2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Phenotypic Characterization of Chronic Kidney Patients Through Hierarchical Clustering

Ronaldo S. Silva Jr[1]., Cindy L. Pereira[1], Naruna A. C. Melo[1], Giovana C. S. Silva[4], Carlos M. Sousa Jr[1]., Nilviane P. S. Sousa[1], Érika C. R. L. Caneiro[3], Allan K. D. B. Filho[1], Ewaldo E. C. Santana[2].

Abstract - **Chronic kidney disease is a major public health problem around the world and this disease early diagnosis is still a great challenge as it is asymptomatic in its early stages. Thus, in order to identify variables capable of assisting CKD diagnosis and monitoring, machine learning techniques and statistical analysis use has shown itself to be extremely promising. For this work, unsupervised machine learning, statistical analysis techniques and discriminant analysis were used.**

Clinical Relevance - **Discriminating variables characterization assist to differentiate groups of patients in different stages of Chronic Kidney Disease and it has important outcomes in the development of future models to aid clinical decision-making, as they can generate models with a greater predictive capacity for Chronic Kidney Disease, predominantly aiding the early diagnosis capacity of this pathology.**

I. INTRODUCTION

Chronic Kidney Disease is characterized by the slow, progressive and irreversible loss of the kidney function, generating an organism metabolic and electrolyte imbalance and it is a rising public health problem, along with a high morbidity and mortality worldwide [1], [2]. In the last decade, the number of cases recorded have been increasing significantly and they are associated with plenty of factors that extend since the ageing to population demographic transition [3].

A recent study [4] has shown that, in Brazil, chronic kidney disease and some main diseases associated with it represent about 7,61% of hospitalizations and 12,97% of hospitalization expenses. In 2015, hospitalizations expenses for every cause were of R$13,8 billion and more than R$2 billion with kidney transplant or dialysis. These numbers provide an important percentage of national expenses with health and there's a constant upward trend, even if only the punctual aspects are taken into consideration, like changes in the country's development profile and population ageing.

Besides, chronic kidney disease presents itself with a risk factor to COVID-19 critical stages development, together with mellitus diabetes, systemic arterial hypertension, pulmonary disease and cardiac disease [5] and according to Yichun Cheng, it has a high association with the fatal outcomes of hospital stays in cases of infection by the Covid-19 virus [6].[1]

This pathology early diagnosis is still considered a challenge since the initial stage is asymptomatic and clinical manifestations are more evident between moderate renal insufficiency to severe. In regards to the risk group definition, diabetics and hypertensive, the most used method of early identification of CKD is the patient's constant monitoring through medical tests that are able to assess the renal function [7].

In clinical practice, kidney excretory function can be measured by the Glomerular Filtration Rate (GFR), anyone who shows a TFG< 60mL/min/1,73m2 for at least three consecutive months is considered to have CKD. In case the person shows a TFG > 60mL/min/1,73m2, they will be considered to have CKD if associated with at least one renal damage marker, for example, albuminuria or polycystic kidney [8].

Overall, people who have CKD have a high physiological and biochemical alterations rate that occur in every stage of the disease, they are associated with the presence of comorbidities like obesity, diabetes, arterial hypertension, dyslipidemia, among others. Therefore, there's an emphasis on the importance of monitoring the risk factors through anthropometric and biochemical determinant variables, as a means to prevent the disease as well as the early diagnosis [9].

From this standpoint, this research aimed to analyze individuals with CKD distribution using unsupervised clustering methods.

II.METHODS

A. Sample

This is a cross-sectional study made with 104 people treated at a Reference Center for Kidney Diseases treatment in São Luís (MA), over 18 years old, of both genders and diagnosed with CKD. CKD stages classification was carried out based on the Ministry of Health Guidelines.

The participants were informed about the purpose of the study and, when agreed, signed the Free and Informed Consent Form. The study was approved by the Human Research Ethics Committee of the Federal University of Maranhão - UFMA according to legal opinion 2,035,753.

B. Inclusion criteria

People included aged 18 years or older, being monitored at the Kidney Disease Prevention Center - HUUFMA and presenting Glomerular Filtration Rate less than or equal to 89 mL / min / 1.73 m^2.

C. Exclusion Criteria

* Research supported by FAPEMA.

[1] Electrical Engineering Department, Biological Information Processing Lab, Federal University of Maranhão, São Luis, MA, Brazil.

[2]Laboratory of Signals Acquisition and Processing, LAPS, State University of Maranhão, São Luís, MA, Brazil.

[3]Presidente Dutra University Hospital - Federal University of Maranhão, São Luís, MA, Brazil.

[4] Language and Literature Department, Federal University of Maranhão, São Luís, MA, Brazil.

Pregnant women or people with physical disabilities that prevented or compromised data collection were excluded. Patients with incomplete collecting data and those who did not agree with the Informed Consent Form.

D. Data Gathering

The measurements were performed by a single researcher with the same calibrated instrument. The measurements were performed in duplicate and the averages were considered for data analysis.

Weight was measured with a calibrated electronic scale (Omron ® HBF 214 LA, Japan) with a 0.1 kg resolution. The height was determined with the aid of a vertical transportable stadiometer with a 0.1 cm resolution (Sanny ®, Brazil). The Body Mass Index (BMI) was obtained through the ratio between weight (Kg) and the square of height (m).

Circumferences were measured with an inelastic anthropometric measuring tape with a 0.1 cm precision (Seca ® 213, Hamburg, Germany). Waist circumference (WC) was measured at the midpoint between the last rib and the iliac crest at minimum breathing. Circumference of the right arm and calf were measured as described in [10]. Hip circumference followed the procedures described by Lohman [11]. The waist-to-height ratio was calculated using the ratio between waist circumference (cm) and height (cm) [12].

All biochemical analysis were performed on cobas 6000 automated device (Roche) following the methodology described by the manufacturer. The gathering was performed under vacuum with a multiple collection system.

Systolic blood pressure (SBP) and diastolic blood pressure (DBP) were measured with the aid of an arm blood pressure monitor (OMRON®, model HEM 7130). The measurement and cut-off point used followed the recommendations of VII Brazilian Guideline for Hypertension [13]in which SBP values $\leq$ 120 mmHg and DBP $\leq$ 80mmHg were considered normal and SBP values> 120 mmHg and of PAD> 80 mmHg were altered.

The variables used were attributed to the study based on indicators for nutritional and health assessment described in the literature and recommended by the World Health Organization (WHO). Anthropometric indicators were consolidated indexes in the assessment of central adiposity and cardiovascular risk [8], [10], [14], [15].

E. Statistical Analysis

To make the clusters definition, it was applied to the agglomerative hierarchical cluster method, it consists of an unsupervised machine learning algorithm, that, a priori, does not require a cluster number specification, but it requires a measure dissimilarity specification between the observations [16]. This approach allows the determination of variables that better discriminate the clusters that were found.

ANOVA one-way, Kruskal – Wallis and Chi-Square tests were executed to analyze the cluster difference between parametric variables, not parametric and categorical, respectively.

A discriminant analysis was applied to identify factors that better discriminate the clusters found, using the Fisher Linear discriminant. Therefore, the dependent variable is represented by the cluster classification found in the hierarchical method and the independent variables were the same used in the cluster analysis. The significance level adopted for the statistical tests was 5%. Python 3.0 and IBM SPSS Statistics programs were utilized for hierarchical analysis and statistics, respectively.

III.RESULTS

The clinical data of 104 people with CKD were available for the hierarchical cluster analysis, which resulted in three well defined clusters. The first cluster included 46 people and presented lower age, weight, waist circumference, hip circumference, systolic blood pressure, urea and waist-to-height ratio average than if compared to the other two discovered clusters. The median glomerular filtration rate for the first cluster, however, was 77,15 mL/min/1,73m², the highest between the two clusters, indicating the presence of people in the initial stages of Chronic Kidney Disease.

The second cluster encompassed the least number of people (n=25) and granted the highest weight, waist circumference, waist-to-height ratio, BMI, hip circumference and systolic blood pressure averages, indicating anthropometric parameters modification as well as higher age, total cholesterol and urea. The median glomerular filtration rate was 63,5 mL/min/1,73m², much lower than the 1st cluster, showing a difference in the Chronic Kidney Disease stage profile between these two clusters.

The third cluster, with 33 people, provided intermediate values, if compared with the first and the second cluster in all the continuous variables, except the glomerular filtration rate, which showed a lower value, that was 63 mL/min/1,73m², suggesting the presence of advanced stages CKD individuals in this cluster. After the statistics test's completion, it was noticed that only the weight, waist circumference and waist-to-height ratio variables exhibited significant differences between the three discovered clusters.

According to Table 1, per use of categorical variables, cluster characteristics discovered are displayed. Cluster 1 presented a woman (95,7%) and normal urea (93,5%) prevalence, as for CKD stages, it had the total of its cluster constituents in the first and second stages, which are the least advanced stages of CKD. In the remaining parameters, the first cluster presented a normoglycemic individual prevalence (78,3%) as well as an altered IMC (60,9%). Also demonstrating a female predominance (92,0%), the second cluster has a BMI prevalence of 100%, WC and WHtR altered variables patients and an elevated hemodynamic profile predominance (68,0%). 92,0% of this cluster doesn't practice physical activity frequently and 16% are already in the fourth and fifth stages of the CKD, already showcasing a significant progression in relation to the first cluster.

Different from the other two clusters, the third group presented a male prevalence of 90,9%, however, akin to the second group, it presents a BMI (63,6%), WHtR (90,9%) and WC (66,7%) altered predominance, as well as 21,2% of this cluster constituents are already on the fourth and fifth stage of CKD, demonstrating a heightened progression when compared to the second group. The biochemical profile showed a preponderance in total cholesterol (78,8%), fasting glycaemia (69,7%), triglycerides (69,7%), normal urea (63,6%) variables.

TABLE 1. Categorical Variables Analysis

Characteristics	n	Cluster 1	Cluster 2	Cluster 3	p
Patient number	104	46	25	33	-
Sex					
Female	70	44 (95,7%)	23 (92,0%)	3 (9,1%)	<0,001
Male	34	2 (4,3%)	2 (8%)	30 (90,9%)	
BMI					
Normal	30	18 (39,1%)	0 (0%)	12 (36,4%)	0,62
Altered	74	28 (60,9%)	25 (100%)	21 (63,6%)	
WHtR					
Normal	23	20 (43,5%)	0 (0%)	3 (9,1%)	< 0,001
Altered	81	26 (56,5%)	25 (100%)	30 (90,9%)	
WC					
Normal	35	24 (52,2%)	0 (0%)	11 (33,3%)	<0,044
Elevaded	69	22 (47,8%)	25 (100%)	22 (66,7%)	
Physical Activity					
Don't do it	85	34 (73,9%)	23 (92,0%)	28 (84,8%)	0,116
Do it	19	12 (26,1%)	2 (8,0%)	5 (15,2%)	
CKD Stages					
2	74	41 (89,1%)	16 (64%)	17 (51,5%)	< 0,001
3	19	5 (10,9%)	5 (20,0%)	9 (27,3%)	
4	9	0 (0%)	3 (12%)	6 (18,2%)	
5	2	0 (0%)	1 (4,0%)	1 (3%)	
Total Cholesterol					
Normal	59	20 (43,5%)	13 (52,0%)	26 (78,8%)	0,002
Elevaded	45	26 (56,5%)	12 (58,0%)	7 (21,2%)	
Fasting Glycaemia					
Normoglycaemia	74	36 (78,3%)	15 (60,0%)	23 (69,7%)	0,388
Prediabetes or high risk	30	10 (21,7%)	10(40,0%)	10 (30,3%)	
Triglycerides					
Normal	64	28 (60,9%)	13 (52,0%)	23 (69,7%)	0,559
Elevaded	40	18 (39,1%)	12 (48,0 %)	10 (30,3%)	
Urea					
Normal	78	43 (93,5%)	14 (56,0%)	21 (63,6%)	0,002
Elevaded	26	3 (6,5%)	11 (44,0%)	12 (36,4%)	
Perfil Hemodynamic					
Normal	55	30 (65,2%)	8 (32,0%)	17 (51,5%)	0,165
Elevaded	49	16 (34,8%)	17 (68,0%)	16 (48,5%)	

IV.DISCRIMINANT ANALYSIS

The multiple discriminant analysis, made in accordance with the use of the same variables applied to the cluster analysis, indicated four variables that have a high discrimination power between the clusters found in the cluster analysis: Waist Circumference, Waist-To-Height Ratio, Hip Circumference and Weight. The discriminant model utilized was 93,3% accurate when the clusters found were the dependent variable.

V. DISCUSSION

The discriminant analysis is a statistical multivariate technique that allows the most relevant variables identification to comprehend the differences between heterogeneous clusters in a determined context, but homogeneous amongst themselves [17]. According to Marôco [18], the discriminant analysis allows the variables identification that best discriminate between two or more clusters, which makes possible to assess the relevance of one or more variables within a data set [19] to identify variables that better represent Brazil States' indebtedness situation and by Cabral [20] to characterize asthma phenotypes in low-income children and teenagers. In this research, the applied discriminant analysis indicated four anthropometric variables (WC, WHtR, HC and Weight) with great capacity for discrimination between clusters.

Anthropometric measurements assessment has been a tool of choice by the clinical evaluation professionals and researchers due to it being a fast, low-cost and practical method. Waist circumference is widely used as it is directly associated with abdominal fat accumulation, that has high metabolic and inflammatory activity, and the highest risk of mortality due to cardiovascular complications [21]. A research involving workers of a Japanese company made a correlation between WC, WHtR, HC measurements as a risk factor for GFR reduction, indicating a better and a worse performance to WHtR and BMI, respectively. Therefore, WC and WHtR are considered efficient variables in central adiposity assessment, enabling cardiovascular risk associated with CKD identification [22].

Studies show that obesity is one of the main causes of comorbidities like insulin resistance, hypertension, hyperlipidemia, atherosclerosis and glucose intolerance, and emphasize that insulin resistance can be involved in dyslipidemia pathophysiology in Chronic Kidney Disease [23].

The current study presented a bigger CKD prevalence in female patients in the first and second clusters, similar to what can be found in [24] works in Goías, with a 70,9% women prevalence. It is proven that women have continuous medical monitoring in comparison to men, that can be associated with a significant predominance of female patients in a conservative stage of CKD, which is explained by the fact that this population tends to be an assiduous participant in regards to health care. It is believed that women prevalence in conservative stage is higher because in this phase, the patient still has the possibility to choose whether to have a medical follow-up or not, which, in more advanced stages, becomes mandatory as it is a way to preserve life [25].

The high kidney damage prevalence in elderly people is associated to the process of aging that results in a significant reduction of GFR, caused by occurrence of natural biological phenomenon cellular senescence in which there's a loss of nephron numbers, besides alterations in the glomerular basement membrane that can lead to increased protein excretion [26].

As for the anthropometric data, the three groups displayed body mass index alterations (BMI), similar data found in [25] studies made with adults and the elderly in which 41% showed overweight/obesity (BMI ≥ 25 kg/m2). Obesity increases risk of comorbidities associated with the development of CKD, like SAH and diabetes, in addition to influencing the disease stages progression by hyperfiltration that happens due to demands of the body weight and the intraglomerular tension increase that deteriorates kidney structures [27].

The higher biochemical parameter and hemodynamic profile control can be justified by follow up with health professionals in reference centers to people with CKD, by enhancing adherence to the treatment, making them reflect about the centrality of its role in conducting the instituted therapy and thus providing better clinical results [28]. In a systematic review, Nicoll [29]evaluated patients with chronic kidney disease efficient care models. Exhibiting evidence that multidisciplinary care in CKD management shows improvement in renal and metabolic results and better adherence to treatment.

VI.CONCLUSION

The phenotype analysis demonstrates the importance of applying discriminating variables according to the Chronic Kidney Disease (CKD) stage, such conclusion may aid intelligent systems development for CKD prediction. Amongst these variables, anthropometric indicators stand out, which are considered low cost and of easy use in clinical practice.

Therefore, new studies employing such parameters are important and necessary for monitoring risk groups and patients with kidney disease based on machine learning techniques.

ACKNOWLEDGMENT

To CAPES and FAPEMA, UFMA University Hospital and the adults who agreed to participate in the study.

REFERENCES

[1] A. W. G. B. Marinho, A. da P. Penha, M. T. Silva, and T. F. Galvão, “Prevalência de doença renal crônica em adultos no Brasil: revisão sistemática da literatura,” *Cad. Saúde Coletiva*, vol. 25, no. 3, pp. 379–388, 2017, doi: 10.1590/1414-462x201700030134.

[2] M. Provenzano *et al.*, “Contribution of predictive and prognostic biomarkers to clinical research on chronic kidney disease,” *Int. J. Mol. Sci.*, vol. 21, no. 16, pp. 1–25, 2020, doi: 10.3390/ijms21165846.

[3] J. W. Stanifer, A. Muiru, T. H. Jafar, and U. D. Patel, “Chronic kidney disease in low- and middle-income countries,” *Nephrol. Dial. Transplant.*, vol. 31, no. 6, pp. 868–874, 2016, doi: 10.1093/ndt/gfv466.

[4] P. R. Alcalde and G. M. Kirsztajn, “Expenses of the Brazilian Public Healthcare System with chronic kidney disease,” *J. Bras. Nefrol.*, vol. 40, no. 2, pp. 122–129, 2018, doi: 10.1590/2175-8239-JBN-3918.

[5] A. S. Schnake-Mahl, M. G. Carty, G. Sierra, and T. Ajayi, “Risk of Severe Covid-19 Complications: Building an Actionable Rules-Based Model for Care Teams,” *NEJM Catal. Innov. Care Deliv.*, vol. 1, no. 3, pp. 1–13, 2020, doi: 10.1056/CAT.20.0116.

[6] Y. Cheng *et al.*, “Kidney disease is associated with in-hospital death of patients with COVID-19,” *Kidney Int.*, vol. 97, no. 5, pp. 829–838, 2020, doi: 10.1016/j.kint.2020.03.005.

[7] J. C. F. Schaefer, M. S. Pereira, C. R. de Jesus, F. Schuelter-Trevisol, and D. J. Trevisol, “Kidney function estimate among subjects aged 18-59 years in Tubarão, Santa Catarina: a population-based study,” *J. Bras. Nefrol.*, vol. 37, no. 2, pp. 185–191, 2015, doi: 10.5935/0101-2800.20150030.

[8] Brasil Miniistério da Saude, “DIRETRIZES CLÍNICAS PARA O CUIDADO AO PACIENTE COM DOENÇA RENAL CRÔNICA – DRC NO SISTEMA ÚNICO DE SAÚDE,” *Secr. Atenção à Saúde*, vol. 1, pp. 1–37, 2014.

[9] L. A. B. P. Peres and T. E. Bettin, “Dislipidemia em pacientes com doença renal crônica,” *Rev. Soc. Bras. Clín. Méd*, pp. 10–13, 2015.

[10] F. V. Salmaso *et al.*, “Analysis of elderly outpatients in relation to nutritional status, sarcopenia, renal function, and bone density,” *Arq Bras Endocrinol Metab*, vol. 58, no. 3, pp. 226–231, 2014.

[11] T. G. Lohman, A. F. Roche, and R. Martorell, “Anthropometric standardization reference manual,” *Champaign Hum. Kinet. books*, vol. 177, 1988.

[12] M. Ashwell and S. D. Hsieh, “Six reasons why the waist-to-height ratio is a rapid and effective global indicator for health risks of obesity and how its use could simplify the international public health message on obesity,” *Int. J. Food Sci. Nutr.*, vol. 56, no. 5, pp. 303–307, 2005, doi: 10.1080/09637480500195066.

[13] S. B. de Cardiologia, *7th Brazilian Guideline of Arterial Hypertension: Presentation*, vol. 107, no. 3. 2016.

[14] P. Jésus *et al.*, “Undernutrition and obesity among elderly people living in two cities of developing countries: Prevalence and associated factors in the EDAC study,” *Clin. Nutr ESPEN*, vol. 21, pp. 40–50, 2017.

[15] M. C. MANCINI, “Diretrizes brasileiras de obesidade,” *Assoc. Bras. Para O Estud. da Obesidade e da Síndrome Metabólica*, vol. 4, 2016.

[16] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Second Edi. 208AD.

[17] A. J. Lacruz, B. L. Américo, and F. Carniel, “Quality indicators in education: Discriminant analysis of the performances in Prova Brasil,” *Rev. Bras. Educ.*, vol. 24, pp. 1–26, 2019, doi: 10.1590/S1413-24782019240002.

[18] J. Marôco, “Statistical Analysis with SPSS application,” *ReportNumber, Lda*, vol. 5, pp. 1–12, 2007.

[19] G. R. de Mello, F. D. Q. Macedo, F. Tavares Filho, and V. Slomski, “Identificando o endividamento dos estados brasileiros: uma proposta através de análise discriminante,” *Enfoque: Reflexão Contábil*, vol. 25, no. 1, pp. 5–14, 2006, doi: 10.4025/enfoque.v25i1.3504.

[20] A. L. B. Cabral, A. W. Sousa, F. A. R. Mendes, and C. R. F. de Carvalho, “Phenotypes of asthma in low-income children and adolescents: cluster analysis,” *J. Bras. Pneumol.*, vol. 43, no. 1, pp. 44–50, 2017, doi: 10.1590/s1806-37562016000000039.

[21] B. C. Fontes, J. S. Dos Anjos, A. P. Black, N. X. Moreira, and D. Mafra, “Effects of Low-Protein Diet on lipid and anthropometric profiles of patients with chronic kidney disease on conservative management,” *J. Bras. Nefrol.*, vol. 40, no. 3, pp. 225–232, 2018, doi: 10.1590/2175-8239-JBN-3842.

[22] K. Odagiri, I. Mizuta, M. Yamamoto, Y. Miyazaki, H. Watanabe, and A. Uehara, “Waist to height ratio is an independent predictor for the incidence of chronic kidney disease,” *PLoS One*, vol. 9, no. 2, 2014, doi: 10.1371/journal.pone.0088873.

[23] A. Meyrier, “Nephrosclerosis: A Term in Quest of a Disease,” *Nephron*, vol. 129, no. 4, pp. 276–282, 2015, doi: 10.1159/000381195.

[24] V. A. G. AMADOR, A. T. V. de S. FREITAS, A. V. NAGHETTINI, E. R. S. PEREIRA, and M. do R. G. PEIXOTO, “Anthropometric measurements and markers of renal function in adults and older adults,” *Rev. Nutr.*, vol. 29, no. 2, pp. 199–209, 2016.

[25] S. M. Rembold, D. L. S. dos Santos, G. B. Vieira, M. S. Barros, and J. R. Lugon, “Demographic profile of individuals with chronic renal disease from a multidisciplinary outpatient clinic of a university teaching hospital [Portuguese].,” *Acta Paul. Enferm.*, vol. 22, pp. 501–504, 2009, [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=2010429709&site=ehost-live.

[26] T. L. M. Amaral, C. de A. Amaral, M. T. L. De Vasconcellos, and G. T. R. Monteiro, “Prevalence and factors associated to chronic kidney disease in older adults,” *Rev. Saude Publica*, vol. 53, pp. 1–11, 2019, doi: 10.11606/S1518-8787.2019053000727.

[27] C. P. Kovesdy, S. L. Furth, and C. Zoccali, “Obesity and kidney disease: Hidden consequences of the epidemic,” *Brazilian J. Med. Biol. Res.*, vol. 50, no. 5, pp. 1–9, 2017, doi: 10.1590/1414-431X20166075.

[28] S. M. de S. B. Lins, J. L. Leite, S. de Godoy, J. M. A. B. Tavares, R. G. Rocha, and F. V. C. e Silva, “Treatment adherence of chronic kidney disease patients on hemodialysis,” *ACTA Paul. Enferm.*, vol. 31, no. 1, pp. 54–60, 2018.

[29] R. Nicoll, L. Robertson, E. Gemmell, P. Sharma, C. Black, and A. Marks, “Models of care for chronic kidney disease: A systematic review,” *Nephrology*, vol. 23, no. 5, pp. 389–396, 2018.