collected medical data (60 real-world medical records) from physical medical records of adult subjects (age ≥ 18) under the treatment of University Hospital Prof. Alberto Antunes of the UFAL, Brazil. The data collection from medical records maintained in a non-electronic format at the hospital was approved by the Brazilian ethics committee of UFAL, and conducted between 2015 and 2016. The dataset comprises 16 subjects with no kidney damage, 14 subjects diagnosed only with CKD, and 30 subjects diagnosed with CKD, AH, and/or DM. In general, the sample included subjects with ages between 31 and 79 years; approximately 94.5% of the subjects were diagnosed with AH, and 58.82% were diagnosed with DM. Table 2 presents a sample of the 60 real-world medical records, related to the four risk classes: low risk (30 records), moderate risk (11 records), high risk (16 records), and very high risk

(3 records). An experienced nephrologist, with more than 30 years of CKD treatment and diagnosis in Brazil, labeled the risk classification based on the KDIGO guideline.

Table 2. Sample of the dataset collected from University Hospital Prof. Alberto Antunes. Each ID represents a specific subject from the dataset.

ID	DM	AH	Crea	Urea	Albu	Age	Gender	GFR	Risk
1	False	True	1.00	27.0	0.4	50	1	89.0	Low
2	False	False	1.04	14.9	9.5	39	1	93.5	Low
3	False	False	0.53	23.0	4.0	59	1	104.6	Low
4	False	True	0.65	34.0	7.10	31	1	118.7	Low
5	False	True	0.65	31.0	4.0	61	1	89.3	Low
6	False	True	0.90	22.0	9.45	45	1	134.2	Low
7	False	True	0.75	29.5	4.90	53	1	82.1	Low
8	False	True	1.40	32.0	14.0	48	0	63.8	Low
9	False	True	0.75	27.0	22.0	73	1	75.4	Low
10	False	False	0.84	43.0	5.5	54	0	84.6	Low

The dataset did not contain duplicated and missing values. We only translated the dataset to English and converted the gender of subjects from string to a binary representation to enable the usage of the J48 decision tree algorithm. In addition, only for the training set, we augmented the dataset to decrease the impact of imbalanced data and improve the data analysis (more 54 records) by duplicating real-world medical records and carefully modifying the features, i.e., increasing each CKD biomarker by 0.5. We selected the constant 0.5 with no other purpose than to differentiate the instances and maintain the new one with the same label of the original. The perturbation of the data did not result in unacceptable ranges of values and incorrect labeling. The total number consisted of 108 for training and 6 records for testing. The experienced nephrologist verified the validity of the augmented data by analyzing each record regarding the correct risk classification (i.e., low, moderate, high, or very high risk). As stated above, the experienced nephrologist also evaluated the 60 real-world medical records. The preprocessed original and augmented datasets are available in a public repository [34].

To increase confidence in the system, after development, we conducted tests using the test set and performance metrics (i.e., correctly classified instances, incorrectly classified instances, precision, Precision-Recall Curve (PRC) area, and Receiver Operating Characteristic (ROC) area). We used the augmented training set to define a decision tree model based on the J48 algorithm. For evaluating the model, we applied the 10-fold cross-validation using the Weka® software. We used the Knowledge Flow interface of the Weka® to handle the 10 folds during the data augmentation to ensure that the test set only contained unseen real data. Afterward, we used the risk assessment model and the Weka® application programming interface for Java to develop the DSS, extending the results presented in [19].

2.2 Simulated scenarios and interviews

Also extending the results presented in [19], we used concepts of knowledge-based systems to design functionality to verify an emergency of a patient with hyperglycemia, hypoglycemia, hyperkalemia, or hypokalemia. We defined a knowledge base by analyzing medical guidelines and interviewing a nephrologist with more than 30 years of teaching and treating patients with CKD and DM. We interviewed the experienced nephrologist to evaluate the knowledge-based system after developing the DSS. We simulated a total of 112 scenarios (i.e., fictitious subjects) considering the knowledge base designed and the risk of hyperglycemia, hypoglycemia, hyperkalemia, and hypokalemia.

To conduct the evaluation, we measured Cohen's kappa statistic by calculating the gross agreement and the kappa concordance index with 95% confidence intervals without adjusting for the bias and prevalence. The analysis considers the following k indices: $k < 0$, no agreement; k between 0 and 0.19, poor agreement; k between 0.20 and 0.39, low agreement; k between 0.40 and 0.59, moderate agreement; k between 0.60 and 0.79, substantial agreement; and k between 0.80 and 1, nearly perfect agreement. We used Cohen's kappa statistic to compare the knowledge-based system's results with the nephrologist evaluation when considering all the simulated scenarios.

3. Results

3.1 DSS methodology

We show in Fig. 1 a schema for the proposed methodology to design DSS for identifying and monitoring CKD in Brazilian communities. Three actors interact with the DSS generated following the methodology: Physician (internal), Patient (internal), and Government Health System (external). This type of methodology is relevant because developing countries such as Brazil usually suffer from precarious primary health care in specific settings, e.g., hard-to-reach and rural settings

3.1.1 System for patients

We divided this methodological step into two main tasks, which may be conducted simultaneously, for designing the front-end and back-end of web-based systems used by patients. The system should contain personal health records (PHR) and risk assessment functionalities. The risk assessment of the monitored chronic disease is based on the machine learning technique of decision tree analysis. A decision tree is suitable for this type of DSS because it is a white-box analysis approach, enabling physicians to double-check the patient's system's risk assessments quickly. In a previous study [4], we compared existing risk assessment models, showing the suitability of decision tree models for the context of developing countries. Once the patient's system evaluates the user's clinical situation, it sends a clinical document, structured using the HL7 CDA, to the physician responsible for monitoring the patient. The HL7 CDA document is an XML file that contains the risk analysis data, a risk analysis decision tree, and the PHR. Being a web-based system, patients can use it in remote and hard-to-reach settings using different devices, such as desktop computers, smartphones, and tablets. The computer-based risk assessment is used to address precarious public health problems, lack of EHRs, and lack of primary care physicians related to remote and hard-to-reach settings in developing countries, e.g., reaching people who live in the Brazilian Amazonas' state. The PHR is continuously sent to a central data server to update the patient's medical records.

3.1.2 System for physicians

We defined two main tasks for this methodological step, which may be conducted simultaneously to design the front-end and back-end web-based systems used by physicians. The system should enable physicians to receive CDA documents from the patient's system to double-check the chronic disease risk assessment, conduct the final diagnosis using the risk analysis data, risk analysis decision tree, and PHR. The decision tree is relevant for physicians to perform a step-by-step verification of the initial risk assessment. From these initial data, in case of uncertainty about the diagnosis, the physician may use the system to include more specific tests in the CDA document and send it to other physicians to get second opinions until a more precise diagnosis is reached. When the physician concludes, the system updates data into the patient's medical records maintained using the clinical data server. This system is relevant to enable the remote evaluation (i.e., a simple analysis of risk assessment results provided by the model) of people who live in remote and hard-to-reach-settings.

3.1.2 Government Health System

Considering the maintenance of PHR, the continuous usage of the system by patients and physicians results in a large and centralized dataset of users under monitoring in remote and hard-to-reach settings. The subsystems handled by patients and physicians use a server subsystem's web services, aiming to update the local data into the centralized dataset. External government health systems can benefit from the centralized dataset by applying machine learning techniques, generating relevant information for planning public policies, e.g., conducting disease awareness marketing tactics for preventing chronic disease, focusing on settings that present high incidence.

3.2 WebMultCare

The proposed system is composed of three main subsystems: Patient, Medical, and Server. The Patient subsystem is composed of functionalities to handle, among other elements, glucose, and blood pressure sensors, acquiring data related to DM and AH. These data are recorded locally and sent to a database by the Server subsystem. When the CKD risk is identified, alerts and the patient's data are sent to the Medical subsystem; a subsystem used by a physician in a healthcare environment. Thus, the Medical subsystem enables physicians to analyze the risk analysis data and the patient's PHR, updating/confirming the patient's clinical condition under monitoring using the Server subsystem. As an example of a health system, the Brazilian SUS can reuse the patient's central data for planning public policies.

The architecture of the WebMultCare was defined following the attribute-driven design method [35], and guided by the architectural drivers' modifiability, portability, scalability, availability, and interoperability. The system is based on the model-view-controller (MVC) pattern and architectural tactics called semantic coherence and information hiding to achieve modifiability and portability. In contrast, we use the client-server architectural pattern and web services to improve scalability, availability, and interoperability.

3.2.1 Patient subsystem

A previous Android version of the Patient subsystem was presented in [19], including formal specifications, effectiveness evaluation, and usability tests. The usability tests showed some limitations that motivated the re-engineering of the subsystem based on web technologies. Additionally, the version presented in this article improves the CKD risk analysis using machine learning and knowledge-based system concepts. For instance, the system provides a new feature to refer patients with specific emergencies (i.e., hyperglycemia, hypoglycemia, hypokalemia, and hyperkalemia) to an adequate healthcare facility using a knowledge base when visiting an unknown location.

The back-end of the Patient subsystem was implemented using Java and web services. The subsystem comprises the following main features: access control, management of ingested drugs, management of allergies, management of examinations, monitoring of AH and DM, execution of risk analysis, generation and sharing CDA documents, and analysis of the emergency. In contrast, the front-end of the Patient subsystem is implemented using HTML 5, Bootstrap, JavaScript, and Vue.js. Fig. 2a illustrates the graphical user interface (GUI) for recording a new CKD test result (the main inputs for the risk assessment model). The user can also upload an XML file containing the test results to present a large number of manual inputs. Once the patient provides the current test results, the main GUI of the Patient subsystem is updated, showing the test results available for the risk assessment.

Fig. 2b illustrates the main GUI of the Patient subsystem, describing the creatinine, urea, albuminuria, and GFR (i.e., the main features used by the risk assessment model). This study reduces the number of required test results to conduct the CKD risk analysis from 5 to 4 compared to the previously published research [19]. This is critical for low-income populations using the Patient subsystem. The subsystem provides a new CKD risk analysis when the patient inputs all CKD features.

During the CKD risk analysis (conducted when all tests are available), and based on the presence/absence of DM, presence/absence of AH, age, and gender, the J48 decision tree algorithm classifies the patient's situation considering four classes: low risk, moderate risk, high risk, and very high risk. In case of moderate risk, high risk, or very high risk, the subsystem packages the classification results as a CDA document, along with the decision tree graphic and general data of the patient. The Patient subsystem alerts the physician responsible for the patient and sends the complete CDA document (i.e., the main output of the DSS) for further clinical analysis. In case of low risk, the Patient subsystem only records the risk analysis results to keep track of the patient's clinical situation. It does not send the physician alert, automating the risk analysis and sharing, previously requested to the users by button events [19]. In this article, the data of 114 records, available in the same CKD dataset used in [4], guided the training of the J48 decision tree algorithm to define the final risk assessment model embedded in the proposed DSS. We experimented with modifying the parameters of the J48 decision tree algorithm to improve accuracy. Thus, we configured the split point, preventing the scanning of the entire dataset for the closest data value (relocation). For the remaining parameters, we used the default values of the J48 Weka© package.

Results presented in a previous study [4] justify the usage of the J48 decision tree algorithm and features (i.e., presence/absence of DM, presence/absence of AH, creatinine, urea, albuminuria, age, gender, and GFR) to conduct risk analyses in developing countries. The physician responsible for the healthcare of a specific patient can, remotely, access the CDA document by Medical subsystem, re-evaluate or confirm the risk analysis (i.e., preliminary diagnosis) provided by the Patient subsystem, and share the data with other physicians to get second opinions. If the physician confirms the preliminary diagnosis, the patient can continue using the Patient subsystem to prevent the CKD progression, including the monitoring of risk factors (DM and AH), CKD stage, and risk level.

Besides, the Patient subsystem includes a knowledge-based system to refer the patient with CKD and risk factors to an adequate healthcare facility at an emergency, as another new contribution from [19]. This feature considers the patient's scenario outside his/her county and does not know the correct facility for treatment, according to the current health situation. Based on semi-structural interviews with an experienced nephrologist that has treated patients in Brazil for more than 30 years, we addressed the following topics: (i) possible emergency care locations; (ii) pathology to be identified; (iii) symptoms; and (iv) associated drugs. For hyperglycemia, hypoglycemia, hyperkalemia, and hypokalemia, the system can refer the patient to emergency care units (ECU) or hospital emergencies, based on the current patient's health condition.

The knowledge base defined for the knowledge-based system comprises data collected from medical guidelines and semi-structured interviews with the nephrologist. The data relates to the symptoms that patients may present and risk factors that can cause health conditions (e.g., specific drugs). Fig. 3 describes the first decisions used to identify the risk of hyperglycemia, hypoglycemia, hyperkalemia, or hypokalemia. Nausea is a symptom shared by all clinical conditions, and each including symptom helps identify a specific condition.

In addition to the symptoms, the excessive consumption of alcohol, and excessive quantity of insulin, may increase the risk of hypoglycemia. For all clinical conditions, the usage of specific drugs may also result in the clinical conditions considered. The possible ingestion of a drug is a relevant indication of the risk of a specific clinical condition. Fig. 4 describes the commonly ingested drugs that may lead to hyperglycemia, hypoglycemia, hyperkalemia, and hypokalemia.

Fig. 5 illustrates, as a tree, a summary of the relationships between the questions presented in the DSS, considered during the identification of hyperglycemia (left side of Fig. 3). We generated the tree from the knowledge base to present an overview of a sample of the knowledge-based system that composes the DSS. Hyperglycemia is a common clinical condition in patients who have DM. The rule base for hypoglycemia, hyperkalemia, and hypokalemia, is defined similarly to the example of Fig. 5, differing by specific tests, symptoms, and ingested drugs. Finally, Fig. 6 shows a view of the GUI of the knowledge-based system in a risk scenario of hyperglycemia. Whenever facing an emergency, the patient can provide information about his/her current clinical condition, enabling the DSS to identify the emergency and recommend a healthcare unit (another example of the DSS). In this case, after asking about specific symptoms, the patient is required to inform if he/she ingested some drugs to increase confidence in the evaluation, following the relationships presented in Fig. 5.

3.2.2 Medical subsystem

On the one hand, the back-end of the Medical subsystem is implemented using Java, Spring MVC framework, and Drools (a business rules management system). The subsystem comprises the following main features: (i) access control; (ii) management of CDA documents; (iii) control version of CDA documents; (iv) sharing of CDA documents; (v) history of CDA documents versions and (vi) re-evaluation of risk analysis. On the other hand, the front-end of the Medical subsystem is implemented using the HTML 5, CSS, JavaScript, Bootstrap, and Java server pages. After validating his/her credentials, the system directs the doctor to the main GUI, displaying a brief presentation of the available features to handle clinical documents.

Two scenarios guide the usage of the Medical subsystem: creating a new clinical document and evaluating an existing clinical document. Fig. 7a illustrates the feature of creating a new clinical document that enables physicians to start evaluating a patient without the dependency on data received from the Patient subsystem. Fig. 7b illustrates that the physicians are requested to provide the risk assessment for patients guided by the classifications proposed in well-accepted international medical guidelines. In contrast, the evaluation of an existing clinical document relies on data received from the Patient subsystem, which performs the risk assessments of patients. The remote monitoring feature is relevant to address precarious public health, the absence of EHRs, and the lack of primary care physicians in Brazilian communities. Suppose a moderate risk, high risk, or very high risk is identified by the Patient subsystem. In that case, physicians receive general data and the risk assessment conducted using the J48 decision tree algorithm, enabling the final evaluation or the interaction with other physicians to improve confidence in a suspicious clinical situation. When clinical documents are available, physicians can perform version control to access current and past documents—the version control helps keep track of the history of clinical evaluations of patients. The re-evaluation of only a subset of patients (referred by the patient's system) can reduce the burden (or inefficiency) of the public health.

3.2.3 Server subsystem

A real-time database supports the central data server, assisting data analysis by patients, physicians, and the government. The Patient and Medical subsystems use web services provided by the Server subsystem to update the PHR of patients as part of the medical records available in a healthcare facility. Therefore, the government can conduct data mining, which is relevant to enable the analysis of a large number of data to support the planning and execution of public health policies. For example, it is possible to identify locations that require educational activities to prevent worsening mortality rates.

3.3 Evaluation

3.3.1 Machine Learning Model

A CKD dataset guided the evaluation of the patient

$s \subset \text{system of the DSS when conduct } \in g \text{ the CKD risk assessment } a \text{ or } d \in g \rightarrow \text{low-risk, moderate risk, high risk, and very high risk. The data } a$

s subsystem. When using the 10-fold cross-validation, the model presented high accuracy (i.e., 95.00%). The 10-fold cross-validation was executed 5 times, showing stability. The J48 decision tree presented a precision of 0.97, ROC area of 0.96, and PRC area of 0.94.

Table 3. Evaluation results, using the performance metrics, of the CKD risk assessment conducted by the patient's subsystem.

Metric	Result
Correctly classified instances (%)	95.00
Incorrectly classified instances (%)	5.00
Precision	0.97
Precision-recall curve area	0.94
Receiver operating characteristic area	0.96

3.3.2 Knowledge-based system

Also, the complete system was presented to the experienced nephrologist, confirming the completeness of the requirements. For instance, the knowledge base and questionings were presented to a nephrologist with more than 30 years of teaching and treating patients with CKD and DM to evaluate the knowledge-based system as part of the DSS. The nephrologist reviewed the knowledge base and questionings, validating the final version of the DSS. The nephrologist analyzed simulated data to evaluate the risk of hyperglycemia, hypoglycemia, hyperkalemia, and hypokalemia. The same data was analyzed using the DSS to compare risk assessments. Table 4 presents a sample of the total 112 simulated scenarios used to evaluate the knowledge-based system, covering 7 of the 21 paths of Fig. 5. This sample relates to fictitious subjects with a risk of hyperglycemia.

Table 4. Sample of hyperglycemia simulated scenarios used to evaluate the knowledge-based system.

ID	Symptoms	Drugs	DM	Glucose	Refer
1	All symptoms	No	-	-	ECU
2	All symptoms	Yes	No	-	ECU
3	All symptoms	Yes	No	120	ECU
4	All symptoms	Yes	No	126	Hospital
5	All symptoms	Yes	Yes	-	ECU
6	All symptoms	Yes	Yes	60	ECU
7	All symptoms	Yes	Yes	250	Hospital

To conduct the evaluation, Cohen's kappa statistic was measured by calculating the gross agreement and the kappa concordance. This task consisted of two steps of evaluations with the experienced nephrologist using Cohen's kappa. In the first step (Table 5), the knowledge-based system only achieved substantial agreement ($k = 0.6821$) and moderate agreement ($k = 0.5962$) with the nephrologist for risk classification and refer, respectively. The main cause of the disagreement was that the nephrologist considered some of the scenarios of hyperkalemia and hypokalemia risks as inconclusive (Table 7, column 2). In the second step, the knowledge base was corrected, and the evaluation resulted in 100% concordance with the opinion of the experienced nephrologist.

Table 5. Results of the first evaluation step of the knowledge-based system by comparing it with the opinion of an experienced nephrologist using the kappa statistic.

Risk	k By Risk	Refer	k By Refer
Hyperglycemia	0.9361	ECU	0.5880
Hypoglycemia	0.8886	Hospital	0.5547
Hyperkalemia	0.5089	Inconclusive	1.0000
Hypokalemia	0.7404		
Inconclusive	0.0290		
Global Kappa	0.6903		0.5962

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

4. Discussion

During monitoring of the CKD, based on the decision tree model, assuming the previous evaluation of DM, a user only needs to periodically conduct two blood tests: creatinine and urea. Albuminuria is measured using a urine test, while GFR can be calculated with the Cockcroft-Gault equation. The reduced number of examinations is relevant for developing countries such as Brazil, given the high levels of poverty of the population. However, it is also relevant to discuss the impacts of the reduced number of features during the training and testing phases.

Table 6 describes the incorrectly classified instances identified when testing the decision tree model. For example, the decision tree model disagreed with the most-experienced nephrologist, stating moderate risk instead of very high risk for the patient with ID 59. However, the model did not conduct any critical underestimations of the risk situation of subjects (e.g., low risk instead of moderate risk). This would be a critical problem because a patient is usually referred to a nephrologist when there is a moderate or higher risk. The remaining misleading classifications are less harmful because they still result in the referral of the patient under evaluation, even slighting overestimating/underestimating the risk.

Table 6. Incorrectly classified instances by the J48 decision tree model.

ID	Nephrologist Evaluation	Decision Tree Evaluation
46	High	Very High
58	Very High	Moderate
59	Very High	Moderate

Besides the use of a reduced number of features and the absence of critical underestimations, another advantage of a decision tree model is the easy interpretation of results. The easy interpretation of CKD risk analyses by nephrologists and primary care physicians who need to conduct further examinations to confirm the clinical situation of a patient is a critical factor for the reuse of the model in real-world situations. For example, Fig. 8 shows the decision tree generated by the J48 decision tree algorithm, which comprises each CKD biomarker considered and the related classification defined by the Weka® software. A physician only follows the decisions to interpret the logic of the classification. From the 8 CKD features, the J48 decision tree used only 4 to classify the risk (i.e., DM, GFR, albuminuria, and age), requiring 1 blood test and 1 urine test when DM is already evaluated, at the cost of 4 incorrectly classified instances. The DM (1), GFR (2), and albuminuria (3) features had the same prediction power for the experienced nephrologist and the decision tree model, considering a scale ranging from 1 (highest priority) to 8 (lowest priority); however, the nephrologist prioritized and used the creatinine (4), urea (5), gender (6), AH (7), and age (8) features.

In addition to the machine learning model, testing all the 112 simulated scenarios of the knowledge-based system (as part of the DSS) covered all possible paths from initial to end nodes of the decision tree (e.g., Fig. 5) representing the complete knowledge base, increasing confidence in the defined rules. Evaluating the knowledge-based system was relevant to increase confidence in the DSS when conducting the risk evaluation of a user with hyperglycemia, hypoglycemia, hyperkalemia, or hypokalemia. This type of functionality can decrease the negative impacts on the health condition of patients during emergency situations. For example, carrying out some palliative control measures during the transit to a health care facility may reduce the impact of low glucose levels in a hypoglycemia emergency.

However, the evaluation of the DSS has some limitations. The size of the dataset used to evaluate the CKD risk is limited. The k-fold cross validation assisted in reducing the impact of this limitation. Besides, the current evaluation of the knowledge-based system did not consider the opinions of patients. The interviews with the experienced nephrologist were relevant to decrease the negative impacts of this limitation.

As future work, the evaluation of the DSS needs to be improved. Testing the risk assessment requires improving the number of subjects in the CKD dataset. Besides, testing the knowledge-based system also requires interviews with patients to evaluate the GUI design, considering usability and user perceptions. We also envision to conduct a clinical study on the usage of the system to compare a group that used the system with a control group.

5. Conclusions

The intelligent web-based DSS presented in this article helps patients, physicians, and the government identify and monitor the CKD and risk factors. We evaluated the DSS using a CKD dataset and interviews with an experienced nephrologist. In addition, the proposed DSS methodology facilitates the identification and monitoring of the CKD, considering low-income populations in Brazil that usually suffer from the lack/precarious primary care. Nowadays, the remote identification and monitoring of chronic diseases are even more relevant, considering epidemics that prevent face-to-face assistance. For example, in Brazil, the COVID-19 epidemic resulted in negative impacts on the healthcare of low-income populations with chronic diseases and increasing mortality rates. The developed web-based DSS addresses this problem by providing remote and continuous healthcare. The methodology and DSS can be reused in other developing countries with similar scenarios.

Declarations

Acknowledgments

The authors thank the support of the Coordination for the Improvement of Higher Education Personnel (known as CAPES) and the National Council for

Scientific and Technological Development (CNPq).

Data Availability Statement

The dataset used and analysed during the current study is available on the Mendeley Data repository [34].

References

- [1] G. C. K. D. Collaboration, Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the global burden of disease study 2017, *The Lancet* 395 (2020) 709–733.
- [2] Abegunde D and Stanciole A, Preventing chronic diseases: a vital investment: Who global report, 2006.
- [3] World Health Organization, World health statistics overview 2019: monitoring health for the sdgs, sustainable development goals, 2019.
- [4] A. Sobrinho, A. C. M. D. S. Queiroz, L. D. D. Silva, E. D. B. Costa, M. E. Pinheiro, A. Perkusich, Computer-aided diagnosis of chronic kidney disease in developing countries: A comparative analysis of machine learning techniques, *IEEE Access* 8 (2020) 25407–25419.
- [5] A. Levey, L. Inker, J. Coresh, Chronic kidney disease in older people, *Journal of the American Medical Association* 314 (2015) 557–558.
- [6] M. Kinaan, H. Yau, S. Q. Martinez, P. Kar, Concepts in diabetic nephropathy: from pathophysiology to treatment, *Journal of Renal And Hepatic Disorders* 1 (2017) 10–24.
- [7] R. de Castro Cintra Sesso, A. A. Lopes, F. S. Thomé, J. R. Lugon, E. de Almeida Burdmann, Brazilian dialysis census, 2009, *Brazilian Journal of Nephrology* 32 (2010) 380–384.
- [8] A. C. Webster, E. V. Nagler, R. L. Morton, P. Masson, Chronic kidney disease, *The Lancet* 389 (2017) 1238–1252.
- [9] R. C. Sesso, A. A. Lopes, F. S. Thomé, J. R. Lugon, D. R. dos Santos, 2010 report of the brazilian dialysis census, *Brazilian Journal of Nephrology* 33 (2011) 442–447.
- [10] R. C. Sesso, A. A. Lopes, F. S. Thomé, J. R. Lugon, C. T. Martins, Brazilian chronic dialysis survey 2016, *Brazilian Journal of Nephrology* 39 (2017) 380–384.
- [11] F. S. Thomé, R. C. Sesso, A. A. Lopes, J. R. Lugon, C. T. Martins, Brazilian chronic dialysis survey 2017, *Brazilian Journal of Nephrology* (2019) 1–7.
- [12] P. D. M. de Menezes Neves, R. de Castro Cintra Sesso, F. S. Thomé, J. R. Lugon, M. M. Nascimento, Brazilian dialysis census: analysis of data from the 2009-2018 decade, *Brazilian Journal of Nephrology* 42 (2020).
- [13] S. Elshahat, P. Cockwell, A. P. Maxwell, M. Griffin, T. O'Brien, C. O'Neill, The impact of chronic kidney disease on developed countries from a health economics perspective: A systematic scoping review, *Plos One* (2020).
- [14] Brazilian Ministry of Health, <http://portalarquivos.saude.gov.br/images/pdf/2016/abril/04/Valores-gastos.pdf> 2016.
- [15] U. Cha'on, K. Wongtrangan, B. Thinkhamrop, S. Tatiyanupanwong, C. Limwattananon, C. Pongskul, T. Panaput, C. Chalermwat, W. Lert-itthiporn, A. Sharma, S. Anutrakulchai, Ckdnet, a quality improvement project for prevention and reduction of chronic kidney disease in the northeast thailand, *BMC Public Health* 20 (2020).
- [16] D. W. Cockcroft, M. H. Gault, Prediction of creatinine clearance from serum creatinine, *Nephron* 16 (1976) 31–41.
- [17] A. S. Levey, J. P. Bosch, J. B. Lewis, T. Greene, N. Rogers, D. Roth, A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation, *Annals of Internal Medicine* 130 (1999) 461–470.
- [18] K. Frazier, Electronic health records, *The American Journal of Nursing* 117 (2017).
- [19] A. Sobrinho, L. D. da Silva, A. Perkusich, M. E. Pinheiro, P. Cunha, Design and evaluation of a mobile application to assist the self-monitoring of the chronic kidney disease in developing countries, *BMC Medical Informatics and Decision Making* 18 (2018).
- [20] G. Yang, M. J. D. Zhibo Pang, M. Dong, Y.-T. Zhang, N. Lovell, A. M. Rahmani, Homecare robotic systems for healthcare 4.0: Visions and enabling technologies, *IEEE Journal of Biomedical and Health Informatics* (2020).
- [21] A. V. Yan Li, M. Randhawa, G. Fick, Designing utilization-based spatial healthcare accessibility decision support systems: A case of a regional health plan, *Decision Support Systems* 99 (2017) 51–63.
- [22] W.-Y. Hsu, A decision-making mechanism for assessing risk factor significance in cardiovascular diseases, *Decision Support Systems* 115 (2018) 64–77.
- [23] S. Walczak, V. Velanovich, Improving prognosis and reducing decision regret for pancreatic cancer treatment using artificial neural networks, *Decision Support Systems* 106 (2018) 110–118.

- [24] K. Topuz, F. D. Zengul, A. Dag, A. Almehti, M. B. Yildirim, Predicting graft survival among kidney transplant recipients: A bayesian decision support model, *Decision Support Systems* 106 (2018) 97–109.
- [25] B. Wang, H.-W. Wang, H. Guo, Q. T. Erik Anderson, T. Wu, R. Falola, T. Smith, P. M. Andrews, Y. Chen, Optical coherence tomography and computer-aided diagnosis of a murine model of chronic kidney disease, *Journal of Biomedical Optics* 22 (2017) 1–11.
- [26] F. F. Jahantigh, B. Malmir, B. A. Avilaq, A computer-aided diagnostic system for kidney disease, *Kidney Research and Clinical Practice* 36(2017) 29–38.
- [27] J. Neves, M. R. Martins, J. Vilhena, J. Neves, S. Gomes, A. Abelha, J. Machado, H. Vicente, A soft computing approach to kidney diseases evaluation, *Journal of Medical Systems* (2015).
- [28] H. Polat, H. D. Mehr, A. Cetin, Diagnosis of chronic kidney disease based on support vector machine by feature selection methods, *Journal of Medical Systems* 41 (2017).
- [29] P. Arulanthu, E. Perumal, An intelligent iot with cloud centric medical decision support system for chronic kidney disease prediction, *International Journal of Imaging Systems and Technology* 30 (2020) 815–827.
- [30] R. H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F. M. Behlen, P. V. Biron, A. S. Shvo, HI7 clinical document architecture, release 2, *Journal of the American Medical Informatics Association* 13 (2006)31–39.
- [31] E. J. Lamb, A. S. Levey, P. E. Stevens, The kidney disease improving global outcomes (kdigo) guideline update for chronic kidney disease: Evolution not revolution, *Clin. Chem.* 59 (2013) 462–465.
- [32] A. Forbes, H. Gallagher, Chronic kidney disease in adults: Assessment and management, *Clinical Medicine* 20 (2020) 128–132.
- [33] L. A. Inker, B. C. Astor, C. H. Fox, T. Isakova, J. P. Lash, C. A. Peralta, M. K. Tamura, H. I. Feldman, Kdoqi us commentary on the 2012 kdigo clinical practice guideline for the evaluation and management of ckd, *Amer. J. Kidney Diseases* 63 (2014) 713–735.
- [34] A. Sobrinho, L. Dias da Silva, A. Perkusich, A. Queiroz, M. Eliete Pinheiro, A brazilian dataset for screening the risk of the chronic kidney disease, 2020.<http://dx.doi.org/10.17632/2gkg7vvcrm.1>.
- [35] U. V. Heesch, A. Jansen, H. P. Breivold, P. Avgeriou, C. Manteuffel ,Platform design space exploration using architecture decision view-points - a longitudinal study, *Journal of Systems and Software* 124(2017) 56–81.

Figures

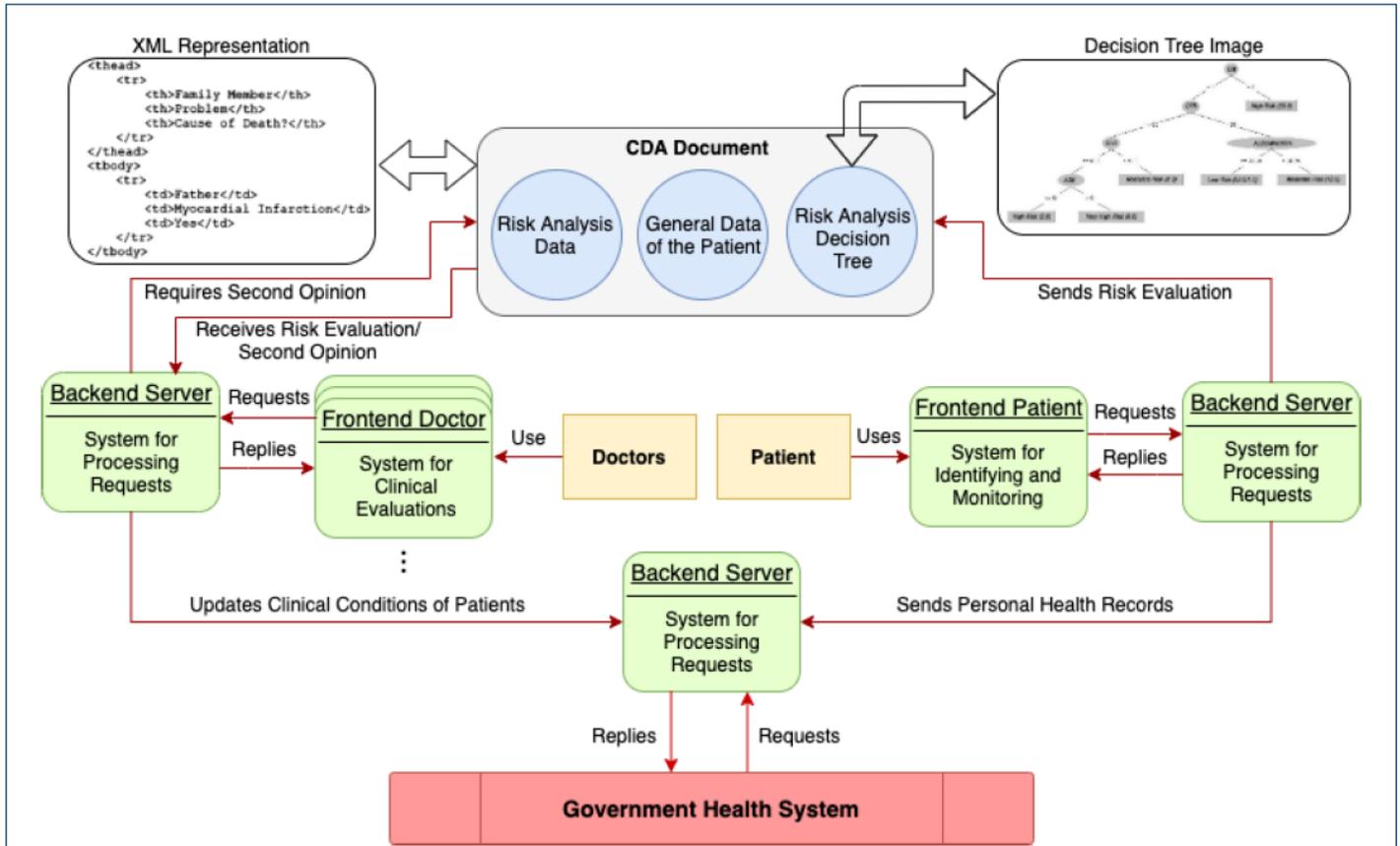


Figure 1 The DSS methodology schema of identification and monitoring of chronic diseases in developing countries.

Figure 1

The DSS methodology schema of identification and monitoring of chronic diseases in developing countries.

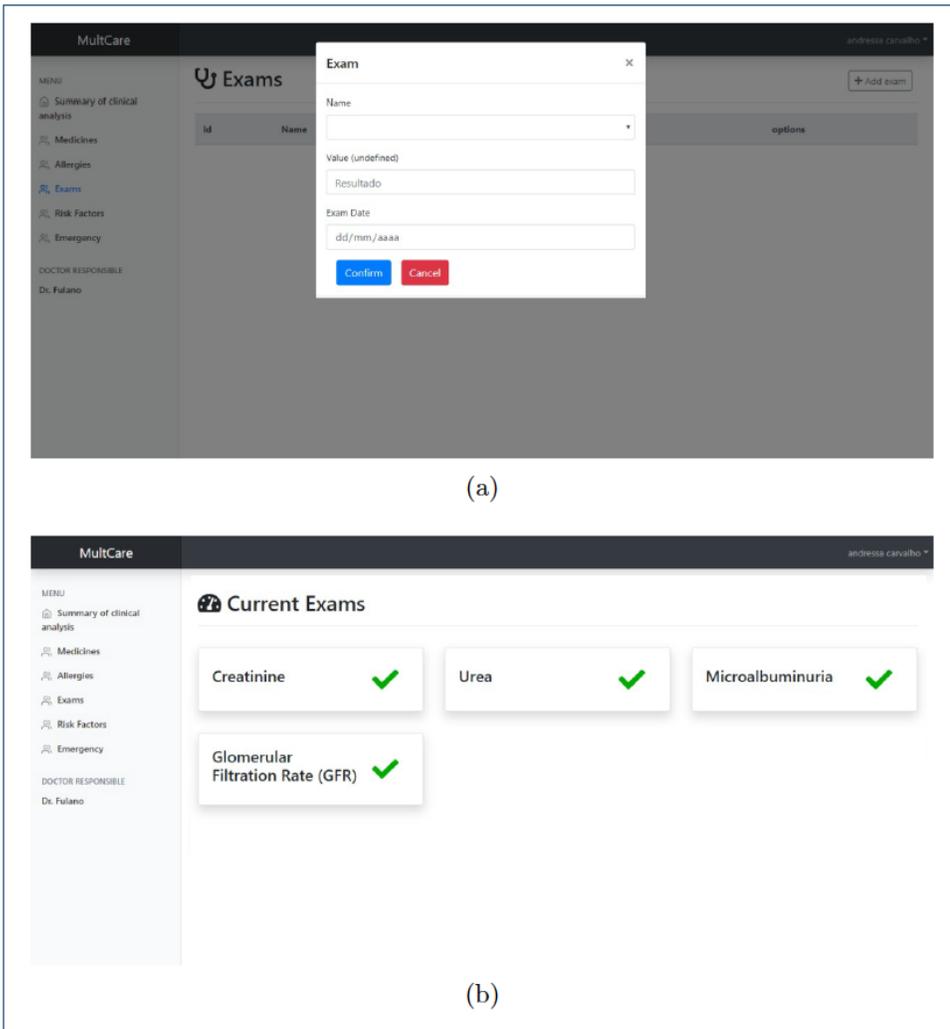


Figure 2

(a) Screenshot of the patient subsystem GUI for recording a new CKD test result. (b) Screenshot of the main GUI for the patient subsystem.

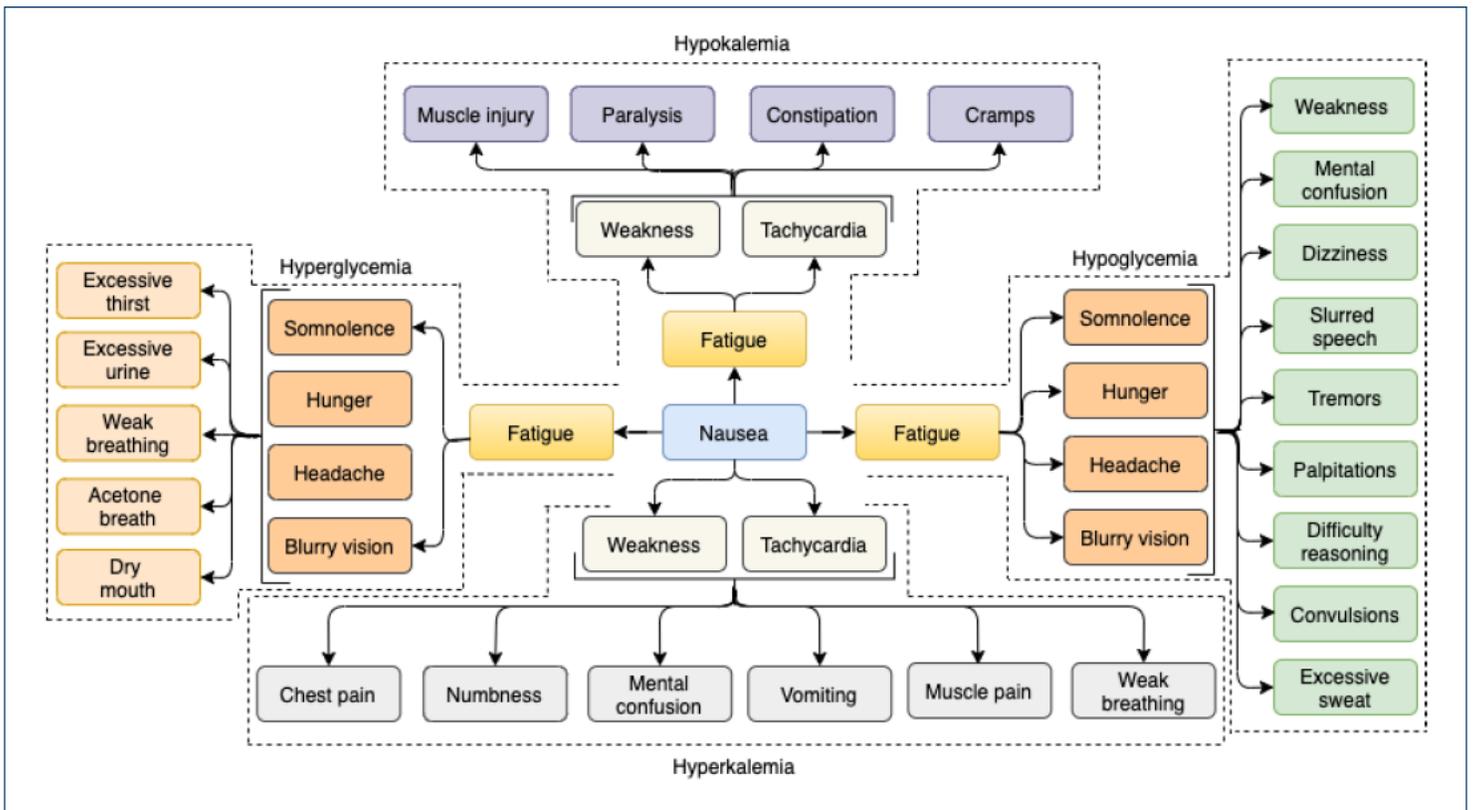


Figure 3

Summary of first decisions to identify the risk of hyperglycemia, hypoglycemia, hyperkalemia, or hypokalemia.

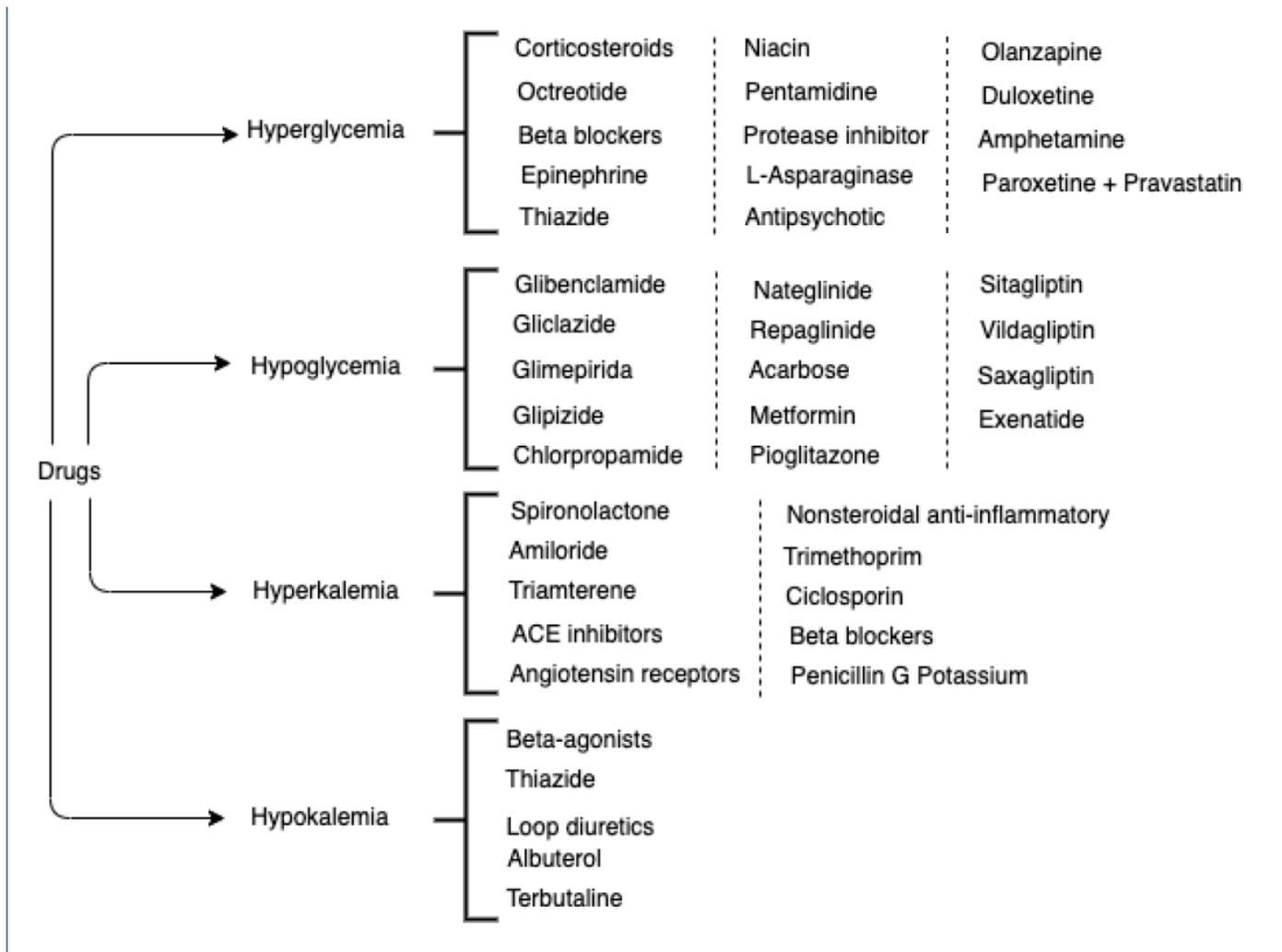


Figure 4
 Drugs that may cause hyperglycemia, hypoglycemia, hyperkalemia, and hypokalemia, collected by analyzing medical guidelines and interviews with a nephrologist.

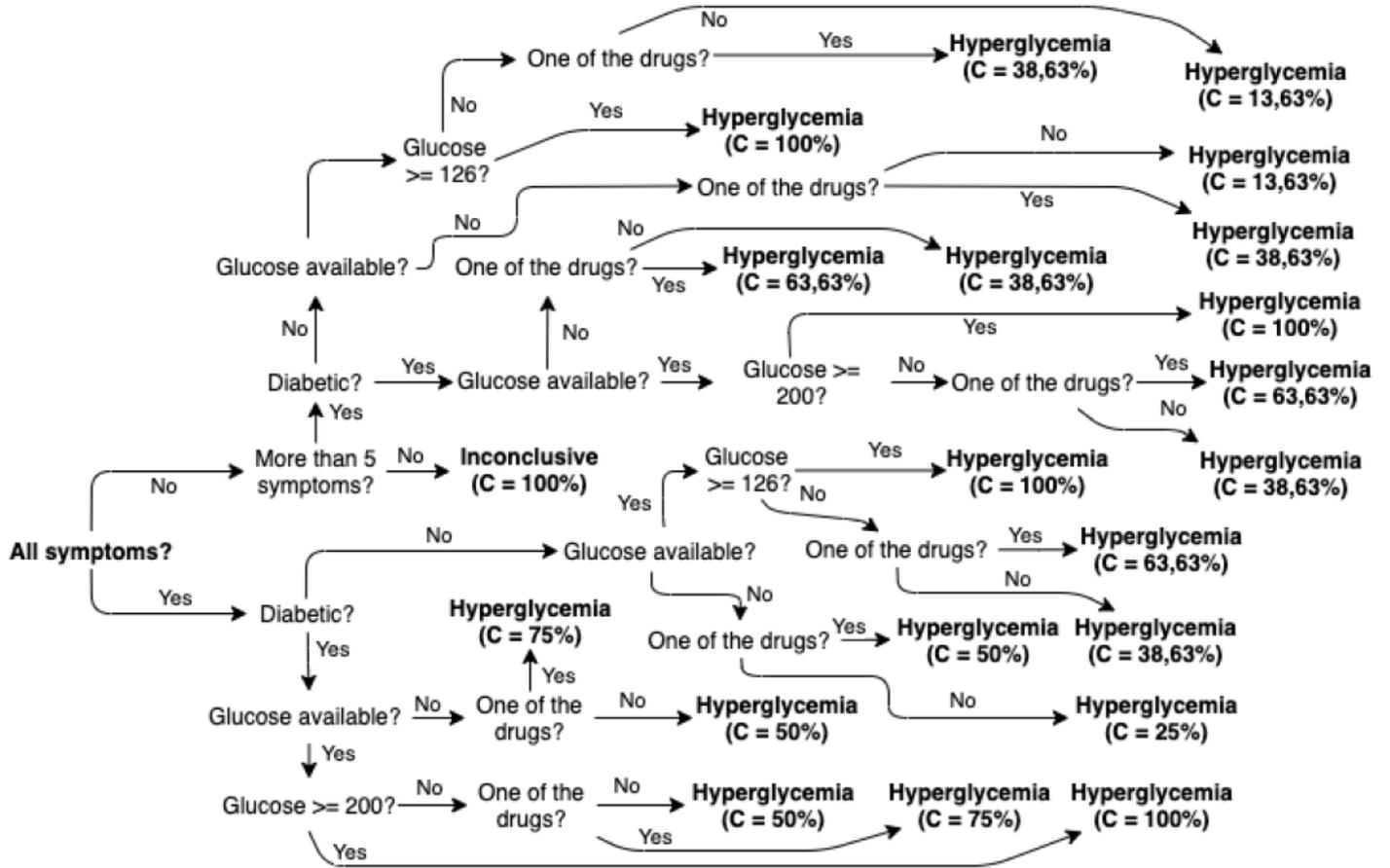
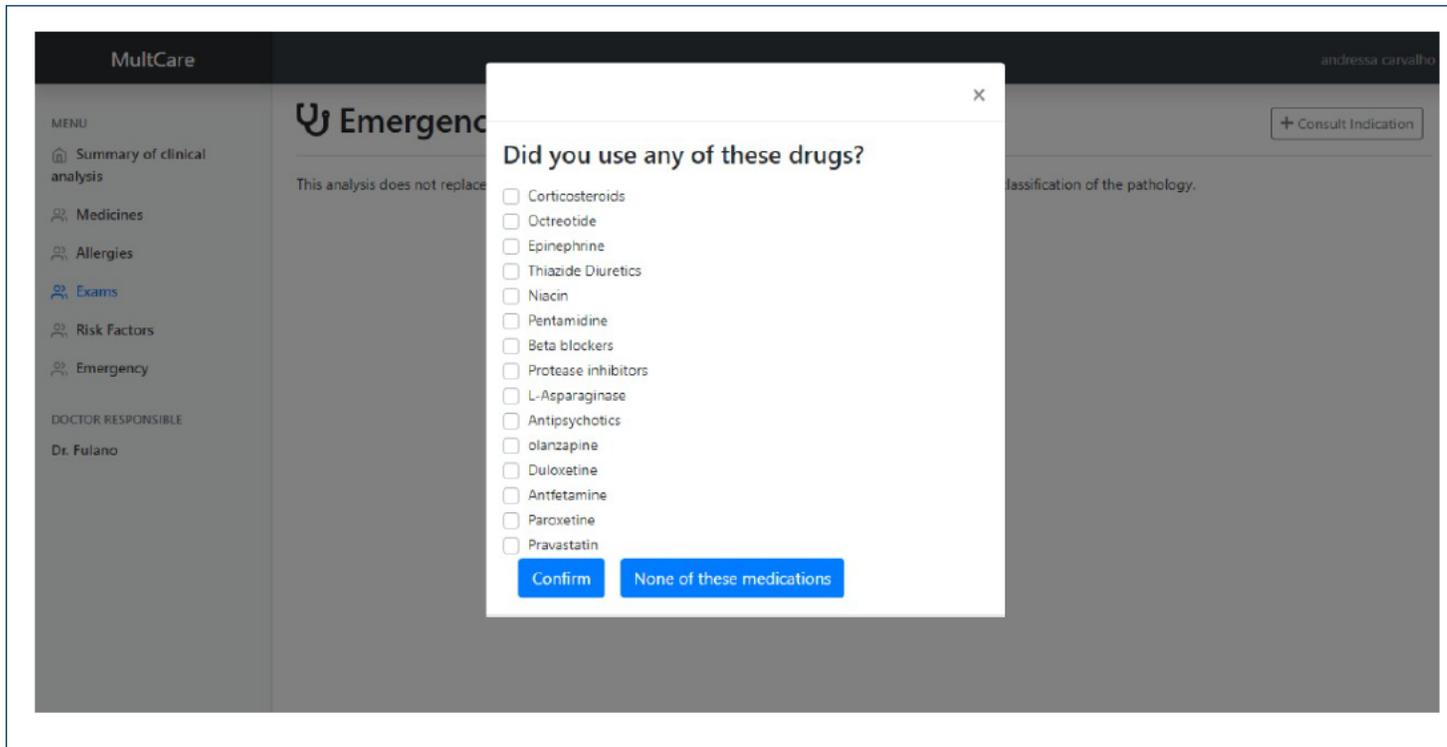


Figure 5

Relationships between decisions used to identify the risk of hyperglycemia.



Sample of GUI of the knowledge-based system, as part of the Patient subsystem.

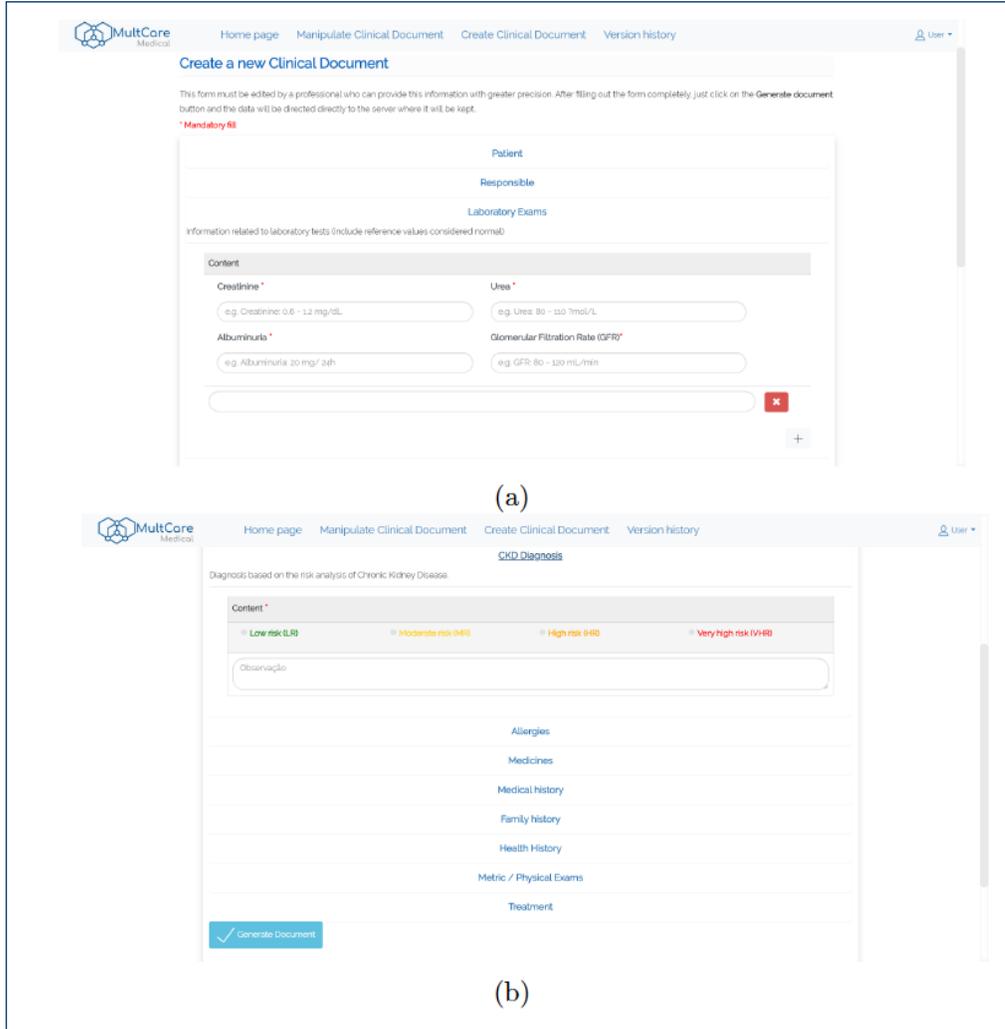


Figure 7

(a) GUI for inserting the test results to create a clinical document. (b) GUI for providing the risk evaluation to create a clinical document.

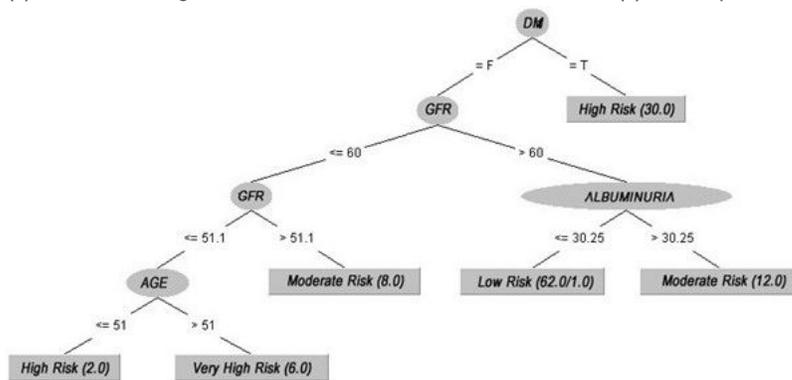


Figure 8

Decision tree for CKD risk evaluation made by J48 algorithm.