Prediction of metabolic syndrome and its associated risk factors in patients with chronic kidney disease using machine learning techniques

Predição de síndrome metabólica e seus fatores de risco associados em pacientes com doença renal crônica utilizando técnicas de *machine learning*

Authors

Jalila Andréa Sampaio Bittencourt[1]
Carlos Magno Sousa Junior[2]
Ewaldo Eder Carvalho Santana[2]
Yuri Armin Crispim de Moraes[1]
Erika Cristina Ribeiro de Lima Carneiro[3]
Ariadna Jansen Campos Fontes[4]
Lucas Almeida das Chagas[5]
Naruna Aritana Costa Melo[6]
Cindy Lima Pereira[1]
Margareth Costa Penha[7]
Nilviane Pires[1]
Edward Araujo Júnior[5]
Allan Kardec Duailibe Barros Filho[1]
Maria do Desterro Soares
Brandão Nascimento[8]

[1]Universidade Federal do Maranhão, Departamento de Engenharia Eletrônica, Laboratório de Processamento da Informação Biológica, São Luiz, MA, Brazil.
[2]Universidade Federal do Maranhão, Departamento de Ciência da Computação, Laboratório de Aquisição e Processamento de Sinais, São Luiz, MA, Brazil.
[3]Universidade Federal do Maranhão, Hospital Universitário, Centro de Prevenção de Doenças Renais, São Luiz, MA, Brazil.
[4]Universidade Federal do Maranhão, Centro de Ciências Biológicas e da Saúde, Laboratório de Imunofisiologia, São Luiz, MA, Brazil.
[5]Universidade Federal de São Paulo, Escola Paulista de Medicina, Departamento de Obstetrícia, São Paulo, SP, Brazil.
[6]Universidade Federal do Maranhão, Laboratório de Ciências Biológicas, Laboratório de Genética e Biologia Molecular, São Luiz, MA, Brazil.
[7]Universidade Ceuma, Departamento de Biomedicina, Laboratório de Ciências Biomédicas, São Luiz, MA, Brazil.
[8]Universidade Federal do Maranhão, Departamento de Medicina, Programa de Pós-Graduação em Saúde do Adulto, São Luiz, MA, Brazil.

ABSTRACT

Introduction: Chronic kidney disease (CKD) and metabolic syndrome (MS) are recognized as public health problems which are related to overweight and cardiometabolic factors. The aim of this study was to develop a model to predict MS in people with CKD. **Methods:** This was a prospective cross-sectional study of patients from a reference center in São Luís, MA, Brazil. The sample included adult volunteers classified according to the presence of mild or severe CKD. For MS tracking, the k-nearest neighbors (KNN) classifier algorithm was used with the following inputs: gender, smoking, neck circumference, and waist-to-hip ratio. Results were considered significant at p < 0.05. **Results:** A total of 196 adult patients were evaluated with a mean age of 44.73 years, 71.9% female, 69.4% overweight, and 12.24% with CKD. Of the latter, 45.8% had MS, the majority had up to 3 altered metabolic components, and the group with CKD showed statistical significance in: waist circumference, systolic blood pressure, diastolic blood pressure, and fasting blood glucose. The KNN algorithm proved to be a good predictor for MS screening with 79% accuracy and sensitivity and 80% specificity (area under the ROC curve – AUC = 0.79). **Conclusion:** The KNN algorithm can be used as a low-cost screening method to evaluate the presence of MS in people with CKD.

Keywords: Artificial Intelligence; Renal Insufficiency, Chronic; Machine Learning; Metabolic syndrome.

RESUMO

Introdução: A doença renal crônica (DRC) e a síndrome metabólica (SM) são reconhecidas como problemas de saúde pública relacionados ao excesso de peso e a fatores cardiometabólicos. O objetivo deste estudo foi desenvolver um modelo para prever a SM em pessoas com DRC. **Métodos:** Este foi um estudo transversal prospectivo de pacientes de um centro de referência em São Luís, MA, Brasil. A amostra incluiu voluntários adultos classificados de acordo com a presença de DRC leve ou grave. Para o rastreamento da SM, o algoritmo de classificação *k-nearest neighbors* (KNN) foi utilizado com os seguintes dados: sexo, tabagismo, circunferência do pescoço e relação cintura-quadril. Os resultados foram considerados significativos com p < 0,05. **Resultados:** Foram avaliados 196 pacientes adultos com média de idade de 44,73 anos, 71,9% do sexo feminino, 69,4% com sobrepeso e 12,24% com DRC. Desses últimos, 45,8% apresentaram SM, a maioria tinha até 3 componentes metabólicos alterados, e o grupo com DRC apresentou significância estatística em: circunferência da cintura, pressão arterial sistólica, pressão arterial diastólica e glicemia de jejum. O algoritmo KNN comprovou ser um bom preditor para a triagem de SM com acurácia e sensibilidade de 79% e especificidade de 80% (área sob a curva ROC – AUC = 0,79). **Conclusão:** O algoritmo KNN pode ser usado como um método de triagem de baixo custo para avaliar a presença de SM em pessoas com DRC.

Descritores: Inteligência Artificial; Insuficiência Renal Crônica; Aprendizado de Máquina; Síndrome Metabólica.

INTRODUCTION

Chronic non-communicable diseases (CNCD) are currently recognized as one of the major public health problems[1]. The World Health Organization (WHO) estimates that CNCDs are responsible for 71% of the 57 million deaths worldwide[2]. In Brazil, CNCDs are responsible for 76.4% of all deaths, with a focus on diseases of the circulatory system (28% of deaths), cancer (18%), diabetes mellitus (5%), and respiratory diseases (6%)[3].

Among the CNCD, chronic kidney disease (CKD), which is characterized by altered renal function, stands out. It is defined as an abnormality in renal structure or function that has been present for more than three months and has health implications. These abnormalities can be represented by a decreased glomerular filtration rate (GFR) <60 ml/min/1.73 m or the presence of one or more markers of kidney injury[4,5].

The prevalence of CKD is still unknown in many countries[6]. However, it has been increasing, mainly as a result of the increasing incidence of obesity, diabetes, and hypertension. In addition, renal function is highly susceptible to age-related changes, with a significantly higher incidence in middle-aged and elderly patients[7,8].

People with CKD tend to have cardiovascular disease (CVD)[8]. CVD is the leading cause of death in patients with chronic kidney disease and is associated with accelerated progression of CKD. These findings support the view that the presence of cardiometabolic risk factors (CRF) and impaired kidney function may increase kidney disease-related risks[9].

In addition, metabolic syndrome (MS) is considered to be a grouping of interrelated risk factors that doubles the risk of CVD in 5 to 10 years[10]. This pathology is described as a set of the CRF, which are usually related to the development of insulin resistance and fat accumulation. These risk factors include arterial hypertension, hypertriglyceridemia, dyslipidemia, hyperglycemia and central obesity[11].

Individuals with NCD generate great financial costs to the public health system as they require treatment to control these diseases, especially in MS and CKD. Patients with end-stage CKD, stage G5 (GFR <15 mL/min/1.73 m²) have severe renal failure leading to complete loss of renal function. At this stage, the therapeutic options are renal replacement therapies (RRT), such as artificial blood purification methods (peritoneal dialysis or hemodialysis) or kidney transplantation[4]. In Brazil, RRT is considered the main treatment and also requires higher costs for health services[12]. Therefore, early detection of such pathologies can delay complications and support the use of appropriate interventions, such as screening tests in high-risk groups[13].

In this case, some data analysis techniques appear to be good solutions that provide more accurate predictions about the individual's health[14]. Therefore, the use of machine learning (ML) techniques appears to be an instrument to help develop and improve new methods for diagnosis and/or screening[15,16].

Therefore, the importance of the study for the prediction of MS in the population with CKD is clear, since this condition leads to more advanced stages and a higher risk of death from cardiovascular events in this population. Therefore, even though patients with CKD have few risk factors for MS, preventive measures must be taken to avoid problems and negative outcomes such as early death.

Thus, given the magnitude of MS and CKD and the complications related to them, efforts should therefore be made to enable studies aimed at early diagnosis of these pathologies. In addition, this study aims to develop a model to predict the risk of MS and associated risk factors in people with CKD.

METHODS

This was a prospective cross-sectional study of patients of both sexes from the Nephrology Reference Center in São Luís, MA, Brazil, between January 2018 and July 2020. The sample consisted of 196 volunteers classified according to their health status. CKD was determined by a glomerular filtration rate <60 mL/min[4], for mild CKD (above this value) or severe CKD. GFR was determined by measuring serum creatinine To calculate the estimated glomerular filtration rate, the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation was used[17].

The suspected diagnosis of MS was defined as proposed by the International Diabetes Federation (IDF)[18] by considering the presence of changes in waist circumference (≥ 90 cm for men and ≥ 80 cm for women), as a mandatory factor in addition to two other altered components. These components may be: triglycerides ≥ 150 mg/dL or treatment

2

Braz. J. Nephrol. (J. Bras. Nefrol.) 2024, 46(4):e20230135

for dyslipidemia; HDL cholesterol < 40 mg/dL for men and < 50 mg/dL for women or treatment for dyslipidemia; systolic blood pressure (SBP) ≥ 130 mmHg, diastolic blood pressure (DBP) ≥ 85 mmHg, or use of antihypertensive medication; and fasting blood glucose ≥ 100 mg/dL or previous diagnosis of diabetes mellitus[19].

Data collection was carried out in two stages. First, volunteers were recruited at the Nephrology Center. Those who agreed to participate in the study voluntarily signed the Informed Consent Form (ICF). Soon after, they were interviewed using a semi-structured questionnaire, followed by an anthropometric and hemodynamic assessment. The laboratory tests were then scheduled for the following day.

Anthropometric, biochemical, hemodynamic, and lifestyle data were evaluated. The semi-structured questionnaire took into account sociodemographic characteristics, lifestyle, and self-reported personal history, such as hypertension and diabetes mellitus. Anthropometric variables were performed in duplicate, and the means were used for data analysis. All variables were measured according to protocols already consolidated in the literature[20]. These variables were: weight, height, arm circumference (AC), waist circumference (WC), hip circumference (HC), neck circumference (NC), and calf circumference (CC).

Anthropometric indices: waist-to-hip ratio (WHR), waist-to-height ratio (WHtR), and body mass index (BMI) were used to determine nutritional status and were based on WHO cutoff points. WC was estimated considering the cutoff points for the South American population, with values of ≥90 cm for men and ≥80 cm for women[19].

Only the k-nearest neighbors (KNN) classifier algorithm, a supervised ML method, was used for MS tracking. It is easy to implement, adaptable, and easy to program. All these advantages had a positive impact on our choice, as KNN alone met our demands[21].

KNN uses the closest data and performs a segmentation of the closest results based on the selected metric by considering a limited margin of error. In this algorithm, the dataset is prepared by removing missing values and normalizing features, which is known as the pre-processing phase. The data is randomly divided into two different sets: the training set and the test set. This technique ensures an adequate representation of the patterns for training and a robust performance evaluation[21].

Therefore, the database was divided into two groups: 80% of the sample was for training and 20% for testing. Based on the patient's clinical information (gender, smoking, WHR, and CC), the classifier algorithm was able to distinguish individuals with and without MS. The classification method was constructed using MATLAB® software version R2021a (MathWorks Inc, Natick, MD, USA).

For the data file and the statistical analysis, the SPSS® software version 25 (SPSS Inc., Chicago, IL, USA) was used. The Kolmogorov-Smirnov test was used to analyze the normality of the data. The variables that were considered normally distributed were analyzed using the Student's t-test. The others were analyzed with the Mann-Whitney U test, and the Receiver Operating Characteristics (ROC) curve was used for the evaluation of the algorithm classifier. In addition, the results were considered statistically significant at $p < 0.05$. All statistical analysis values are described in the tables in the Results section.

This study is part of a larger project ("umbrella" or "root"), entitled "Prediction of Chronic Kidney Disease Using Artificial Neural Networks". It was also approved by the Ethics and Research Committee of the Federal University of Maranhão, according to the CAAE opinion number 67030517.5.0000.5087. In addition, all participants signed the informed consent form. It is worth mentioning that this work uses original population data generated exclusively for this research, in addition to providing new data for the umbrella project. Furthermore, it is worth mentioning that both neural networks and the KNN are artificial intelligence computational models that work with data processing for classification and prediction.

RESULTS

A total of 196 adult patients with a mean age of 44.73 ± 15.96 years were evaluated, of whom 71.9% (n = 141) were female and 79.1% (n = 155) identified themselves as non-white. In addition, 65.8% (n = 129) of patients reported no physical activity, 29.1% (n = 57) consumed alcohol, and 3.6% (n = 7) smoked. Regarding underlying diseases, 31.1% (n = 61) reported having systemic arterial hypertension (SAH) and 29.6% (n = 58) had diabetes mellitus (Table 1). The analysis of anthropometric data revealed that

more than half of the individuals were overweight (69.4%), with a mean BMI of 27.44 ± 5.10 kg/m^2 (Tables 1 and 2).

Table 3 shows the general characteristics of the sample, which was stratified according to the presence of mild and severe CKD. In this Table, the severe CKD patients had significantly higher values of systolic and diastolic blood pressure, age, CP, fasting blood glucose, and urea (p < 0.05) than those with mild CKD. Similarly, the obesity indicator variables (WC, BMI, WHR, WHtR) also showed a higher prevalence in the severe CKD group.

For clinical, epidemiological, didactic, and conceptual purposes, CKD is classified into six functional stages according to the patient's degree of renal function based on glomerular filtration rate, ranging from normal/high condition to dialysis or transplantation. The Kidney Disease: Improving Global Outcomes (KDIGO) guides the estimation of GFR from serum creatinine as the best method to diagnose, classify and monitor progression of CKD[4]. GFR was categorized from G1 to G5 as shown in Table 4.

Regarding the staging and classification of CKD, Table 4 shows that a large part of the sample is in the initial stages (G1 and G2), all from the mild CKD group. In contrast, only 1.4% (n = 4) is in the end-stage and requires some type of renal replacement therapy.

Table 5 shows the prevalence values of risk factors for metabolic syndrome stratified by the presence of mild or severe CKD. The analysis of clinical data showed that, in percentage values, all altered components of MS were found to be prevalent in the severe CKD group, with the exception of total cholesterol (TC). Of the investigated sample, 45.8% of patients with severe CKD had MS, and the majority had up to 3 altered metabolic components.

When we look at the table stratified by the presence of CKD (Table 3), the group of individuals with kidney problems had higher mean/median values in all parameterscompared to healthy individuals, with statistical significance in WC, systolic blood pressure, diastolic blood pressure and fasting blood glucose.

Regarding the software developed through ML, the implemented algorithm was the KNN with the following entries: gender, smoking, WHR, and

TABLE 1	SOCIODEMOGRAPHIC AND LIFESTYLE CHARACTERISTICS OF THE SAMPLE		
Variables	Category	N	%
Skin color*	White	41	20.9%
	Not white	155	79.1%
Gender*	Women	141	71.9%
	Man	55	28.1%
Nutritional status (BMI)*	Eutrophic	60	30.6%
	Overweight	136	69.4%
Smoking*	No	189	96.4%
	Yes	7	3.6%
Alcoholism*	No	139	70.9%
	Yes	57	29.1%
Physical activity practice*	No	129	65.8%
	Yes	67	34.2%
Presence of CKD*	No	172	87.8%
	Yes	24	12.2%
Presence of MS*	No	150	76.5%
	Yes	46	23.5%
DM self-reported*	No	138	70.4%
	Yes	58	29.6%
Self-reported SAH*	No	135	68.9%
	Yes	61	31.1%

Abbreviations – BMI: body mass index; CKD: chronic kidney disease; MS: metabolic syndrome; DM: diabetes mellitus; SAH: systemic arterial hypertension. Notes – *values are described as percentage (%) and frequency (n).

TABLE 2	ANTHROPOMETRIC AND HEMODYNAMIC CHARACTERISTICS OF THE SAMPLE
Variables	Averages
Age (years)[†]	44.73 ± 15.96
Height (m)[†]	1.58 ± 0.08
Body weight (kg)[†]	68.63 ± 13.79
WC (cm)[†]	87.22 ± 12.95
AC (cm)[†]	30.46 ± 4.08
NC (cm)[†]	35.16 ± 3.53
CC (cm)[†]	35.66 ± 3.63
BMI (kg/m^2)[†]	27.44 ± 5.10
WHtR[†]	0.55 ± 0.08
WHR[†]	0.85 ± 0.08
SBP[†]	124.06 ± 23.99
DBP[†]	77.37 ± 12.06

Abbreviations – WC: waist circumference; AC: Arm circumference; NC: neck circumference; CC: calf circumference; BMI: body mass index; WHtR: waist to height ratio; WHR: waist: hip ratio; SBP: systolic blood pressure; DBP: diastolic blood pressure; m: meter; cm: centimeter. Notes – [†]Mann-Whitney U test; Data are described as mean ± (standard deviation).

TABLE 3	ANTHROPOMETRIC, HEMODYNAMIC, AND LABORATORY CHARACTERISTICS OF THE SAMPLE, STRATIFIED BY THE PRESENCE OF CHRONIC KIDNEY DISEASE (CKD)		
Variables	Mild CKD	Severe CKD	p-value
	(n = 172)	(n = 24)	
Age (years)[†]	40.5 (29.25 – 54)	60 (51.25 – 68)	<0.001
Height (m)[†]	1.57 (1.51 – 1.63)	1.61 (1.55 – 1.68)	0.153
Body weight (kg)[†]	66.2 (58.8 – 76.57)	70.45 (64.17 – 80.45)	0.090
WC (cm)[†]	86.16 ± 12.79	94.85 ± 11.7	0.002
AC (cm)[†]	30 (28 – 33.3)	30 (28.12 – 31.87)	0.816
NC (cm)[†]	34.9 (32 – 38)	36.75 (34 – 39.75)	0.039
CC (cm)[†]	35.5 (33 – 38)	34.25 (32 – 36.75)	0.196
BMI (kg/m^2)[†]	26.54 (23.71 – 29.91)	27.67 (25.58 – 31.42)	0.285
WHR[†b]	0.84 ± 0.08	0.93 ± 0.08	<0.001
WHtR[†]	0.54 ± 0.08	0.59 ± 0.07	0.012
SBP[†]	117.5 (107 – 131)	134 (123.25 – 165.5)	<0.001
DBP[†]	75 (70 – 82)	86.5 (76.25 – 91)	0.019
TC[†]	178.12 ± 45.39	163.13 ± 44.78	0.131
Triglycerides[†]	108 (75.25 – 164.75)	128 (84.25 – 174.5)	0.216
HDL-cholesterol[†]	44 (38 – 54)	43 (34 – 53.75)	0.575
LDL-cholesterol[†]	105.73 ± 36.41	89.80 ± 41.89	0.050
Blood glucose[†]	79.00 (77.50 – 87.50)	81.00 (75.00 – 88.00)	0.001
Urea[†]	24.5 (20.25 – 31)	54.5 (39.25 – 79)	<0.001

Abbreviations – BMI: body mass index; WC: waist circumference; AC: Arm circumference; CC: calf circumference; WHR: waist:hip ratio; WHtR: waist to height ratio; SBP: systolic blood pressure; DBP: diastolic blood pressure; TC: total cholesterol; GFR: glomerular filtration rate; m: meter; cm: centimeter. Notes – [†]Mann-Whitney U test; Values are described as mean/median ± SD (standard deviation).

TABLE 4	STAGING AND CLASSIFICATION OF CHRONIC KIDNEY DISEASE (CKD) IN THE SAMPLE (N = 196)			
Stages	Classification		Frequency[*]	
G1	TFG >90	(normal or increased)	117	41.1%
G2	TFG 60–89	(slightly diminished)	55	19.3%
G3a	TFG 45–59	(mild to moderate decrease)	10	3.5%
G3b	TFG 30–44	(moderate to severe decrease)	5	1.8%
G4	TFG 15–29	(severely diminished)	5	1.8%
G5	TFG <15	(renal insufficiency)	4	1.4%

Abbreviation – GFR: glomerular filtration rate. Notes – [*]values are described in frequency (n) and percentage (%).

NC. The development of the algorithm was labeled based on the MS components according to the IDF criteria[18,19].

There is no fixed ratio that works in all scenarios in the KNN algorithm. The database was divided into two groups: 80% of the sample was allocated for training and 20% for testing. This was done taking into account the fact that, with very large datasets, it may be feasible to use ratios such as 70/30 or even 90/10[22]. All these divisions were tested, and the one that showed the best response was 80/20, as it is a medium sized database. The KNN had 79% accuracy and sensitivity and a specificity of 80%.

Figure 1 shows the graphical representation of the ROC curve (area under the ROC curve – AUC = 0.79) generated by plotting the sensitivity (true positive rate) on the y-axis against the specificity (false positive rate) on the x-axis. Thus, the KNN proved to be a good classifier for predicting MS. For a diagnostic test to be considered accurate, a curve in the upper left

TABLE 5   PREVALENCE OF METABOLIC SYNDROME AND ITS COMPONENTS IN THE SAMPLE

Variables	Category	Mild CKD		Severe CKD		p-value
		N	%	N	%	
MS*	Not	137	79.70%	13	54.20%	0.006
	Yes	35	20.30%	11	45.80%	
WC*	Normal	68	39.50%	6	25.00%	0.169
	Changed	104	60.50%	18	75,00%	
BP*	Normal	113	65.70%	7	29.20%	0.001
	Changed	59	34.30%	17	70.80%	
TC*	Normal	108	62.80%	17	70,80%	0.443
	Changed	64	37.20%	7	29.20%	
TG*	Normal	121	70.30%	15	62.50%	0.434
	Changed	51	29.70%	9	37.50%	
HDL*	Not	75	43.60%	10	41.70%	0.858
	Yes	97	56.40%	14	58.30%	
FG*	Normal	149	86.60%	15	62.50%	0.003
	Changed	23	13.40%	9	37.50%	

Abbreviations – WC: waist circumference; BP: blood pressure; high TC: high total cholesterol; TG: triglycerides; HDL: high density lipoprotein; GJ: fasting blood glucose. Note – *Chi-square test.
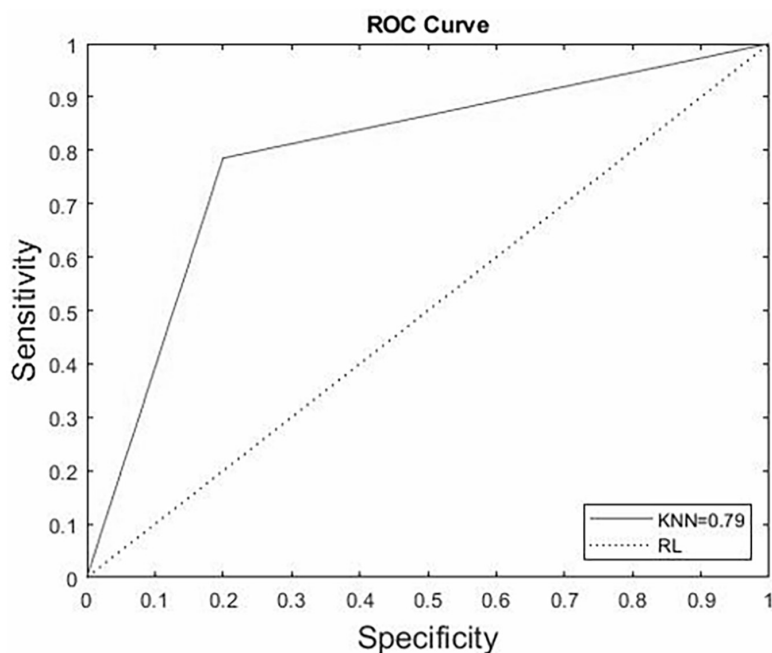

Figure 1. Area under the ROC curve is demonstrating the discriminatory power of the k-nearest neighbors' algorithm in predicting metabolic syndrome in the test set.

triangle must be above the reference line. The closer this curve is to this corner, i.e. closer to 1, the better the method's performance[23,24].

In this first stage of the study, we decided to use these anthropometric input parameters as they are inexpensive, easy to use and already recommended in the literature. However, it is believed that using more input variables can increase the specificity of the classifier algorithm.

The software is simple and easy to use. It has four fields for entering the patient's clinical data and three buttons. These buttons are labeled as
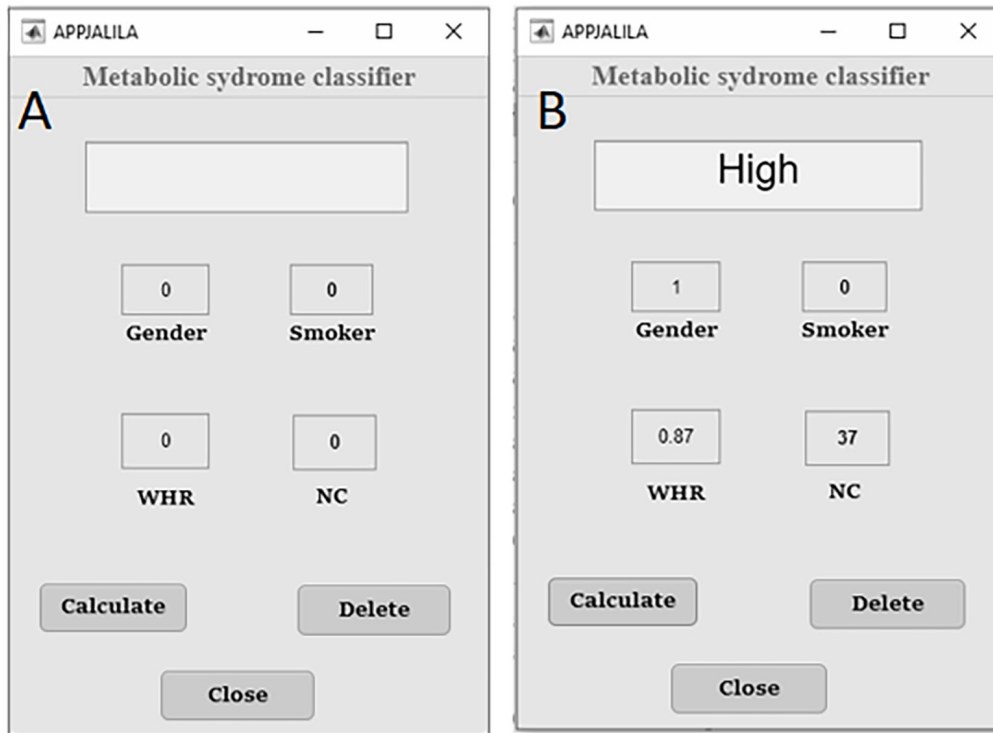
Figure 2. User interface layout (A) and example with input parameters (B).

follows: calculate (provides the patient classification after analyzing the clinical data), clear (deletes the entered data on the screen), and close (terminates the operation of the software). The conclusion from the analysis of the algorithm is presented in the "Result" field and described as follows: high or normal for predicting MS risk in CKD patients, as shown in Figure 2.

DISCUSSION

Corroborating our results, several studies reported a higher incidence of MS in women with severe CKD and a higher prevalence in the elderly[19–25]. Age is considered a risk factor for both MS and CKD. In this study, the high prevalence of MS in elderly people can be described by functional limitations, an increasingly sedentary lifestyle, and reduced physical activity, as described in other reports[26].

The results of our study reveal that abdominal obesity, altered blood pressure, and high blood glucose were associated with CKD. In addition, the main underlying diseases in the study population were SAH and DM. These findings confirm the results of studies in which central obesity was associated with CKD, regardless of general obesity and increased BMI[25,27,28].

In MS, abdominal obesity is one of the main components responsible for insulin resistance, which in turn can lead to progressive loss of renal function[29]. This occurs because obesity directly affects hemodynamics and renal structure, as substantial evidence shows[30].

Among CNCD, SAH and DM were reported as the main underlying diseases in our study. These pathologies are among the most prevalent risk factors for the development of CKD and are responsible for most cases. In Brazil, according to the Census of Hemodialysis Centers, hypertensive nephropathy (34%) and diabetes (31%) are the main underlying diseases in patients undergoing hemodialysis in 2019[31].

In this study, MS (according to the IDF definition) is associated with an increased risk of CKD[18]. These findings call for greater attention to policies and interventions, such as lifestyle changes, that should aim to reduce the prevalence of MS and its adverse outcomes. The literature suggests that efforts to raise awareness of prevention strategies should start early, when any of the constituent components of MS is present[19].

Thus, public health strategies are important for the prevention of CNCD in general. To this end, the

Strategic Action Plan for Confronting CNCDs 2011–2022 was launched to promote the development of public policies. Itaimed at the prevention and control of CNCDs and their grievances in order to reduce the premature mortality rates (30 to 69 years) by 2% per year and reduce the prevalence of their risk factors[32].

In this study, MS had a higher prevalence in the severe CKD group. It is already known that MS negatively impacts the progression of CKD. Therefore, MS and its associated factors need to be identified early. However, a disadvantage in MS evaluation is the use of invasive variables that are included in all criteria for MS diagnosis[18,33].

In developing countries, as in Brazil, the population's difficulty in accessing primary health care services, specialized consultations, and complementary exams contribute to the underreporting of CNCD, including MS and CKD. Therefore, the development of complementary, cost-effective and easy-to-use diagnostic methods is needed to facilitate patient access to early diagnosis[14–16].

In this sense, the use of additive manufacturing techniques appears to be an instrument to assist in the development of new low-cost screening methods without the use of invasive variables[34]. Therefore, one of the proposals of this study was the development of a software with these characteristics, such as the KNN algorithm, which can be used in the evaluation of CDK patients with MS.

The KNN method performs binary classification by not only giving the outputs, when there is pathology and when there is no pathology, but also performing the expected classification based on the evaluation of the collected data. Several studies use the KNN method to classify groups for application to biological data, with satisfactory results, indicating that it is an efficient method for the present work[21,34].

With the demand for methods that can facilitate diagnoses and optimize the care of healthcare professionals, several studies have been developed to address the application of ML in healthcare. These include the following: prediction of cardiovascular disease over a 10-year period[35]; predictive models for undiagnosed DM[36]; and Parkinson's disease diagnosis from patient writing, by using image processing techniques[37].

In the present study, based on the evaluation metrics of the chosen classifier and the area under the ROC curve (AUC = 0.79), KNN was found to be a good predictor of MS tracking in clinical practice. In Rosa's study evaluating the prediction of metabolic syndrome in antiretroviral users, the KNN algorithm also proved to be a good predictor by providing AUC = 0.78, a result similar to ours[38].

This result can be attributed to input parameters such as NC and WHR, for example. This statement is contradicted by several findings in the literature, which report the importance of anthropometric variables for clinical practice by associating NC[37,39] and WHR[40,41] as good predictors of MS.

Similarly, smoking is an established indicator in the literature for cardiometabolic risk analysis[42,43]. Smoking remains the leading cause of preventable death worldwide and a crucial factor for the development of CNCDs such as cancer, cardiovascular, and pulmonary diseases. Gender, in turn, has been associated with MS and CKD, with higher prevalence among women[43,44].

The advantages of the KNN are: its simplicity of implementation, its very effective performance in different situations and areas (engineering, health, education, among others), its ease of interpretation, and its ideal suitability for small or medium-sized databases. Furthermore, it directly produces the decision rule without estimating the densities that are conditioned on the classes, making it a good choice for classification problems where closely related patterns in the feature space may belong to the same class[21].

Thus, the KNN algorithm proved to be an attractive tool that can be implemented in environments with few resources. It can also support the clinical management of MS in the general population, especially in the CKD population, given the association between the number of MS components and the progression of CKD. In addition, it can have a positive impact on the quality of life of this population.

However, some limitations were found throughout the course of the study. One of the initial limitations was due to the COVID-19 pandemic, which made data collection difficult since the study population was considered at risk for the coronavirus. The number of variables can be seen as another limiting factor, because it is believed that using more input variables can increase the specificity of the classifier algorithm. Another additional limitation can be the inclusion of competitive risks, where patients may suffer from other types of events that prevent the

8

Braz. J. Nephrol. (J. Bras. Nefrol.) 2024, 46(4):e20230135

observation of the event under study. Furthermore, as this was a cross-sectional study, it was not possible to continue monitoring patients over time, and it was not possible to establish a cause-effect relationship.

This contributed to a limitation of the sample, as the results found should only be considered for the population in question. Thus, as a recommendation for future investigations in this topic, we emphasize the importance of developing the classifier algorithm at this time. For this, a more robust sample can be used by expanding it to other populations with other associated comorbidities, by increasing the number of input variables, and by developing a longitudinal study. Also, a validation study of the implemented software, an incorporation of additional statistical tests, and a comparison of different algorithms can be provided, with or without the same proposed methods.

Conclusion

In summary, there was a significant prevalence of MS in the CKD population, demonstrating the importance of early detection of this syndrome. This occurs because of the fact that MS involves a series of cardiometabolic factors, and the more components present, the greater the risk of CKD progression. The KNN classifier algorithm can be used as a screening method with high sensitivity and low cost. In addition, it can be used to screen for MS in primary healthcare units and low-resource settings by contributing to the early detection of MS in the general population, especially in CKD. The use of the classifier can help health professionals in decision making by contributing to preventive healthcare, especially in CKD, to reduce treatment costs and even avoid negative outcomes and early death.

Considering its important usefulness for the public health field, another possible area of future research could be to carry out larger studies with the KNN classifier and its variants. Also, other low-cost parameters could be used to predict the risk of harmful clinical conditions in groups with and no risk of chronic non-communicable diseases.

Acknowledgments

Authors' Contributions

JASB carried out the writing of this study and was a collaborator in the development of the software and data collection. CMSJ collaborated with the deployment of prediction software for MS. EECS was a collaborator in the development of the software. YACM assisted in the correction, spelling, and translation of the text. ECRLC was a mentor in the software's data research. AJCF collaborated in data collection. LAC collaborated in data collection. NACM collaborated in data collection. CLP collaborated in data collection and assisted in carrying out the statistical analysis. MCP collaborated in data collection. EAJ carried out the writing of the final version and submission. AKBDF was the research coordinator and collaborator in the development of the program. NP supervised the software's data research and assisted in the writing of the text. MDSBN was the program's data research advisor and assisted in the writing of the text. All authors read and approved the final manuscript.

Conflict of Interest

The authors declare that they have no conflict of interest related to the publication of this manuscript.

References

1. Boutayeb A, Boutayeb S, Boutayeb W. Multi-morbidity of non communicable diseases and equity in WHO Eastern Mediterranean countries. Int J Equity Health. 2013;12(1):60. doi: http://doi.org/10.1186/1475-9276-12-60. PubMed PMID: 23961989.

2. World Health Organization. Global Health Estimates 2016: deaths by cause, age, sex, by country and by region, 2000–2016 [Internet]. Geneva: WHO; 2018 [cited 2024 June 10]. Available from: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates

3. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância de Doenças e Agravos não Transmissíveis e Promoção da Saúde. SAÚDE BRASIL 2018: uma análise da situação de saúde e das doenças e agravos crônicos: desafios e perspectivas. Brasília: Ministério da Saúde; 2019 [cited 2024 June 10]. 424 p. Available from: https://bvsms.saude.gov.br/bvs/publicacoes/saude_brasil_2018_analise_situacao_saude_doencas_agravos_cronicos_desafios_perspectivas.pdf

4. International Society of Nephrology. KDIGO – Kidney Disease: Improving Global Outcomes. Clinical Practice Guideline for Lipid Management in Chronic Kidney Disease. Kidney Int Suppl. 2013;3(1):163. Available from: https://kdigo.org/wp-content/uploads/2017/02/KDIGO-2013-Lipids-Guideline-English.pdf

5. Webster AC, Nagler EV, Morton RL, Masson P. Chronic Kidney Disease. Lancet. 2017;389(10075):1238–52. doi: http://doi.org/10.1016/S0140-6736(16)32064-5. PubMed PMID: 27887750.

6. Saran R, Robinson B, Abbott KC, Agodoa LYC, Bhave N, Bragg-Gresham J, et al. US Renal Data System 2017 Annual Data Report: Epidemiology of Kidney Disease in the United States. Am J Kidney Dis. 2018;71(3, Suppl 1):A7. doi: http://doi.org/10.1053/j.ajkd.2018.01.002. PubMed PMID: 29477157.

7. Aguiar LK, Prado RR, Gazzinelli A, Malta DC. Factors associated with chronic kidney disease: epidemiological survey of the National Health Survey. Rev Bras Epidemiol. 2020;23:e200044. doi: http://doi.org/10.1590/1980-549720200044. PubMed PMID: 32520099.

8. Park YJ, Kim JM. Klotho and postmenopausal hormone replacement therapy in women with chronic kidney disease. J Menopausal Med. 2018;24(2):75–80. doi: http://doi.org/10.6118/jmm.2018.24.2.75. PubMed PMID: 30202755.

9. Pei F, Zhou Z, Li Y, Ren Y, Yang X, Liu G, et al. Chronic kidney disease in Chinese postmenopausal women: a cross-sectional survey. Niger J Clin Pract. 2017;20(2):153–7. doi: http://doi.org/10.4103/1119-3077.198314. PubMed PMID: 28091429.

10. Tune JD, Goodwill AG, Sassoon DJ, Mather KJ. Cardiovascular consequences of Metabolic Syndrome. Transl Res. 2017;183: 57–70. doi: http://doi.org/10.1016/j.trsl.2017.01.001. PubMed PMID: 28130064.

11. Silveira Rossi JL, Barbalho SM, Reverete de Araujo R, Bechara MD, Sloan KP, Sloan LA. Metabolic syndrome and cardiovascular diseases: going beyond traditional risk factors. Diabetes Metab Res Rev. 2022;38(3):e3502. doi: http://doi.org/10.1002/dmrr.3502. PubMed PMID: 34614543.

12. Silva SB, Caulliraux HM, Araújo CA, Rocha E. Cost comparison of kidney transplant versus dialysis in Brazil. Cad Saude Publica. 2016; 32(6):S0102-311X2016000605005. doi: http://doi.org/10.1590/0102-311X00013515. PubMed PMID: 27383457.

13. Jesus NM, Souza GF, Mendes-Rodrigues C, Almeida Neto OP, Rodrigues DDM, Cunha CM. Quality of life of individuals with chronic kidney disease on dialysis. J Bras Nefrol. 2019;41(3):364–74. doi: http://doi.org/10.1590/2175-8239-jbn-2018-0152. PubMed PMID: 30720851.

14. Popoola PA, Tapamo JR, Assounga AG. Cluster analysis of mixed and missing chronic kidney disease data in KwaZulu-Natal Province, South Africa. IEEE Access 2021;9:52125–43. doi: http://doi.org/10.1109/ACCESS.2021.3069684.

15. Chen X, Wang X, Zhang K, Fung KM, Thai TC, Moore K, et al. Recent advances and clinical applications of deep learning in medical image analysis. Med Image Anal. 2022;79:102444. doi: http://doi.org/10.1016/j.media.2022.102444. PubMed PMID: 35472844.

16. Acs B, Rantalainen M, Hartman J. Artificial intelligence as the next step towards precision pathology. J Intern Med. 2020;288(1):62–81. doi: http://doi.org/10.1111/joim.13030. PubMed PMID: 32128929.

17. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro 3rd AF, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009;150(9):604–12. doi: http://doi.org/10.7326/0003-4819-150-9-200905050-00006. PubMed PMID: 19414839.

18. Alberti KG, Zimmet P, Shaw J. Metabolic syndrome-a new world-wide definition. A Consensus Statement from the International Diabetes Federation. Diabet Med. 2006;23(5):469–80. doi: http://doi.org/10.1111/j.1464-5491.2006.01858.x. PubMed PMID: 16681555.

19. Moore JX, Chaudhary N, Akinyemiju T. Metabolic syndrome prevalence by race/ethnicity and sex in the United States, National Health and Nutrition Examination Survey, 1988–2012. Prev Chronic Dis. 2017;14:E24. doi: http://doi.org/10.5888/pcd14.160287. PubMed PMID: 28301314.

20. Lohman TG, Roche AF, Martorell R. Anthropometric standardization reference manual: Abridged edition. Champaign, IL: Human Kinetics Books; 1991.

21. Uddin S, Haque I, Lu H, Moni MA, Gide E. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. Sci Rep. 2022;12(1):6256. doi: http://doi.org/10.1038/s41598-022-10358-x. PubMed PMID: 35428863.

22. Joseph VR. Optimal ratio for data splitting. Stat Anal Data Min. 2022;15(4):531–8. doi: http://doi.org/10.1002/sam.11583.

23. Borges LSR. Diagnostic accuracy measures in cardiovascular research. Int J Cardiovasc Sci. 2016;29(3):218–22. doi: http://doi.org/10.5935/2359-4802.20160030.

24. Hoo ZH, Candlish J, Teare D. What is an ROC curve? Emerg Med J. 2017;34(6):357–9. doi: http://doi.org/10.1136/emermed-2017-206735. PubMed PMID: 28302644.

25. Comini LO, de Oliveira LC, Borges LD, Dias HH, Batistelli CRS, da Silva LS, et al. Individual and Combined Components of Metabolic Syndrome with Chronic Kidney Disease in Individuals with Hypertension and/or Diabetes Mellitus Accompanied by Primary Health Care. Diabetes Metab Syndr Obes. 2020;13:71–80. doi: http://doi.org/10.2147/DMSO.S223929. PubMed PMID: 32021353.

26. Mankowski RT, Aubertin-Leheudre M, Beavers DP, Botoseneanu A, Buford TW, Church T, et al. Sedentary time is associated with the metabolic syndrome in older adults with mobility limitations — The LIFE Study. Exp Gerontol. 2015;70:32–6. doi: http://doi.org/10.1016/j.exger.2015.06.018. PubMed PMID: 26130060.

27. Bakhshayeshkaram M, Roozbeh J, Heidari ST, Honarvar B, Dabbaghmanesh MH, B Lankarani K. Relationships between various components of metabolic syndrome and chronic Kidney Disease in Shiraz, Iran. Int J Endocrinol Metab. 2019;17(2):e81822. doi: http://doi.org/10.5812/ijem.81822. PubMed PMID: 31372168.

28. Madero M, Katz R, Murphy R, Newman A, Patel K, Ix J, et al. Comparison between Different Measures of Body Fat with Kidney Function Decline and Incident CKD. Clin J Am Soc Nephrol. 2017;12(6):893–903. doi: http://doi.org/10.2215/CJN.07010716. PubMed PMID: 28522656.

29. Ding C, Yang Z, Wang S, Sun F, Zhan S. The associations of metabolic syndrome with incident hypertension, type 2 diabetes mellitus and chronic kidney disease: a cohort study. Endocrine. 2018;60(2):282–91. doi: http://doi.org/10.1007/s12020-018-1552-1. PubMed PMID: 29492904.

30. Hall JE, Brands MW, Dixon WN, Smith Jr MJ. Obesity-induced hypertension. Renal function and systemic hemodynamics. Hypertension. 1993;22(3):292–9. doi: http://doi.org/10.1161/01.HYP.22.3.292. PubMed PMID: 8349321.

31. Neves PDMM, Sesso RCC, Thomé FS, Lugon JR, Nasicmento MM. Brazilian dialysis census: analysis of data from the 2009–2018 decade. J Bras Nefrol. 2020;42(2):191–200. doi: http://doi.org/10.1590/2175-8239-jbn-2019-0234. PubMed PMID: 32459279.

32. Malta DC, Andrade SSCA, Oliveira TP, Moura L, Prado RRD, Souza MFM. Probability of premature death for chronic non-communicable diseases, Brazil and Regions, projections to 2025. Rev Bras Epidemiol. 2019;22:e190030. doi: http://doi.org/10.1590/1980-549720190030. PubMed PMID: 30942336.

33. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). JAMA. 2001;285(19):2486–97. doi: http://doi.org/10.1001/jama.285.19.2486. PubMed PMID: 11368702.

34. Lichtenberger JP, Tatum PS, Gada S, Wyn M, Ho VB, Liacouras P. Using 3D printing (additive manufacturing) to produce low-cost simulation models for medical training. Mil Med.

2018;183(Suppl 1):73–7. doi: http://doi.org/10.1093/milmed/usx142. PubMed PMID: 29635555.

35. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One. 2017;12(4):e0174944. doi: http://doi.org/10.1371/journal.pone.0174944. PubMed PMID: 28376093.

36. Olivera AR, Roesler V, Iochpe C, Schmidt MI, Vigo Á, Barreto SM, et al. Comparison of machine-learning algorithms to build a predictive model for detecting undiagnosed diabetes – ELSA-Brasil: accuracy study. Sao Paulo Med J. 2017;135(3):234–46. doi: http://doi.org/10.1590/1516-3180.2016.0309010217. PubMed PMID: 28746659.

37. Pereira CR, Pereira DR, Silva FA, Masieiro JP, Weber SA, Hook C, et al. A new computer vision-based approach to aid the diagnosis of Parkinson's disease. Comput Methods Programs Biomed. 2016;136:79–88. doi: http://doi.org/10.1016/j.cmpb.2016.08.005. PubMed PMID: 27686705.

38. Rosa CRAA. Predição de síndrome metabólica em usuários de antirretrovirais utilizando modelos computacionais [tese]. São Luís: Universidade Federal do Maranhão; 2019.

39. Kroll C, Mastroeni SSBS, Czarnobay SA, Ekwaru JP, Veugelers PJ, Mastroeni MF. The accuracy of neck circumference for assessing overweight and obesity: a systematic review and meta-analysis. Ann Hum Biol. 2017;44(8):667–77. http://doi.org/10.1080/03014460.2017.1390153. PubMed PMID: 29037078.

40. Sokol A, Wirth MD, Manczuk M, Shivappa N, Zatonska K, Hurley TG, et al. Association between the dietary inflammatory index, waist-to-hip ratio and metabolic syndrome. Nutr Res. 2016;36(11):1298–303. doi: http://doi.org/10.1016/j.nutres.2016.04.004. PubMed PMID: 27865615.

41. Wang H, Liu A, Zhao T, Gong X, Pang T, Zhou Y, et al. Comparison of anthropometric indices for predicting the risk of metabolic syndrome and its components in Chinese adults: a prospective, longitudinal study. BMJ Open. 2017;7(9):e016062. doi: http://doi.org/10.1136/bmjopen-2017-016062. PubMed PMID: 28928179.

42. Kim BJ, Kang JG, Kim BS. Association between secondhand smoke exposure and new-onset hypertension in self-reported never smokers verified by cotinine. Korean J Intern Med (Korean Assoc Intern Med). 2021;36(6):1377–88. doi: http://doi.org/10.3904/kjim.2021.214. PubMed PMID: 34742177.

43. Lee K, Giovannucci EL, Kim J. The effect of smoking and sex on the association between long-term alcohol consumption and metabolic syndrome in a middle-aged and older population. J Epidemiol. 2021;31(4):249–58. doi: http://doi.org/10.2188/jea.JE20190328. PubMed PMID: 32378517.

44. Lin IT, Lee MY, Wang CW, Wu DW, Chen SC. Gender differences in the relationships among metabolic syndrome and various obesity-related indices with nonalcoholic fatty liver disease in a taiwanese population. Int J Environ Res Public Health. 2021;18(3):857. doi: http://doi.org/10.3390/ijerph18030857. PubMed PMID: 33498329.