

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

PONTO FOCAL - MARANHÃO

LUANA MONTEIRO ANAISSE AZOUBEL

**ESTIMATIVA DA CREATININA SÉRICA EM ADULTOS HÍGIDOS E RENAIIS
CRÔNICOS ATRAVÉS DE TÉCNICAS DE APRENDIZADO DE MÁQUINA
UTILIZANDO VARIÁVEIS NÃO INVASIVAS**

São Luís
2022

LUANA MONTEIRO ANAISSE AZOUBEL

**ESTIMATIVA DA CREATININA SÉRICA EM ADULTOS HÍGIDOS E RENAIIS
CRÔNICOS ATRAVÉS DE TÉCNICAS DE APRENDIZADO DE MÁQUINA
UTILIZANDO VARIÁVEIS NÃO INVASIVAS**

Tese apresentada ao Programa de Pós-Graduação em Biotecnologia da Rede Nordeste de Biotecnologia, ponto focal do Maranhão, como requisito parcial para obtenção do Título de Doutora em Biotecnologia.

Orientador: Prof. Dr. Allan Kardec Duailibe Barros Filho

São Luís
2022

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Monteiro Anaisse Azoubel, Luana.

Estimativa da creatinina sérica em adultos hígidos e renais crônicos através de técnicas de aprendizado de máquina utilizando variáveis não invasivas / Luana Monteiro Anaisse Azoubel. - 2022.

111 f.

Coorientador(a): Nilviane Pires Silva Sousa.

Orientador(a): Allan Kardec Duailibe Barros Filho.

Tese (Doutorado) - Programa de Pós-graduação em Biotecnologia - Renorbio/ccbs, Universidade Federal do Maranhão, São Luís, 2022.

1. Aprendizado de Máquina. 2. Creatinina sérica. 3. Estimador. 4. Função renal. 5. Regressão linear múltipla. I. Duailibe Barros Filho, Allan Kardec. II. Pires Silva Sousa, Nilviane. III. Título.



Universidade Federal do Maranhão
Av. dos Portugueses, 1966, Cidade Universitária – 65080-805 São Luís-MA
Telefone (98) 3272- 9531 E-mail: renorbio@ufma.br
Homepage: <http://www.renorbio.org.br>



FOLHA DE APROVAÇÃO DE DEFESA DE TESE

ALUNA: Luana Monteiro Anaisse Azoubel

TÍTULO DO PROJETO: ESTIMATIVA DA CREATININA SÉRICA EM ADULTOS HÍGIDOS E RENAIIS CRÔNICOS ATRAVÉS DE TÉCNICAS DE APRENDIZADO DE MÁQUINA UTILIZANDO VARIÁVEIS NÃO INVASIVAS.

PROFESSOR ORIENTADOR: Allan Kardec Duailibe Barros Filho

BANCA EXAMINADORA:

CONCEITO

ASSINATURA

Prof. Dr. Allan Kardec Duailibe Barros Filho – UFMA
(Presidente)

Prof^a. Dr^a. Maria do Desterro Soares Brandão
Nascimento – UFMA
(Titular)

Prof^a. Dr^a. Gisele Cutrim de Oliveira Santos – UEMA
(Titular)

Prof. Dr. Ewaldo Eder Carvalho Santana – UEMA
(Titular)

Prof^a. Dr^a. Cláudia Regina de Andrade Arrais Rosa – UFMA
(Titular)

DATA DA APROVAÇÃO: 4 de fevereiro de 2022
HORÁRIO: 08:00h

LUANA MONTEIRO ANAISSE AZOUBEL

**ESTIMATIVA DA CREATININA SÉRICA EM ADULTOS HÍGIDOS E RENAIIS
CRÔNICOS ATRAVÉS DE TÉCNICAS DE APRENDIZADO DE MÁQUINA
UTILIZANDO VARIÁVEIS NÃO INVASIVAS**

Tese apresentada ao Programa de Pós-Graduação em Biotecnologia da Rede Nordeste de Biotecnologia, ponto focal do Maranhão, como requisito parcial para obtenção do Título de Doutora em Biotecnologia.

Aprovada em ____ / ____ / ____

BANCA EXAMINADORA

Prof. Dr. Allan Kardec Duailibe Barros (Orientador)

Universidade Federal do Maranhão/RENORBIO (Orientador)

Profa. Dra. Maria do Desterro Soares Brandão Nascimento

Universidade Federal do Maranhão/RENORBIO (1º Examinador)

Profa. Dra. Giselle Cutrim de Oliveira Santos

Universidade Estadual do Maranhão-UEMA (2º Examinador)

Prof. Dr. Ewaldo Eder Carvalho Santana

Universidade Estadual do Maranhão-UEMA (3º Examinador)

Prof. Dra. Cláudia Regina de Andrade Arrais Rosa

Universidade Federal do Maranhão-UFMA (4º Examinador)

*"Tu, ó Senhor Deus és tudo o que tenho.
O meu futuro está nas tuas mãos; tu
diriges minha vida. Como são boas as
bençãos que me dás! Como são
maravilhosas!"*

Salmo 16, versículo 5.

A Deus, por renovar minhas forças a cada dia.
Ao meu esposo, por todo apoio e carinho.

AGRADECIMENTOS

A Deus, toda a honra e glória por me permitir chegar até aqui, com saúde e ânimo, e por seu infinito amor.

Ao meu esposo, pois me faz sentir amada cada dia da minha vida, por todo apoio incondicional, por toda a paciência.

À minha mãe Nilce Monteiro pelo incentivo aos estudos desde a primeira infância e aos meus irmãos Lorena Anaisse e Lorenzo Anaisse, para cada momento difícil em que me deram conselhos e tanto carinho.

Ao meu orientador, prof. Allan Kardec Barros, pela oportunidade de crescimento profissional e pessoal, e todo o aprendizado que levarei para minha vida.

À minha co-orientadora prof. Nilviane Pires, por todo apoio e ensinamentos ao longo desta jornada, por acreditar em meu potencial, e por sempre me oferecer palavras de motivação.

À minha mentora Érika Carneiro e coordenadora do Centro de Prevenção de Doenças Renais, anexo do Hospital Universitário do Maranhão, a quem tenho muita admiração e que acreditou no desenvolvimento do nosso projeto, nos recebeu com confiança e sempre foi uma grande incentivadora de minhas pesquisas.

Ao Hospital Universitário, instituição de referência, que possibilitou que esse projeto fosse realizado, abrindo suas portas, pensando no melhor para seus pacientes. Ao serviço de Nefrologia e à toda a equipe do Centro de Prevenção de Doenças Renais, que nos apoiaram em todo processo.

Aos colegas do Laboratório de Processamento da Informação Biológica – PIB, que foram essenciais para que esse projeto ocorresse, em especial o amigo Carlos Magno, pelos ensinamentos sobre aprendizado de máquina.

Aos usuários do serviço de Nefrologia por concordarem em participar de nosso estudo, disponibilizando seu tempo e confiando em nosso trabalho.

A FAPEMA, pelo financiamento concedido, permitindo a realização deste estudo.

RESUMO

A Doença renal crônica (DRC) tem se mostrado um importante problema de saúde pública. Estima-se que atualmente 11,7 a 15,1% da população mundial seja acometida por esta condição. Tal enfermidade traz consigo impactos para a qualidade de vida do indivíduo pois está associada a falência renal e outros desfechos adversos. A DRC não é uma doença de fácil detecção, e as barreiras para realização do diagnóstico vão além dos sintomas tardios, como a escassez de nefrologistas e difícil acesso aos exames específicos. A estimativa da taxa de filtração glomerular (TFG) através do uso da creatinina é o método mais recomendado para a prática clínica, porém é obtida por análise bioquímica de amostra sanguínea. Neste contexto, com intuito de facilitar o diagnóstico desta doença, têm se utilizado técnicas de Aprendizado de Máquina, pois seus algoritmos são capazes de aprender e analisar padrões e utilizá-los para solucionar demandas específicas no âmbito da saúde. Portanto, o objetivo desse estudo foi desenvolver um modelo matemático de alta confiabilidade e fácil manuseio, capaz de estimar a creatinina sérica para determinar a TFG, apenas com indicadores não invasivos e de baixo custo. Participaram deste estudo 116 indivíduos adultos hígidos e renais crônicos. Para testar a normalidade dos dados foi utilizado o teste de Kolmogorov-Smirnov, no software SPSS®. Para estimativa da creatinina foi utilizado o método de regressão linear múltipla com as características de entrada: gênero, idade, PAS e PAD, e correlação de Pearson para comparar os valores estimados e reais de creatinina e TFG, no programa MATLAB®. A estatística descritiva foi apresentada para ambos os grupos em valores absolutos para gênero e etnia, e em média e desvio padrão (DP) para: idade, IMC, estatura, peso, circunferência de cintura, PAS, PAD, Creatinina sérica e TFG. O modelo matemático utilizado apresentou correlação forte tanto para creatinina: $r=0,72$ com $DP=0,14$; quanto para TFG: $r=0,87$ com $DP=13,2$. O modelo computacional implementado neste estudo foi eficiente em estimar a creatinina sérica, demonstrando forte correlação entre o valor estimado e o real, o mesmo ocorreu para a TFG, porém com melhor desempenho. Portanto, o software desenvolvido neste estudo é capaz de estimar a função renal e pode ser um grande aliado para medidas de prevenção e tratamento precoce da DRC na população economicamente vulnerável.

Palavras-chave: Creatinina sérica. Função renal. Estimador. Aprendizado de Máquina. Regressão linear múltipla.

ABSTRACT

Chronic kidney disease (CKD) is an important public health problem. It is estimated that currently 11.7 to 15.1% of the world population is affected by this condition. This disease has impacts on the quality of life as it is associated with kidney failure and other adverse outcomes. CKD is not an easy-to-detect disease, and the barriers to making the diagnosis go beyond late symptoms, such as the scarcity of nephrologists and difficult access to specific tests. Estimation of glomerular filtration rate (GFR) through the use of creatinine is the most used method in clinical practice, but it is obtained by biochemical analysis of a blood sample. In this scenario, in order to facilitate the diagnosis of this disease, Machine Learning techniques have been used, as their algorithms are capable of learning and analyzing patterns and using them to solve specific demands in the health field. Thus, the objective of this study was to develop a mathematical model of high reliability and easy handling, capable of estimating the serum creatinine to determine the GFR, only with non-invasive and low-cost indicators. 116 healthy adults and chronic renal patients participated in this study. To test the normality of the data, the Kolmogorov-Smirnov test was used in the SPSS® software. To estimate creatinine, the multiple linear regression method was used with the input characteristics: gender, age, SBP and DBP, and Pearson's correlation to compare the estimated and actual values of creatinine and GFR, in the MATLAB® program. Descriptive statistics of the sample were presented as absolute values for gender and ethnicity, and as mean and standard deviation (SD) for: age, BMI, height, weight, waist circumference, SBP, DBP, serum creatinine and GFR. The mathematical model used showed a strong correlation for both creatinine: $r=0.72$ with $SD=0.14$; and GFR: $r=0.87$ with $SD=13.2$. The computational model implemented in this study was efficient in estimating the serum creatinine, showing a strong correlation between the estimated and the real value, the same occurred for the GFR, but with better performance. Therefore, the software developed in this study is able to estimate renal function and can be a great ally for preventive measures and early treatment of CKD in the low-income population.

Keywords: Serum creatinine. Kidney function. Estimator. Machine Learning. Multiple linear regression.

LISTA DE ILUSTRAÇÕES

Figura 1 – Censo brasileiro de diálise	15
Figura 2 – Hierarquia do Aprendizado de máquina	24
Figura 3 – Fluxograma do modelo proposto	30
Figura 4 – Layout de interface do software.....	31
Figura 5 – Relação entre creatinina sérica real e estimada pelo modelo proposto ..	34
Figura 6 – Relação entre TFG real e estimada utilizando o valor da creatinina estimada pelo modelo proposto	35

LISTA DA TABELAS

Tabela 1 – Classificação da pressão arterial e definição da hipertensão arterial para indivíduos a partir de 18 anos	21
Tabela 2 – Variáveis de entrada de modelo.....	29
Tabela 3 – Características Sócio-demográficas, antropométricas, laboratoriais e hemodinâmicas da amostra, hígidos e renais crônicos	33
Tabela 4 – Resultado da RLM, coeficiente (β) e significância (p) para cada variável do modelo	34

LISTA DA QUADROS

Quadro 1 – Estadiamento da DRC de acordo com a TFG	19
---	-----------

LISTA DA ABREVIATURAS

AM - Aprendizado de Máquina

CKD-EPI - Chronic Kidney Disease Epidemiology Collaboration

Crea - creatinina

DEXA - *Dual-energy X-ray Absorptiometry*

DRC - Doença Renal Crônica

HAS - Hipertensão Arterial Sistêmica

HUUFMA - Hospital Universitário da Universidade Federal do Maranhão

IMC - Índice de Massa Corporal

KDIGO - Kidney Disease Improving Global Outcomes

MDRD - Modificação da Dieta em Doença Renal

OMS - Organização Mundial de Saúde

PAD - Pressão arterial diastólica

PAS - Pressão arterial sistólica

PMP - Por milhão de população

RLM - Regressão linear múltipla

TCLE - Termo de Consentimento Livre e Esclarecido

TFG - Taxa de filtração glomerular

TRS - Terapia renal substitutiva

SUMÁRIO

1	INTRODUÇÃO	15
2	OBJETIVOS	17
2.1	Objetivo Geral	17
2.2	Objetivos específicos	17
3.	FUNDAMENTAÇÃO TEÓRICA	18
3.1	Doença renal crônica	18
3.2	Creatinina	19
3.3	Método para determinar a TFG	20
3.4	Fatores de risco relacionados à DRC	21
3.4.1	Hipertensão arterial sistêmica	21
3.4.2	Idade	22
3.4.3	Gênero	22
3.5	Aprendizado de Máquina	23
4	MATERIAIS E MÉTODOS	27
4.1	Modelo do Estudo e População	27
4.2	Cálculo Amostral	27
4.3	Crítérios de Inclusão e Não Inclusão	28
4.4	Instrumentos de Coleta e Avaliação dos Dados	28
4.4.1	Variáveis Sociodemográficas	28
4.4.2	Determinação dos Índices Antropométricos	28
4.4.3	Perfil Hemodinâmico	28
4.4.4	Exames laboratoriais	29
4.5	Escolha das Características de Entrada do Algoritmo	29
4.6	Implementação do Método de Aprendizado de Máquina	29
5	RESULTADOS	31
5.1	Desenvolvimento do Software	31
5.2	Análise estatística	32
5.3	Aspectos éticos	32
6	DISCUSSÃO	36
7	CONCLUSÃO	40
	REFERÊNCIAS	41
	ANEXOS	49

APÊNDICE.....	103
----------------------	------------

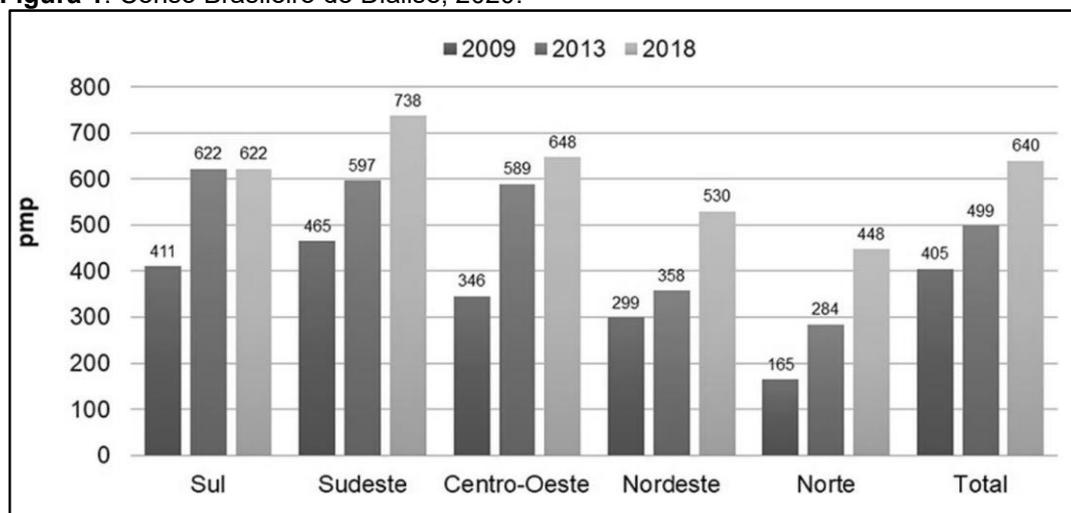
1 INTRODUÇÃO

A Doença renal crônica (DRC) tem se mostrado um importante problema de saúde pública, que vem crescendo tanto em países desenvolvidos como em desenvolvimento. Estudos recentes estimam que entre 11,7 - 15,1% da população mundial seja acometida por esta condição, e que existam entre 4,9 - 7 milhões de indivíduos que necessitam de terapia renal substitutiva (LV et al, 2019; HIMMELFARB et al, 2020). Aproximadamente 850 milhões de pessoas possuem algum tipo de doença renal, sendo a (grande) maioria a doença em estágio crônico (MURTON et al, 2020).

Segundo revisão sistemática, a média de brasileiros com DRC com algum grau de disfunção renal varia entre 3-6 milhões (MARINHO et al, 2017), e em diálise estima-se 133,464 indivíduos, de acordo com o último censo da Sociedade Brasileira de Nefrologia (NEVES et al, 2020).

Na região Nordeste do país, há uma prevalência estimada atualmente em 530 pacientes por milhão da população (pmp) e no Maranhão especificamente 276 pmp.

Figura 1. Censo Brasileiro de Diálise, 2020.



Fonte: NEVES, 2020.

Tal enfermidade traz consigo impactos para a qualidade de vida do indivíduo pois está associada à falência renal e outros desfechos adversos como doenças cardiovasculares, desnutrição, alta taxa de hospitalização e maior mortalidade por outras causas em comparação à população em geral (GO et al, 2004; MCMANUS & WYNTER-MINOTT, 2017). Um fator agravante é que a DRC não é uma doença de fácil detecção, pois seus sintomas mais evidentes, como edema, fadiga, falta de ar e hematúria, só aparecem em estágios muito avançados, onde já há uma perda

significativa da função renal (ROMAGNANI, et al 2017; CHEN et al, 2019; RAI et al, 2019).

Sabe-se que quanto mais precoce for o diagnóstico da DRC melhores serão os resultados para o paciente, pois assim, será possível o início antecipado do tratamento e conseqüentemente um retardamento na progressão da doença e atenuação de seus efeitos deletérios, redução de risco para eventos cardiovasculares e menor mortalidade nesta população (LEVEY et al, 2002; RAI et al, 2019).

As barreiras para realização do diagnóstico da DRC não são apenas os sintomas tardios, mas também a escassez de médicos nefrologistas e o difícil acesso aos serviços públicos de saúde para realização de exames específicos (BASTOS & KIRSZTAJN, 2011; MARINHO et al, 2017). Um dos métodos mais utilizado e recomendado na literatura é a estimativa da taxa de filtração glomerular através do uso da creatinina (LEVEY et al, 2002), e para se obter esse marcador, é necessário coletar amostras de sangue ou urina, para posterior análise bioquímica.

Neste contexto, com intuito de aprimorar e desenvolver técnicas capazes de prever e diagnosticar doenças em geral de forma mais eficiente, muitos estudiosos, e incluindo a Organização Mundial da Saúde (OMS), já fazem uso das técnicas de Aprendizado de Máquina (AM) (ALDHYANI et al, 2020; BIKIA et al, 2020; MATHUR et al, 2020; WILLIAMSON et al, 2020). O AM é um segmento da inteligência artificial, que estuda o desenvolvimento de ferramentas e técnicas capazes de solucionar problemas e auxiliar diagnósticos, fazer classificação e predição das mais diversas condições de saúde (NITHYA & ILANGO, 2017). Os algoritmos de AM são capazes de aprender e analisar padrões e utilizá-los para solucionar demandas específicas (WITTEN & FRANK, 2002).

Portanto, pretende-se neste estudo desenvolver um modelo matemático capaz de estimar a creatinina sérica e a partir dela calcular a taxa de filtração glomerular, apenas com indicadores não invasivos, que tenham alta confiabilidade, sejam de fácil aplicabilidade e baixo custo, facilitando o diagnóstico precoce e permitindo que o tratamento inicie ainda nos primeiros estágios da doença, promovendo uma desaceleração no ritmo de progressão da doença (CHENG et al, 2017).

2 OBJETIVOS

2.1 Objetivo Geral

Desenvolver modelo computacional capaz de prever creatinina sérica utilizando indicadores não invasivos e de baixo custo.

2.2 Objetivos específicos

- Estimar creatinina sérica
- Calcular a taxa de filtração glomerular a partir da creatinina estimada.
- Verificar dados sociodemográficos, laboratoriais, antropométricos e hemodinâmicos
- Desenvolver um *software* para prever a creatinina sérica e calcular a função renal estimada.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 Doença renal crônica

A incidência da DRC é um problema crescente, e está entre as 20 principais causas de mortes pelo mundo (LIU et al, 2019). Em sua maioria, essas mortes ocorrem em países onde há uma maior concentração de pessoas com baixa e média renda (BRADSHAW et al, 2019).

A DRC caracteriza-se pela diminuição gradual da função renal, quando o rim em um período mínimo de três meses apresenta taxa de filtração glomerular abaixo de $60\text{ml}/\text{min}/1,73\text{m}^2$, e/ou dano renal, e não tem a capacidade de manter sua funcionalidade adequada. Quando a função renal diminui para 5% (estágio 5 da DRC), o corpo perde a capacidade de manter o metabolismo em funcionamento normal, sendo necessário realizar uma terapia renal substitutiva, como diálise ou transplante renal (PIROJSAKUL et al, 2015; CHENG et al, 2017). Porém, o que gera grande preocupação é o fato da DRC ser detectada em geral somente em estágios mais avançados, quando a função renal diminui para 30%, isso ocorre pois em suas fases iniciais seus sintomas não são específicos e acabam sendo confundidos com outras doenças, resultando em um diagnóstico tardio (CHENG et al, 2017; RAI et al, 2019).

Divide-se a DRC em 5 estágios principais, determinados pela estimativa da taxa de filtração glomerular (eTFG). Entende-se como estágio 1: $e\text{TFG} \geq 90$ $\text{ml}/\text{min}/1.73\text{m}^2$, o estágio 2: $e\text{TFG} = 60\text{--}89$ $\text{ml}/\text{min}/1.73$ m^2 , no estágio 3: $e\text{TFG} = 30\text{--}59$ $\text{ml}/\text{min}/1.73$ m^2 , no estágio 4: $e\text{TFG} = 15\text{--}29$ $\text{ml}/\text{min}/1.73$ m^2 , e por fim o estágio 5: $e\text{TFG} < 15$ $\text{ml}/\text{min}/1.73\text{m}^2$ (LEVEY et al, 2002; RAI et al, 2019).

Nos estágios 1 e 2 da DRC a função renal encontra-se preservada, porém é possível identificar marcadores de dano renal, como a hematúria, proteinúria e albuminúria, já nos estágios seguintes (3 - 5) mesmo sem lesão no órgão, a TFG encontra-se prejudicada, aumentando o risco para problemas cardiovasculares e mortalidade (LEVIN et al, 2013).

Quadro 1. Estadiamento da DRC de acordo com a TFG.

Estágo	TFG (ml/min/1,73m²)	Situação
1	≥90	Normal ou alta
2	60-89	Ligeiramente diminuído
3a	45-59	Ligeiramente a moderadamente diminuído
3b	30-44	Moderadamente a severamente diminuído
4	15-29	Severamente diminuído
5	<15	Falência renal

Abreviações: DRC doença renal crônica; TFG taxa de filtração glomerular. Fonte: adaptado de KDIGO, 2013.

3.2 Creatinina

A creatinina é um subproduto do metabolismo muscular, proveniente da degradação da proteína fosfocreatina armazenada nas fibras musculares. Sofre influência dos hábitos alimentares, pela ingestão de proteína animal por exemplo, da quantidade de massa muscular e até mesmo pelo uso de medicamentos que podem bloquear a secreção tubular (PASALA & CARMODY, 2017). Por não possuir valor para o organismo, não é reabsorvida ou metabolizada, é apenas filtrada pelos rins para ser eliminada através do túbulo proximal, pela urina.

No entanto, quando o funcionamento do rim está prejudicado não há uma excreção adequada desse subproduto, o que culmina no aumento dos seus níveis séricos, por esse motivo a creatinina se mostra como um importante biomarcador para avaliar o funcionamento do rim (PECOITS-FILHO, 2004).

Porém, por sofrer influência de hábitos alimentares, uso de fármacos e da quantidade de massa muscular, a análise isolada da creatinina não é o método mais indicado para determinar a função renal. Para isso foram desenvolvidas fórmulas que consideram juntamente à creatinina sérica variáveis como gênero, idade e etnia, o que explica o fato de mesmo com valores de creatinina sérica iguais, os indivíduos que apresentam características sociodemográficos diferentes obtenham resultado distinto da TFG estimada (LEVIN et al, 2013; DELANAYE et al, 2017).

Os valores de referência de creatinina sérica em adultos no Brasil são: mulheres: 18 a 59 anos < 0,5 a < 1,1 mg/d e para homens: 18 a 59 anos < 0,6 a < 1,3 mg/dl. (MALTA et al, 2019; SZWARCOWALD et al, 2019).

3.3 Método para determinar a TFG

Na prática clínica alguns métodos com alta especificidade, como depuração de inulina, iohexol, 125I-iotalamato e 51Cr-EDTA, são impraticáveis devido às suas complexidades e alto custo (HERGET-ROSENTHAL et al, 2007; PRATES et al, 2007). Portanto, clinicamente a TFG é estimada através de indicadores bioquímicos corriqueiros, sendo o mais indicado dentre eles o método que utiliza a creatinina sérica em conjunto às equações de predição (NGUYEN et al, 2015; CHENG et al, 2017). Segundo recomendação da *Kidney Disease Improving Global Outcomes* (KDIGO), a classificação da DRC deve ser determinada, preferencialmente, a partir da estimativa TFG e da relação albuminúria/creatininúria (LEVIN et al, 2013).

Com intuito de amenizar esse problema, pesquisadores e profissionais da saúde optaram por utilizar equações que estimassem a TFG. Essas equações utilizam o valor da creatinina sérica por ser excretada pelos rins, e posteriormente é associada a variáveis como gênero, idade, etnia, estatura e massa corporal, que influenciam diretamente a função renal. (LEVEY & ECKFELDT, 2017; INAGUMA et al, 2017).

Dentre as equações mais conhecidas estão a de Cockcroft-Gault, a MDRD (Modificação da Dieta em Doença Renal) e a mais amplamente utilizada e recomendada pela diretriz da KDIGO, chamada de Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI), que utiliza a dosagem da creatinina sérica, a idade, o gênero e a etnia do indivíduo. A equação CKD-EPI pode ser representada em uma única fórmula ou pode ser descrita de formas diferentes, por sofrer adaptações de acordo com gênero, nível de creatinina sérica e etnia, porém para melhor compreensão descrevemos neste estudo a fórmula geral, apresentada a seguir:

$$eTFG = 141 \times \min\left(\frac{Cre}{\kappa}, 1\right)^\alpha \times \max\left(\frac{Cre}{\kappa}, 1\right) - 1.209 \times 0,993 \text{ Idade} \times 1,018 [\text{se mulher}] \times 1,159 [\text{se negro}]^1$$

¹No qual: Cre: é creatinina sérica, κ : é 0,7 para mulheres e 0,9 para homens, α : é -0,329 para mulheres e -0,411 para homens, min: indica o mínimo de Cre/ κ ou 1; max: indica o máximo de Cre / κ ou 1 (LEVIN et al, 2013; CHANG et al, 2019).

3.4 Fatores de risco relacionados à DRC

3.4.1 Hipertensão arterial sistêmica (HAS)

No caso da HAS, especificamente, entende-se que é o maior fator de risco para a lesão renal e progressão da DRC, além de ser também uma consequência comum (TOWNSEND & TALER, 2015). A prevalência da HAS aumenta com a progressão de estágios da DRC, por exemplo no caso do estágio 5, o mais avançado, aproximadamente 86% dos pacientes são hipertensos (BIDANI & GRIFFIN, 2004; MUNTNER et al, 2010; TEDLA et al, 2011; INAGUMA et al, 2017).

Isso ocorre, pois, o controle inadequado da pressão arterial (PA) pode causar alterações vasculares devido ao aumento de substâncias vasoconstritoras e estresse mecânico, culminando na aceleração da perda da função renal, no caso da nefroesclerose hipertensiva maligna. Já se tratando da nefroesclerose benigna, ocorre uma perda progressiva da função renal em decorrência de outros fatores, como a hialinose, fibrose intimal concêntrica, enrugamento da membrana basal, atrofia de glomérulos (BROWN & WHITWORTH, 1992; ISSA et al, 2021).

A HAS, segundo a European Society of Cardiology, é definida por pressão arterial sistólica (PAS) ≥ 140 mmHg e/ou pressão diastólica (PAD) ≥ 90 mmHg, como pode-se observar no Quadro 2.

Tabela 1. Classificação da pressão arterial e definição da hipertensão arterial para indivíduos a partir de 18 anos.

Classificação	PAS (mmHg)		PAD (mmHg)
PA ótima	< 120	e	< 80
PA normal	120-129	e/ou	80-84
Pré-hipertensão	130-139	e/ou	85-89
HA Estágio 1	140-159	e/ou	90-99
HA Estágio 2	160-179	e/ou	100-109
HA Estágio 3	≥ 180	e/ou	≥ 110

Abreviações: PA pressão arterial; HA hipertensão arterial; mmHg milímetros de mercúrio. Fonte: adaptada de WILLIAMS et al, 2018.

3.4.2 Idade

A idade também apresenta-se como fator de risco para DRC, pois há um declínio biológico natural da TFG com o passar dos anos, chegando a ser observada uma diminuição de 8ml/min da TFG para cada década quando se trata de indivíduos com idade avançada, esta condição também está associada à alta prevalência de fatores de risco de ordem patológica que contribuem para o declínio da função renal (STEVENS et al, 2010; YOUNG et al, 2016; TUĞCU & ATAŞ, 2021).

Dentre esses fatores de risco destacam-se: Hipertensão arterial sistêmica que pode causar nefroesclerose hipertensiva; a diabetes que pode culminar em nefropatia diabética, as doenças cardiovasculares que estão entre as principais causas de óbitos entre renais crônicos, e obesidade que está relacionada à lipotoxicidade renal, hiperativação do sistema renina-angiotensina-aldosterona e resistência insulínica (STEVENS & LEVEY, 2005; WINEARLS & GLASSOCK, 2011; MALLAPPALLIL et al, 2014; JAROSZYNSKI et al, 2016; BAMBA et al, 2020).

Segundo estudo epidemiológico, que verificou a prevalência de múltiplas doenças crônicas, foi observado que a partir dos 65 anos a média de doenças chegava a 3.5 por indivíduo (SCHNEIDER et al, 2009). Ademais, o indivíduo com múltiplas morbidades apresenta maior probabilidade de prejuízo da saúde no geral (WOLFF et al, 2002).

3.4.3 Gênero

O gênero também demonstra ter um papel importante no risco para desenvolvimento da DRC, estudos prévios afirmam que para o gênero masculino a doença se apresenta de forma mais grave, com maior velocidade no declínio da função renal, maior risco de progressão para insuficiência renal crônica e mais alta taxa de mortalidade em comparação ao gênero feminino (NEUGARTEN et al, 2000; ERIKSEN & INGEBRETSEN, 2006; GOLDBERG et al, 2016). Dados recentes sugerem que a incidência de Insuficiência renal crônica (a fase mais avançada da DRC) em homens é maior, independentemente da faixa-etária (KRAMER et al, 2018; VALDIVIELSO et al, 2019).

Segundo estudos de Minutolo et al., dentre a população idosa com DRC em estágios moderado a avançado, concluiu-se que em homens o risco chega a ser 50% maior para a progressão da doença em comparação às mulheres, ou seja, o declínio

é bem mais acentuado para indivíduos do gênero masculino, independente do estágio em que se encontram, e mesmo para os acompanhados pelo médico nefrologista (MINUTOLO et al, 2020). Isso pode se dar pelo fato de que em homens há uma progressão mais acelerada da nefropatia diabética, doença renal policística autossômica dominante e a glomerulopatia membranosa em relação às mulheres (COGGINS et al, 1998), aliado ao fato de que homens apresentam hábitos alimentares menos saudáveis que as mulheres, o que gera efeitos negativos por exemplo no perfil glicêmico e na composição corporal que são conhecidos fatores de risco para DRC (CARRERO et al, 2018).

3.5 Aprendizado de Máquina

O AM é uma área da inteligência artificial que estuda o desenvolvimento de algoritmos capazes de aprender automaticamente com os dados que forem fornecidos. Tais técnicas baseiam-se em analisar conjuntos de dados, identificar padrões e criar soluções para tarefas requisitadas (DAS et al, 2017; MEHTA et al, 2019; ALDHYANI et al, 2020).

Pode-se dividir o AM em três categorias gerais: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço. No aprendizado supervisionado, os algoritmos são treinados com dados rotulados, objetivando aprender uma função específica, e comumente utilizam técnicas de classificação e regressão. Já o não-supervisionado, busca encontrar padrões em dados sem rótulos. O aprendizado por reforço, ocorre com a interação com um ambiente e altera seu comportamento para obter uma recompensa (MEHTA et al, 2019; SIDEY-GIBBONS et al, 2019).

Figura 2. Hierarquia do aprendizado de máquina.



Fonte: Adaptado de MONARD & BARANAUSKAS, 2003.

É importante citar que o aprendizado supervisionado é o mais utilizado para desenvolvimento desses aplicativos de AM (UDDIN et al, 2019). Os métodos supervisionados são treinados a partir de uma associação de dados de entrada com um desfecho já conhecido. Fazendo um paralelo com a medicina, os dados clínicos (exemplo: IMC, circunferência de cintura, PA, glicemia) de um indivíduo seriam os dados de entrada e a doença (exemplo: hipertensão, diabetes, câncer) seria o desfecho conhecido. A partir de então, após ser treinado, o algoritmo será capaz de fazer associações e previsões de desfechos com novos dados (SIDEY-GIBBONS et al, 2019).

O AM é composto por algoritmos de previsão, classificação e clustering, e através destes é possível descobrir padrões ocultos nos bancos de dados, para isso utilizam uma variedade de métodos de otimização, estatística e probabilidade, como por exemplo, árvore de decisão, support vector machine, regressão linear, k-vizinhos mais próximos e outros (MITCHELL, 2013; ALDHYANI et al, 2020).

Estes algoritmos possuem aplicação nas mais diversas áreas como, comércio, indústria, saúde e acadêmica, como exemplos práticos temos a detecção de fraudes de cartão de crédito, comportamento de consumo de um indivíduo, reconhecimento de fala, detecção de invasão de redes e rastreamento de doenças (SIDEY-GIBBONS

et al, 2019; JORDAN & MITCHELL, 2015). Atualmente, devido à evolução computacional, algorítmica e acesso facilitado a banco de dados, os aplicativos de AM alcançam resultados notáveis em técnicas de classificação e regressão (CILIBERTO et al, 2018)

Há tempos os algoritmos de AM vêm sendo utilizados como alternativa aos modelos limitados padrões, inclusive pela OMS, para auxiliar na detecção de diversas doenças, tais como diabetes, câncer, dengue, hepatites e doenças cardiovasculares, mostrando-se como uma ferramenta crucial para predição, e assim possibilitando um tratamento precoce e melhor qualidade dos cuidados com a saúde (OBERMEYER & EMANUEL, 2016; NITHYA & ILANGO, 2017).

Frente à realidade socioeconômica do Brasil, a população economicamente vulnerável encontra dificuldades de acesso a laboratórios de análises clínicas, com profissionais especializados e insumos apropriados. Torna-se inviável, em muitas ocasiões, a aplicação de métodos padrão ouro para fechar diagnósticos, e lançar mão de técnicas de AM possibilitará predições de doenças de forma prática, barata, rápida e com a exatidão necessária. Essa técnica poderá ser utilizada em locais sem acesso à internet, bastando o agente de saúde ter acesso a um aparelho digital como um *smartphone*, que possibilite o download do software, depois disso bastará fornecer os dados de entrada para que o algoritmo gere a classificação para diagnóstico (CHIU et al, 2012; BELL, 2020).

Dentre diversas técnicas de AM empregadas para desenvolver um algoritmo, a regressão linear múltipla (RLM) é uma opção largamente utilizada na elaboração de modelos computacionais, os quais conseguem avaliar situações em que é fundamental observar se há uma relação entre o desfecho e as variáveis independentes (SIDEY-GIBBONS et al, 2019; DUTTA & CHATTERJEE, 2020). A regressão é usada no AM para predizer algo, por exemplo, a expectativa de vida de um indivíduo ou a desvalorização de um automóvel. Para isso basta organizar o banco de dados em variáveis independentes e desfecho (SIDEY-GIBBONS et al, 2019).

A técnica de RLM utiliza o método de mínimos quadrados para reduzir o erro entre os resultados, e busca determinar uma estimativa do produto real através da equação:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \varepsilon$$

No qual, y representa a variável dependente, x_i a variável independente, β_0 é a constante de regressão, β_i indica o coeficiente de regressão, e ε é o fator de erro (AKAN et al, 2015).

É possível que um algoritmo incorra em um fenômeno chamado *overfitting*, que é quando o modelo consegue ter desempenho ótimo na predição da amostra de treino, porém é ineficiente na amostra de teste, pois o mesmo aprendeu as regras além do necessário e não consegue generalizar o modelo para outras amostras. No intuito de evitar isso, existem diversas técnicas a serem utilizadas, como *k-fold cross validation*, *train/test split*, e outras (SIDEY-GIBBONS et al, 2019).

4 MATERIAIS E MÉTODOS

4.1 Modelo do Estudo e População

Este estudo faz parte de um projeto maior denominado: "Predição da Doença Renal Crônica utilizando redes neurais artificiais". Concordearam em participar desse estudo 151 usuários dos serviços do Centro de Prevenção de Doenças Renais Crônicas, anexo do Hospital Universitários do Maranhão e seus acompanhantes, no período de março de 2018 a novembro de 2019, com faixa etária entre 18 e 60 anos, de ambos gêneros, na condição de hígidos ou renais crônicos. Porém, um total de 35 participantes foram excluídos da amostra final por não terem concluído as etapas necessárias, impossibilitando a verificação de variáveis importantes, como a creatinina sérica, PAS e PAD, portanto a amostra final deste estudo foi de 116 indivíduos.

4.2 Cálculo Amostral

A população estudada foi constituída por indivíduos atendidos em um centro de referência, segundo dados fornecidos pela instituição, N=600 pacientes. Para a realização do cálculo amostral, utilizou-se a equação proposta por Rea & Parker (2000). Considerou-se um intervalo de confiança de 95%, um erro absoluto amostral de 5,0%, e prevalência de DRC em adultos 10% (Chen et al., 2017) e adição de 10% para possíveis perdas ou recusas, de acordo com a equação:

$$n = (Z\alpha^2 \cdot P \cdot Q \cdot N) / (Z\alpha^2 \cdot P \cdot Q + (N - 1) \cdot E^2)^2.$$

Por fim, a amostra calculada foi de aproximadamente 112 pacientes. Somando-se 10% referente às possíveis perdas ou recusas, o tamanho mínimo da amostra corresponde a 123 pacientes.

²Onde: n – amostra calculada (número de pacientes); Z α – variável correspondente ao nível de confiança (1,96); P – prevalência dos fatores associados a causa (10%); Q – variável associada aos indivíduos que não pertencem à categoria de interesse (1 – P = 0,911); N – tamanho da população de referência (600 pacientes); E – erro amostral (5,0%).

4.3 Critérios de Inclusão e Não Inclusão

Inclusão: faixa etária entre 18 e 60 anos e ser capaz de responder individualmente e verbalmente as perguntas da anamnese aplicada pelos pesquisadores.

Não Inclusão: incapacidade física que impossibilitasse ou comprometesse as medidas antropométricas necessárias; gestantes ou lactantes; aqueles que possuíam comprometimento cognitivo causados por doenças como Alzheimer ou distúrbios mentais.

4.4 Instrumentos de Coleta e Avaliação dos Dados

4.4.1 Variáveis Sócio-demográficas

Foram coletadas, idade decimal: calculada pela diferença entre a data de nascimento e data da coleta de dados; gênero: autodeclarado pelo participante; gravidez: autodeclarada pela participante; nível de atividade física: autodeclarada; etnia/cor: autodeclarada com base no Instituto Brasileiro Geografia e Estatística.

4.4.2 Determinação dos Índices Antropométricos

As medidas antropométricas incluindo peso, altura foram obtidas utilizando técnicas padronizadas e por pesquisadores previamente treinados (Lohman et al., 1988). O peso foi aferido em balança eletrônica calibrada (Omron Healthcare®, HBF-214, Japan) com resolução de 0,1 kg, com o avaliado em posição ortostática, descalço e utilizando roupas leves. A estatura foi determinada através de estadiômetro transportável vertical com resolução de 0,1 cm (Seca® 213, Hamburg, Germany), com o avaliado descalço, em posição ortostática, com a cabeça, os glúteos e calcanhares em contato com a parede. O índice de massa corporal (IMC) foi obtido através da razão entre peso (Kg) e o quadrado da altura (m).

4.4.3 Perfil hemodinâmico

A mensuração e classificação da pressão arterial (PA) foi baseada no protocolo das Diretrizes Brasileiras de Hipertensão Arterial (ISSA et al, 2021), onde a medida da PA foi realizada pelo método indireto, com técnica auscultatória e com uso de esfigmomanômetro aneróide devidamente calibrado, foi explicado o procedimento ao paciente e este foi deixado em repouso por pelo menos cinco minutos em ambiente

calmo e confortável. O paciente foi instruído a não conversar durante a medida. Possíveis dúvidas foram esclarecidas antes e após o procedimento, o participante permaneceu na posição sentada, com as pernas descruzadas, pés apoiados no chão, dorso recostado na cadeira e relaxado. O braço deveria estar na altura do coração (nível do ponto médio do esterno ou quarto espaço intercostal), livre de roupas, apoiado, com a palma da mão voltada para cima e o cotovelo ligeiramente fletido.

4.4.4 Exames laboratoriais

Amostras de sangue foram coletadas para determinar a creatinina sérica (mg/dl) dos participantes, as quais foram analisadas no Laboratório de Bioquímica do Hospital Universitário do Maranhão.

4.5 Escolha das Características de Entrada do Algoritmo

A escolha das variáveis de entrada foi realizada considerando que as mesmas têm conhecida relação com o dano renal e diminuição da taxa de filtração glomerular, e também possuem baixo custo e fácil mensuração, as variáveis escolhidas estão demonstradas do Quadro 3 (TOWNSEND & TALER, 2015; GOLDBERG et al, 2016; YOUNG et al, 2016; CHENG et al, 2017; CHANG et al, 2019).

Tabela 2. Variáveis de entrada do modelo.

Variável	Abreviação	Unidade
Gênero	-	-
Idade	-	Anos
Pressão arterial sistólica	PAS	mmHg
Pressão arterial diastólica	PAD	mmHg

Abreviações: RLM-regressão linear múltipla; PAS-pressão arterial sistólica; PAD-pressão arterial.

4.6 Implementação do Método de Aprendizado de Máquina

Para desenvolvimento do modelo computacional proposto foi utilizado o software interativo de alta performance MATLAB®. O banco de dados foi dividido de forma randomizada, em 5 sub etapas ($k=5$), no intuito de evitar um *overfitting*, utilizou-se o teste *K-fold cross validation*, que divide igualmente (em *equal* ou *nearly equal k segments*) os dados utilizados em duas etapas: treinamento e teste. Divididas em "folds" tanto a fase de treinamento como a de teste foram cumpridas através das interações k . Em cada interação, deixou-se um "fold" para teste e treinou-se o modelo

nos $k-1$ folds restantes. Portanto, do total de participantes, 80% foram escolhidos aleatoriamente para compor o grupo da etapa de treinamento e 20% foram para a etapa de teste.

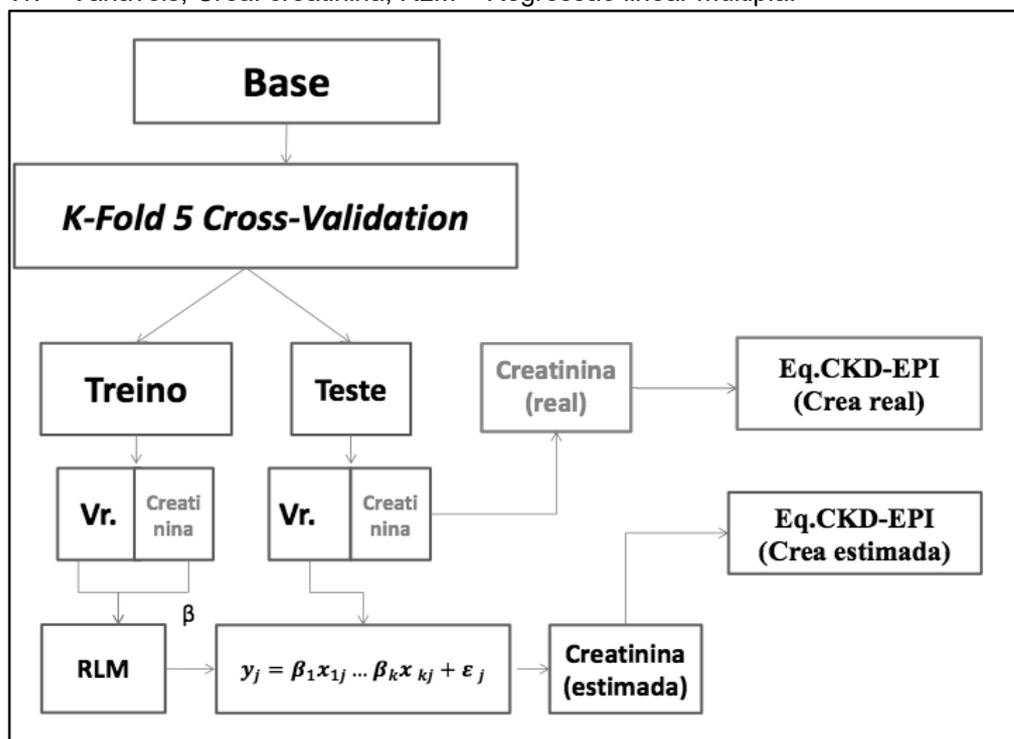
Em nosso modelo foi empregado o método de aprendizagem supervisionada de RLM para determinar quais coeficientes β e significância das variáveis escolhidas (CHANG et al, 2015; LOPES et al, 2019). A partir disso, as variáveis utilizadas para entrada do modelo foram determinadas (a idade, o gênero, a PAS e PAD), e variável de saída foi creatinina.

Em seguida, após estimar o valor da creatinina, aplicou-se à equação:

$$\begin{aligned} CKD - EPI: eTFG \\ &= 141 \times \min(Cre/\kappa, 1)^\alpha \times \max(Cre/\kappa, 1) \\ &- 1.209 \times 0,993 \text{ Idade} \times 1,018 [\text{se mulher}] \times 1,159 [\text{se negro}] \end{aligned}$$

Na qual se ajusta o valor da creatinina com os requisitos necessários para cada indivíduo, estimando assim a TFG, conforme está demonstrado no fluxograma do modelo representado na Figura 3.

Figura 3. Fluxograma do modelo proposto: y_j é o valor estimado j - ésimo pacientes $j = 1, \dots, n$ (n número de pacientes da base de teste); x_{1j}, \dots, x_{kj} são os valores das k variáveis independentes para o j - ésimo paciente da base de teste; β_j é um parâmetro que mede a contribuição da i - ésima variável independente x_i no modelo; $i = 1, \dots, k$ encontrada a partir do treino; ε_j erro (desvio) aleatório; Vr. - Variáveis; Crea: creatinina; RLM - Regressão linear múltipla.



Fonte: Autora.

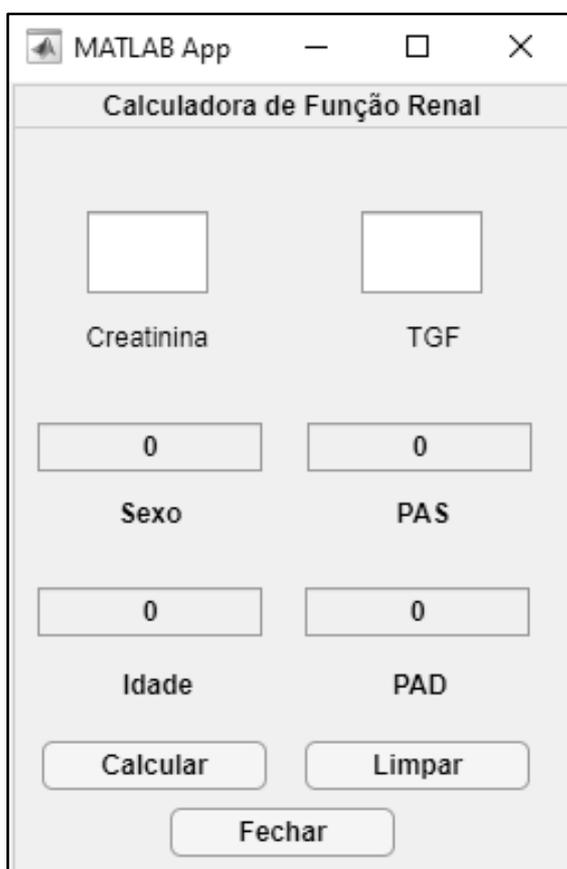
5 RESULTADOS

5.1 Desenvolvimento do software

Com o objetivo de facilitar e popularizar a utilização do modelo matemático proposto neste estudo, que estima a creatinina sérica e a taxa de filtração glomerular, desenvolveu-se um programa de computador capaz de interagir de maneira simples com o usuário.

A interface (disponível para visualização na Figura 4) é de fácil utilização, basta acrescentar as variáveis de entrada (PAS, PAD, gênero e idade) nos campos indicados, em seguida pressionar o botão "calcular" para obter o resultado da creatinina e da taxa de filtração glomerular estimadas, que ficarão disponíveis em seus respectivos campos, e caso o usuário deseje apagar os dados preenchidos e fazer nova análise basta clicar no botão "limpar", por fim caso deseje sair do software basta clicar no botão "fechar". Para construção desse software foi utilizada a ferramenta GUIDE (*Graphical User Interface Development Environment*) do MATLAB®.

Figura 4. Layout de interface do software, visualizada pelo usuário. Abreviações: TFG - taxa de filtração glomerular; PAS-pressão arterial sistólica; PAD-pressão arterial.



Fonte: Autora

5.2 Análise estatística

Para o tratamento estatístico dos dados foi utilizado o software SPSS (*Statistical Package for the Social Sciences, Inc., Chicago, IL, USA*) versão 25.0, para testar a normalidade dos dados utilizou-se o teste de *Kolmogorov-Smirnov*. A estatística descritiva da amostra estudada, foi apresentada em média e desvio padrão para ambos os grupos (hígidos e DRC). Para avaliar o desempenho do modelo proposto foi utilizada a correlação de *Pearson*, onde se comparou os valores estimados de creatinina e TFG com os seus valores reais.

5.3 Aspectos éticos

O presente estudo possui aprovação do Comitê de ética da Universidade Federal do Maranhão com CAEE nº 67030517.5.0000.5087 (Anexo A). Os voluntários só foram incluídos após serem informados, em detalhes, sobre o estudo e preencherem corretamente o termo de Consentimento Livre e Esclarecido (Apêndice A). Ficou garantida a confidencialidade dos dados e a qualquer momento o voluntário poderia desistir de participar do estudo.

A tabela 1 apresenta as características sociodemográficas, antropométricas, laboratoriais e hemodinâmicas dos 116 participantes do estudo, com valores demonstrados em média e desvio padrão.

Através da análise por regressão linear múltipla, observou-se que há uma associação significativa entre as variáveis escolhidas para entrada do modelo e a creatinina sérica (real), conforme demonstrado na tabela 3.

Tabela 3. Características Sócio-demográficas, antropométricas, laboratoriais e hemodinâmicas da amostra, hígidos e renais crônicos.

Características	Hígidos (n=92)	DRC (n=24)	Total (n=116)
Gênero			
<i>Feminino</i>	65	15	80
<i>Masculino</i>	27	9	36
Etnia			
<i>Branços</i>	33	2	35
<i>Pretos e pardos</i>	27	22	49
Idade (anos)	39,9± 15,5	53±13,3	40,9±14,6
Estatura (cm)	1,57±0,08	1,58±0,08	1,57±0,08
Massa corporal (kg)	65,8±14,19	70,9±13,5	85,0±14,1
IMC (kg/m ²)	26,4±5,2	27,9±4,6	27,15±5,2
CC (cm)	83,6±13,4	91±12,1	85,0±13,0
PAS (mmHg)	118,8±21,2	131,3±24,9	120,4±21,4
PAD (mmHg)	75,7±11,8	79,9±12,6	77,4±12,1
Creatinina sérica (mg/dl)	0,73±0,21	1,37±1,15	0,79±0,26
TFG (ml/min/1,73m ²)	109,4±17,5	62,7±22,2	102,4±23,1

Abreviações: PAS-pressão arterial sistólica; PAD-pressão arterial diastólica; CC - circunferência de cintura; IMC - Índice de massa corpórea; TFG - taxa de filtração glomerular.

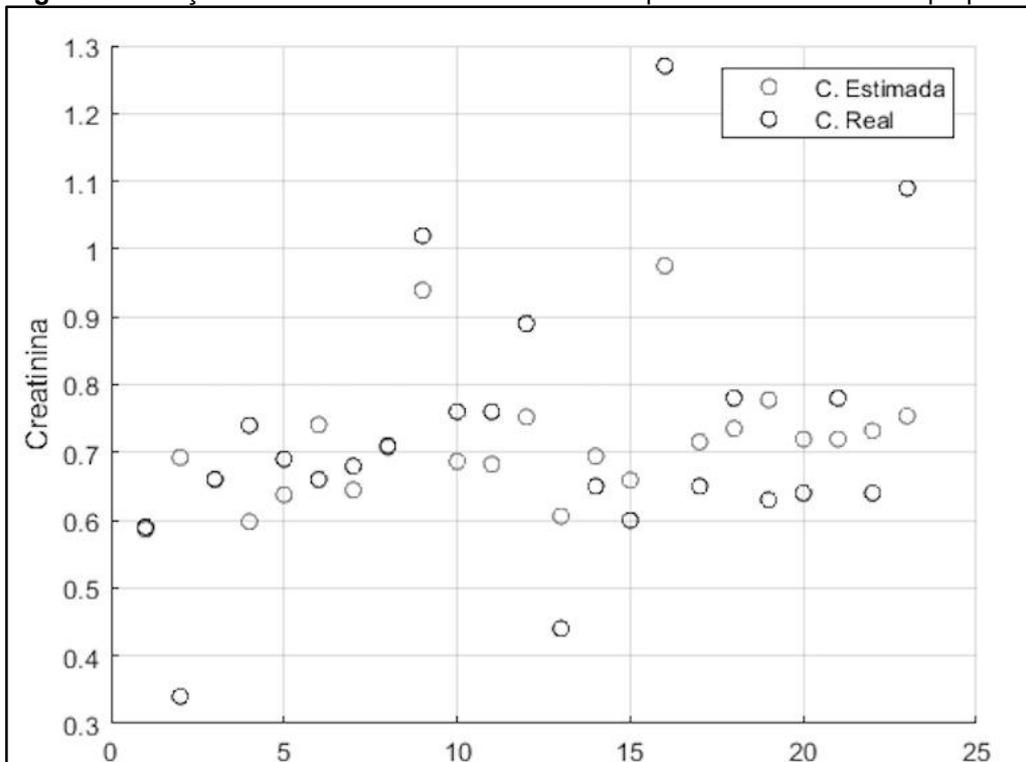
É possível verificar na Figura 5, que o modelo proposto foi eficiente em prever a creatinina, visto que ao comparar os valores da creatinina estimada aos valores de creatinina real observou-se uma forte correlação ($r=0,72$ e $DP=0,14$). Também, verificou-se uma correlação importante ($r=0,87$ e $DP=13,2$) entre TFG calculada baseando-se na creatinina estimada pelo modelo proposto e a TFG real, reforçando o bom desempenho do algoritmo, conforme apresentado na Figura 6.

Tabela 4. Resultado da RLM, coeficiente (β) e significância (p) para cada variável do modelo.

Variável	β (beta)	p -valor
Intercept	0,9859	0,000
Gênero	-0,24513	0,000
Idade	-0,00252	0,002
PAS	-0,00312	0,000
PAD	0,00536	0,000

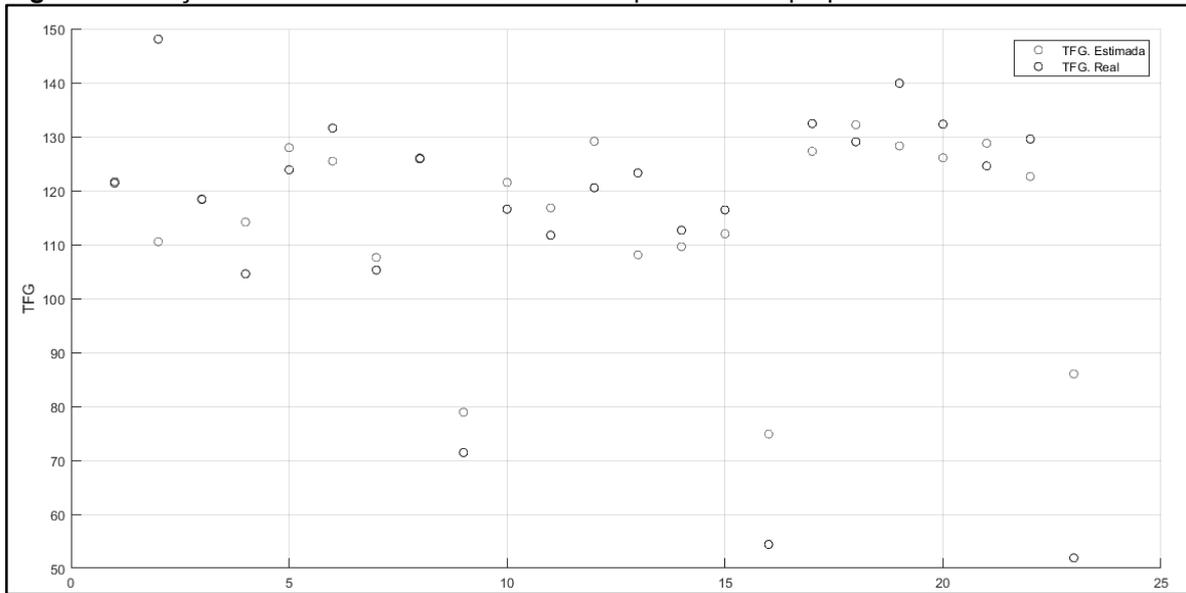
Abreviações: RLM - regressão linear múltipla; PAS - pressão arterial sistólica; PAD - pressão arterial diastólica; significância estatística p -valor $< 0,05$.

Figura 5. Relação da creatinina real com a estimada pelo modelo estatístico proposto.



Abreviações: C - creatinina.

Figura 6. Relação entre TFG real e a TFG estimada pelo método proposto.



Abreviações: TFG - taxa de filtração glomerular.

6 DISCUSSÃO

A creatinina sérica é um marcador bioquímico essencial para determinar a função renal e importante aliada para triagem e prevenção da DRC, porém não deve ser considerada de forma isolada. Por esse motivo, é vinculada à equações de estimativa da TFG, como a Cockcroft-Gault, MDRD e a CKD-EPI que ajustam para idade, gênero, superfície corpórea e etnia (HERGET-ROSENTHAL et al, 2007; LEVIN et al, 2013). A creatinina sérica estimada pelo modelo proposto neste estudo, demonstrou uma correlação forte com a creatinina sérica real, o mesmo ocorreu para TFG estimada, que obteve melhor desempenho, devido aos fatores de correção da equação de estimativa utilizada.

No cenário atual, os métodos para estimar a creatinina em sua maioria são invasivos (dependem de amostras sanguíneas ou urinárias) o que causa desconforto para o paciente, não oferecem praticidade e também necessitam de laboratórios de análises bioquímicas com aparatos tecnológicos caros e mão-de-obra especializada (PEAKE & WHITING, 2006). Até o momento, não existem modelos computacionais para prever a creatinina sérica a partir de variáveis não invasivas, o que existem são modelos que estimam a TFG, porém, sempre baseados em indicadores bioquímicos, como por exemplo a creatinina sérica e/ou cistatina C.

No modelo proposto neste estudo, utilizamos variáveis de simples aferição, e que já estão inseridas na prática clínica, com o intuito de desenvolver uma ferramenta computacional capaz de estimar a creatinina sérica de forma imediata, e por conseguinte determinar a TFG, buscando auxiliar na predição e diagnóstico da DRC. As variáveis de entrada do nosso modelo foram: gênero, idade, PAS e PAD, os quais são conhecidos fatores de risco para o prejuízo da função renal, todas demonstraram desempenho matemático satisfatório quando aplicada a técnica de RLM (TOWNSEND & TALER, 2015; INAGUMA et al, 2017; VALDIVIELSO et al, 2019; TUĞCU & ATAŞ, 2021).

A PAS e a PAD, especificamente, ainda não são utilizadas por outros métodos de determinação da creatinina sérica, possivelmente pois os mesmos possuem foco em variáveis bioquímicas, no entanto foram escolhidas para este modelo matemático por serem essenciais para diagnóstico da HAS, doença que consta na literatura como a principal causa da DRC, juntamente com a diabetes mellitus (BAKRIS et al, 2020).

Essa forte relação da HAS com a DRC pode ser explicada pelos mecanismos hormonais envolvidos no controle de longo prazo da pressão arterial, que tem papel fundamental dos rins, a desregulação do sistema renina-angiotensina-aldosterona culmina em níveis exacerbados de angiotensina II, diminuição na produção de substâncias vasodilatadoras, e causa desequilíbrio do hormônio antidiurético aumentando os níveis de sódio no sangue. Tal desequilíbrio gera um aumento na pressão intraglomerular causando danos à estrutura e mesmo na função dos rins, sendo uma das consequências a elevação da creatinina sérica (ZELLER et al, 2007; HAMRAHIAN et al, 2016).

Outro fator importante em nosso modelo é a variável idade, sabe-se que com o seu avanço há diminuição da TFG, isso se dá por alguns fatores que o envelhecimento está atrelado, como por exemplo a redução de fluxo sanguíneo para os rins, fator inerente à progressão da idade, também pode-se apontar a polifarmácia em decorrência do desenvolvimento de doenças como HAS, dislipidemia, obesidade (MALTA et al, 2019).

Assim como as outras variáveis de entrada, o gênero contribui para a predição da creatinina de forma mais fidedigna, visto que há uma conhecida diferença dos parâmetros da creatinina para homens e mulheres, assim como a evolução da DRC entre ambos. Apesar da literatura ainda possuir divergências sobre a questão do gênero e a DRC, ainda é mais aceito que em homens a doença parece evoluir mais rápido (GRAMS et al, 2017; NEUGARTEN et al 2019). Em seus estudos Kang et al, explicam essa vantagem do gênero feminino devido à influência do estrogênio endógeno especificamente, que demonstra ser um fator de proteção para a DRC em mulheres, concluíram que quanto maior a fase reprodutiva feminina (maior exposição ao estrogênio) menor os riscos de desenvolver dano renal (KANG et al, 2020)

Atualmente, outras técnicas de determinar creatinina têm sido desenvolvidas, e uma delas é a coleta da creatinina na saliva, que é um método menos invasivo que a coleta sanguínea e menos dispendioso que os tradicionais (VENKATAPATHY et al, 2014; TEMILOLA et al, 2019), no entanto, o modelo utilizado neste estudo mostra-se mais vantajoso, pois estima a creatinina sérica utilizando unicamente variáveis não invasivas simples, de fácil coleta e baixo custo, mediante aplicações de técnicas de AM.

No presente estudo, obteve-se uma correlação forte ($r=0,72$) entre creatinina estimada e a real. Cabe ressaltar que no estudo de Venkatapathy et al, que utilizou a creatinina salivar para estimar a creatinina sérica, e foi encontrada correlação semelhante de $r=0,73$, o que reforça a vantagem do nosso modelo, que alcançou resultado equiparado a aquele com uso de método que exige análise laboratorial da amostra coletada (VENKATAPATHY et al, 2014)

Em nosso estudo a TFG obteve correlação de $r=0,87$, vale dizer que em estudo recente com 230 indivíduos renais crônicos (estágios 1 - 5), que utilizou a creatinina salivar para determinar a TFG, obteve uma correlação de $r=0,82$ com a TFG real (TEMILOLA et al, 2019), demonstrando que o modelo proposto aqui atingiu melhor desempenho.

No que diz respeito à predição da TFG, Taylor et al, analisaram métodos já utilizados na literatura, como as equações de Cockcroft-Gault e a MDRD, e verificaram que os mesmos apresentaram correlação de $r=0,74$ e $r=0,87$, respectivamente, em comparação à TFG medida por iohexol, (TAYLOR et al, 2006), observou-se que a correlação do nosso modelo foi mais importante que a da primeira equação, e equivalente à segunda, a qual está entre as principais e mais recomendadas pela KDIGO para determinar TFG, juntamente à CKD-EPI, que foi a utilizada para testar o desempenho do nosso modelo.

Por conseguinte, outro estudo com 10 indivíduos, que desenvolveu um método para estimar a TFG, utilizou como variáveis para sua equação a massa muscular e a creatinina, a primeira obtida por meio de DEXA (*Dual-energy X-ray Absorptiometry*) e a última na amostra de sangue e análise bioquímica, alcançou um coeficiente de correlação de $r=0,97$ ao ser comparado aos valores encontrados através de indicadores bioquímicos com alta especificidade (TAYLOR et al, 2006). Apesar do modelo aqui proposto alcançar uma correlação de $r=0,87$, ainda é vantajoso nos aspectos clínico e econômico, visto que o método do referido estudo utiliza a creatinina sérica real, e adiciona a massa magra (medida direta) para ajuste, exames que tem custo elevado e não estão amplamente disponíveis, o que não torna viável para população economicamente vulnerável.

A dificuldade em obter indicadores com boa especificidade, como alguns marcadores bioquímicos e exames de imagem, é um fator preocupante, uma vez que a demora em estabelecer um diagnóstico de DRC pode ser determinante para um pior

prognóstico. Deste modo, o modelo proposto em nosso estudo, apresenta-se como uma ferramenta relevante para auxiliar no diagnóstico precoce, ao demonstrar uma forte correlação com o método validado para a prática clínica, tanto para estimar a creatinina quanto a TFG (aplicando o valor de creatinina sérica estimada à equação CKD-EPI), utilizando tão somente variáveis de simples coleta, não invasivas e de baixo custo, rotineiras na prática clínica.

A RLM é uma técnica amplamente utilizada em diversas áreas do conhecimento e propriamente na medicina, como uma ferramenta capaz de auxiliar diagnósticos de saúde, facilitando e agilizando o acesso aos resultados. Tal ferramenta se propõe a oferecer as informações de forma mais célere e com a precisão necessária para ser válida e ter sua utilidade na promoção de saúde da população.

Neste estudo, o modelo matemático proposto poderá ser amplamente utilizado em consultórios médicos da rede SUS como forma de iniciar mais precocemente o tratamento de saúde, já que neste âmbito os usuários dispõem de menor acesso à exames tanto pela demora em conseguir realizá-los em decorrência da alta demanda populacional, quanto pela escassez de recursos (seja de insumos ou de mão de obra). Também, na atenção básica, através de estratégias de saúde da família (programa implantado no Brasil desde 1994), também será possível lançar mão desta ferramenta para auxílio no diagnóstico e até mesmo em estudos epidemiológicos com o objetivo de determinar características populacionais relacionadas à saúde.

7 CONCLUSÃO

O modelo computacional implementado neste estudo foi eficiente em estimar a creatinina sérica, demonstrando forte correlação entre o valor estimado e o real para esta variável. O mesmo ocorreu para a TFG calculada a partir da creatinina estimada pelo nosso modelo, demonstrou uma alta correlação com a TFG calculada pela creatinina real. Entende-se, portanto, que o software desenvolvido neste estudo foi capaz de prever a função renal, e pode ser um grande aliado para medidas de prevenção e tratamento precoce da população economicamente vulnerável, auxiliando também na diminuição dos gastos públicos com o tratamento tardio da DRC.

REFERÊNCIAS

AKAN, Recep; KESKIN, S. Nilay; UZUNDURUKAN, Soner. Multiple regression model for the prediction of unconfined compressive strength of jet grout columns. **Procedia Earth and Planetary Science**, v. 15, p. 299-303, 2015.

ALDHYANI, Theyazn HH; ALSHEBAMI, Ali Saleh; ALZHRANI, Mohammed Y. Soft Clustering for Enhancing the Diagnosis of Chronic Diseases over Machine Learning Algorithms. *Journal of healthcare engineering*, v. 2020, 2020.

BAKRIS, George; YANG, Y. Fred; PITT, Bertram. Mineralocorticoid receptor antagonists for hypertension management in advanced chronic kidney disease: BLOCK-CKD trial. **Hypertension**, v. 76, n. 1, p. 144-149, 2020.

BAMBA, Ryo et al. The visceral adiposity index is a predictor of incident chronic kidney disease: a population-based longitudinal study. **Kidney and Blood Pressure Research**, v. 45, n. 3, p. 407-418, 2020.

BASTOS MG, KIRSZTAJN GM. Chronic kidney disease: importance of early diagnosis, immediate referral and structured interdisciplinary approach to improve outcomes in patients not yet on dialysis. *J Bras Neurol*. 2011;33(1):93-108. PMID:21541469.

BELL, J. **Machine Learning: Hands-on for developers and technical professionals**. [S.l.]: John Wiley & Sons, 2014. P 45.

BELL, Jason. **Machine learning: hands-on for developers and technical professionals**. John Wiley & Sons, 2020

BIDANI, Anil K.; GRIFFIN, Karen A. Pathophysiology of hypertensive renal damage: implications for therapy. **Hypertension**, v. 44, n. 5, p. 595-601, 2004.

BIKIA, Vasiliki et al. Noninvasive estimation of aortic hemodynamics and cardiac contractility using machine learning. *Scientific Reports*, v. 10, n. 1, p. 1-17, 2020.

BRADSHAW, Christina et al. Early detection of chronic kidney disease in low-income and middle-income countries: development and validation of a point-of-care screening strategy for India. **BMJ global health**, v. 4, n. 5, p. e001644, 2019.

BROWN, Mark A.; WHITWORTH, Judith A. Hypertension in human renal disease. **Journal of hypertension**, v. 10, n. 8, p. 2, 1992.

BURTON, James O. et al. Association of anthropometric obesity measures with chronic kidney disease risk in a non-diabetic patient population. **Nephrology Dialysis Transplantation**, v. 27, n. 5, p. 1860-1866, 2012.

CARRERO, Juan Jesus et al. Sex and gender disparities in the epidemiology and outcomes of chronic kidney disease. *Nature Reviews Nephrology*, v. 14, n. 3, p. 151, 2018.

CHANG, Fengming; CHEN, Hsing-Chung; LIU, Hsiang-Chuan. Double k-folds in SVM. In: **2015 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing**. IEEE, 2015. p. 384-387.

CHANG, Hsueh-Lu et al. A predictive model for progression of CKD. **Medicine**, v. 98, n. 26, 2019.

CHEN, N. A. N. et al. Challenging chronic kidney disease: experience from chronic kidney disease prevention programs in Shanghai, Japan, Taiwan and Australia. **Nephrology**, v. 15, p. 31-36, 2010. LIU, Ling et al. Waist height ratio predicts chronic kidney disease: a systematic review and meta-analysis, 1998–2019. **Archives of Public Health**, v. 77, n. 1, p. 55, 2019.

CHEN, Teresa K.; KNICELY, Daphne H.; GRAMS, Morgan E. Chronic kidney disease diagnosis and management: a review. **Jama**, v. 322, n. 13, p. 1294-1304, 2019.

CHENG, Li-Chen; HU, Ya-Han; CHIOU, Shr-Han. Applying the temporal abstraction technique to the prediction of chronic kidney disease progression. **Journal of medical systems**, v. 41, n. 5, p. 85, 2017.

CHIU, RK; CHEN, RY; WANG, SA; JIAN, SJ. **Intelligent systems on the cloud for the early detection of chronic kidney disease**. In Machine Learning and Cybernetics (ICMLC), 2012 International Conference on Vol. 5, pp. 1737-1742. IEEE. 2012.

CILIBERTO, Carlo et al. Quantum machine learning: a classical perspective. **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 474, n. 2209, p. 20170551, 2018.

COGGINS CH, BREYER Lewis J, CAGGIULA AW, CASTALDO LS, KLAHR S, WANG SR. Differences between women and men with chronic renal disease. **Nephrol Dial Transplant**. 1998;13(6):1430-1437

DAS, Kajaree; BEHERA, Rabi Narayan. A survey on machine learning: concept, algorithms and applications. **International Journal of Innovative Research in Computer and Communication Engineering**, v. 5, n. 2, p. 1301-1309, 2017.

DE BOER, Ian H. et al. Obesity and change in estimated GFR among older adults. **American Journal of Kidney Diseases**, v. 54, n. 6, p. 1043-1051, 2009.

DELANAYE, Pierre; CAVALIER, Etienne; POTTEL, Hans. Serum creatinine: not so simple!. **Nephron**, v. 136, n. 4, p. 302-308, 2017.

DUTTA, Anirban; CHATTERJEE, Biswapati. Development of prediction model through linear multiple regression for the prediction and analysis of the GSM of embroidered fabric. **Research Journal of Textile and Apparel**, 2020.

ERIKSEN, B. O.; INGEBRETSEN, O. C. The progression of chronic kidney disease: a 10-year population-based study of the effects of gender and age. **Kidney international**, v. 69, n. 2, p. 375-382, 2006.

EVANS, Philip D. et al. Anthropomorphic measurements that include central fat distribution are more closely related with key risk factors than BMI in CKD stage 3. **PLoS One**, v. 7, n. 4, p. e34699, 2012.

GO, Alan S. et al. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *New England Journal of Medicine*, v. 351, n. 13, p. 1296-1305, 2004.

GOLDBERG, Idam et al. The role of gender in chronic kidney disease. **Emj**, v. 1, n. 2, p. 58-64, 2016.

GRAMS, Morgan E. et al. Risks of adverse events in advanced CKD: the chronic renal insufficiency cohort (CRIC) study. **American Journal of Kidney Diseases**, v. 70, n. 3, p. 337-346, 2017.

HAMRAHIAN, Seyed Mehrdad; FALKNER, Bonita. Hypertension in chronic kidney disease. *Hypertension: from basic research to clinical practice*, p. 307-325, 2016.

HERGET-ROSENTHAL, Stefan; BÖKENKAMP, Arend; HOFMANN, Walter. How to estimate GFR-serum creatinine, serum cystatin C or equations?. **Clinical biochemistry**, v. 40, n. 3-4, p. 153-161, 2007.

HIMMELFARB, Jonathan et al. The current and future landscape of dialysis. **Nature Reviews Nephrology**, v. 16, n. 10, p. 573-585, 2020.

HUIDOBRO, Juan Pablo; TAGLE, Rodrigo; GUZMÁN, Ana María. Creatinina y su uso para la estimación de la velocidad de filtración glomerular. **Revista médica de Chile**, v. 146, n. 3, p. 344-350, 2018.

INAGUMA, Daijo et al. Risk factors for CKD progression in Japanese patients: findings from the Chronic Kidney Disease Japan Cohort (CKD-JAC) study. **Clinical and experimental nephrology**, v. 21, n. 3, p. 446-456, 2017.

ISSA, Aurora Felice Castro et al. Diretrizes Brasileiras de Hipertensão Arterial–2020. **Arq Bras Cardiol**, v. 116, n. 3, p. 516-658, 2021

ISSA, Aurora Felice Castro et al. Diretrizes Brasileiras de Hipertensão Arterial–2020. **Arq Bras Cardiol**, v. 116, n. 3, p. 516-658, 2021.

JAROSZYNSKI, Andrzej et al. Association of anthropometric measures of obesity and chronic kidney disease in elderly women. **Ann Agric Environ Med**, v. 23, n. 4, p. 636-640, 2016.

JORDAN, Michael I.; MITCHELL, Tom M. Machine learning: Trends, perspectives, and prospects. **Science**, v. 349, n. 6245, p. 255-260, 2015.

KANG, Shin Chan et al. Association of reproductive lifespan duration and chronic kidney disease in postmenopausal women. In: **Mayo Clinic Proceedings**. Elsevier, 2020. p. 2621-2632.

Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO clinical practice guideline for the evaluation and management of chronic kidney disease. **Kidney Inter Suppl** 2013; 3: 1-150.

KRAMER, Anneke et al. The european renal association–european dialysis and transplant association (ERA-EDTA) registry annual report 2015: a summary. **Clinical kidney journal**, v. 11, n. 1, p. 108-122, 2018.

LEVEY, Andrew S. et al. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. **American Journal of Kidney Diseases**, v. 39, n. 2 SUPPL. 1, 2002

LEVEY, Andrew S.; ECKFELDT, John H. Estimating glomerular filtration rate using serum creatinine. **Clinical chemistry**, v. 63, n. 6, p. 1161-1162, 2017.

LEVIN, Adeera et al. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. **Kidney international supplements**, v. 3, n. 1, p. 1-150, 2013

LOPES, Marcus Vinicius et al. Attesting compliance of biodiesel quality using composition data and classification methods. **Neural Computing and Applications**, v. 31, n. 2, p. 539-551, 2019.

LV, Ji-Cheng; ZHANG, Lu-Xia. Prevalence and disease burden of chronic kidney disease. **Renal Fibrosis: Mechanisms and Therapies**, p. 3-15, 2019.

MALLAPPALLIL, Mary et al. Chronic kidney disease in the elderly: evaluation and management. **Clinical practice (London, England)**, v. 11, n. 5, p. 525, 2014.

MALTA, Deborah Carvalho et al. Avaliação da função renal na população adulta brasileira, segundo critérios laboratoriais da Pesquisa Nacional de Saúde. **Revista Brasileira de Epidemiologia**, v. 22, 2019.

MARINHO, Ana Wanda Guerra Barreto et al. Prevalência de doença renal crônica em adultos no Brasil: revisão sistemática da literatura. **Cadernos Saúde Coletiva**, v. 25, n. 3, p. 379-388, 2017.

MATHUR, Pankaj et al. Artificial Intelligence, Machine Learning, and Cardiovascular Disease. **Clinical Medicine Insights: Cardiology**, v. 14, p. 1179546820927404, 2020

MCMANUS, M. Sue; WYNTER-MINOTT, Shushanne. Guidelines for chronic kidney disease: Defining, staging, and managing in primary care. **The Journal for Nurse Practitioners**, v. 13, n. 6, p. 400-410, 2017.

MEHTA, Pankaj et al. A high-bias, low-variance introduction to machine learning for physicists. **Physics reports**, v. 810, p. 1-124, 2019.

MINUTOLO, Roberto et al. Sex Differences in the Progression of CKD Among Older Patients: Pooled Analysis of 4 Cohort Studies. **American Journal of Kidney Diseases**, v. 75, n. 1, p. 30-38, 2020.

MITCHELL, R.; MICHALSKI, J.; CARBONELL, T. **An artificial intelligence approach**. Berlin: Springer, 2013.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, v. 1, n. 1, p. 32, 2003.

MUNTNER, Paul et al. Hypertension awareness, treatment, and control in adults with CKD: results from the Chronic Renal Insufficiency Cohort (CRIC) Study. **American Journal of Kidney Diseases**, v. 55, n. 3, p. 441-451, 2010.

MURTON, Molly et al. Burden of chronic kidney Disease by KDIGO categories of glomerular filtration rate and albuminuria: a systematic review. **Advances in therapy**, p. 1-21, 2020.

NEVES, Precil Diego Miranda de Menezes et al. Censo Brasileiro de Diálise: análise de dados da década 2009-2018. **Brazilian Journal of Nephrology**, v. 42, p. 191-200, 2020.

NEUGARTEN, Joel; ACHARYA, Anjali; SILBIGER, Sharon R. Effect of gender on the progression of nondiabetic renal disease: a meta-analysis. **Journal of the American Society of Nephrology**, v. 11, n. 2, p. 319-329, 2000.

NEUGARTEN, Joel; GOLESTANEH, Ladan. Influence of sex on the progression of chronic kidney disease. In: **Mayo Clinic Proceedings**. Elsevier, 2019. p. 1339-1356.

NGUYEN, Minh T. et al. Estimating glomerular filtration rate in obese subjects. **Obesity Research & Clinical Practice**, v. 9, n. 2, p. 152-157, 2015.

NITHYA, B.; ILANGO, V. Predictive analytics in health care using machine learning tools and techniques. In: **2017 International Conference on Intelligent Computing and Control Systems (ICICCS)**. IEEE, 2017. p. 492-499.

OBERMEYER, Ziad; EMANUEL, Ezekiel J. Predicting the future—big data, machine learning, and clinical medicine. **The New England journal of medicine**, v. 375, n. 13, p. 1216, 2016.

PASALA, Swetha; CARMODY, J. Bryan. How to use... serum creatinine, cystatin C and GFR. **Archives of Disease in Childhood-Education and Practice**, v. 102, n. 1, p. 37-43, 2017.

PEAKE, Michael; WHITING, Malcolm. Measurement of serum creatinine—current status and future goals. **Clinical biochemist reviews**, v. 27, n. 4, p. 173, 2006.

PECOITS-FILHO, Roberto. Diagnóstico de doença renal crônica: avaliação da função renal. **J Bras Nefrol**, v. 26, n. 3, p. 4-5, 2004.

PIROJSAKUL, Kwanchai; MATHEWS, Nisha; SEIKALY, Mouin G. Chronic kidney disease in children: recent update. **The Open Urology & Nephrology Journal**, v. 8, n. 1, 2015.

PRATES, Aline Bodanese et al. Avaliação da filtração glomerular através da medida da cistatina C sérica. **Jornal brasileiro de nefrologia= Brazilian journal of nephrology. São Paulo. Vol. 29, n. 1 (2007), p. 48-55, 2007.**

RAI, Pradeep Kumar et al. Chronic Kidney Disease among Middle-Aged and Elderly Population: A cross-sectional screening in a Hospital Camp in Varanasi, India. **Saudi Journal of Kidney Diseases and Transplantation**, v. 30, n. 4, p. 795, 2019.].

REA, L. M.; PARKER, R. A. Desenvolvendo perguntas para pesquisas. _____. **Metodologia de pesquisa: do planejamento à execução. São Paulo: Pioneira**, p. 57-75, 2000.

ROMAGNANI, Paola et al. Chronic kidney disease. *Nature reviews Disease primers*, v. 3, n. 1, p. 1-24, 2017.

SCHNEIDER, Kathleen M.; O'DONNELL, Brian E.; DEAN, Debbie. Prevalence of multiple chronic conditions in the United States' Medicare population. **Health and quality of life outcomes**, v. 7, n. 1, p. 1-11, 2009.

SIDEY-GIBBONS, Jenni AM; SIDEY-GIBBONS, Chris J. Machine learning in medicine: a practical introduction. **BMC medical research methodology**, v. 19, n. 1, p. 1-18, 2019.

SILVA, Maria Inês Barreto et al. Waist-to-height ratio: an accurate anthropometric index of abdominal adiposity and a predictor of high HOMA-IR values in nondialyzed chronic kidney disease patients. **Nutrition**, v. 30, n. 3, p. 279-285, 2014.

STEVENS, Lesley A.; LEVEY, Andrew S. Chronic kidney disease in the elderly—how to assess risk. 2005.

STEVENS, Lesley A.; VISWANATHAN, Gautham; WEINER, Daniel E. Chronic kidney disease and end-stage renal disease in the elderly population: current prevalence, future projections, and clinical significance. **Advances in chronic kidney disease**, v. 17, n. 4, p. 293-301, 2010.

[Szwarcwald, Célia Landmann et al. Valores de referência para exames laboratoriais de colesterol, hemoglobina glicosilada e creatinina da população adulta brasileira. *Revista Brasileira de Epidemiologia* [online]. 2019, v. 22, n. Supl 02 [Acessado 9

Setembro 2021] , e190002.supl.2. Disponível em: <<https://doi.org/10.1590/1980-549720190002.supl.2>>. Epub 07 Out 2019. ISSN 1980-5497. <https://doi.org/10.1590/1980-549720190002.supl.2>]

TAYLOR, Timothy P. et al. Glomerular filtration rate can be accurately predicted using lean mass measured by dual-energy X-ray absorptiometry. **Nephrology Dialysis Transplantation**, v. 21, n. 1, p. 84-87, 2006.

TEDLA, F. M. et al. Hypertension in chronic kidney disease: navigating the evidence. **International journal of hypertension**, v. 2011, 2011.

TEMILOLA, Dada Oluwaseyi et al. Salivary creatinine as a diagnostic tool for evaluating patients with chronic kidney disease. **BMC nephrology**, v. 20, n. 1, p. 1-6, 2019.

TOWNSEND, Raymond R.; TALER, Sandra J. Management of hypertension in chronic kidney disease. **Nature Reviews Nephrology**, v. 11, n. 9, p. 555-563, 2015.]

TUĞCU, Murat; ATAŞ, Dilek Barutçu. Chronic kidney disease progression in aged patients. **International Urology and Nephrology**, p. 1-7, 2021.

UDDIN, Shahadat et al. Comparing different supervised machine learning algorithms for disease prediction. **BMC medical informatics and decision making**, v. 19, n. 1, p. 1-16, 2019.

UDDIN, Shahadat et al. Comparing different supervised machine learning algorithms for disease prediction. **BMC medical informatics and decision making**, v. 19, n. 1, p. 1-16, 2019.

VALDIVIELSO, José Manuel; JACOBS-CACHÁ, Conxita; SOLER, María José. Sex hormones and their influence on chronic kidney disease. **Current opinion in nephrology and hypertension**, v. 28, n. 1, p. 1-9, 2019.

VENKATAPATHY, Ramesh et al. Salivary creatinine estimation as an alternative to serum creatinine in chronic kidney disease patients. **International journal of nephrology**, v. 2014, 2014.

WILLIAMS, Bryan et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology (ESC) and the European Society of Hypertension (ESH). **European heart journal**, v. 39, n. 33, p. 3021-3104, 2018.

WILLIAMSON, PhD Tyler et al. Developing and Validating a Primary Care EMR-based Frailty Definition using Machine Learning. **International Journal of Population Data Science**, v. 5, n. 1, 2020

WINEARLS, Christopher G.; GLASSOCK, Richard J. Classification of chronic kidney disease in the elderly: pitfalls and errors. **Nephron Clinical Practice**, v. 119, n. Suppl. 1, p. c2-c4, 2011.

WITTEN, Ian H.; FRANK, Eibe. Data mining: practical machine learning tools and techniques with Java implementations. **Acm Sigmod Record**, v. 31, n. 1, p. 76-77, 2002.

WOLFF, Jennifer L.; STARFIELD, Barbara; ANDERSON, Gerard. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. **Archives of internal medicine**, v. 162, n. 20, p. 2269-2276, 2002.

YOUNG, Bessie A. et al. Risk factors for rapid kidney function decline among African Americans: the Jackson Heart Study (JHS). **American Journal of Kidney Diseases**, v. 68, n. 2, p. 229-239, 2016.

YUN, Hae-Ryong et al. Obesity, metabolic abnormality, and progression of CKD. **American Journal of Kidney Diseases**, v. 72, n. 3, p. 400-410, 2018.

ZELLER, Carlos Borelli et al. Hipertensão arterial e órgão-alvo: a importância do tratamento. **Revista Brasileira de Hipertensão**, v. 14, n. 1, p. 60-62, 2007.

ANEXOS

ANEXO A – APROVAÇÃO DO COMITÊ DE ÉTICA.

UFMA - UNIVERSIDADE
FEDERAL DO MARANHÃO



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: PREDIÇÃO DA DOENÇA RENAL CRÔNICA UTILIZANDO REDES NEURAIS ARTIFICIAIS

Pesquisador: ALLAN KARDEC DUAILIBE BARROS FILHO

Área Temática:

Versão: 1

CAAE: 67030517.5.0000.5087

Instituição Proponente: Hospital Universitário da Universidade Federal do Maranhão/HU/UFMA

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 2.035.753

Apresentação do Projeto:

O projeto apresentado trata de um estudo clínico, prospectivo, observacional. O estudo prevê a utilização de amostras de indivíduos adultos de 20 a 50 anos, de ambos os sexos, saudáveis ou com patologia de base. No contexto geral de análise, serão avaliados pacientes com Doença Renal Crônica (DRC) em desenvolvimento de modelo de predição da patologia.

Objetivo da Pesquisa:

Desenvolver modelos computacionais capazes de prever os estágios da DRC com base nos fatores de risco envolvidos na progressão da doença.

Avaliação dos Riscos e Benefícios:

Os riscos dizem respeito à punção sanguínea, que pode causar um leve desconforto e o aparecimento de hematomas, o que não interfere na saúde geral dos participantes.

Os benefícios serão de curto e médio prazo, como a detecção precoce de fatores de risco relacionados ao desenvolvimento da doença renal crônica; avaliação da doença renal oculta através de questionário; e o desenvolvimento de software biomédico para predição da doença renal crônica de forma precoce.

Comentários e Considerações sobre a Pesquisa:

O projeto trata de uma temática relevante na saúde pública, a doença renal crônica (DRC). É um

Endereço: Avenida das Portuguesas, 1968 CEB Velho

Bairro: Bloco C.Sala 7, Comitê de Ética **CEP:** 65.060-040

UF: MA **Município:** SAO LUIS

Telefone: (98)3272-8708 **Fax:** (98)3272-8708 **E-mail:** cepufma@ufma.br

ANEXO B - REGISTRO DE SOFTWARE

  
<p>REPÚBLICA FEDERATIVA DO BRASIL Ministério Da Indústria, Comércio Exterior e Serviços Instituto Nacional da Propriedade Industrial</p> <p>Diretoria de Patentes, Programas de Computador e Topografias de Circuitos Integrados</p>
<p>Certificado de Registro de Programas de Computador</p>
<p>Processo nº: BR 51 2018 000138-6</p> <p>O Instituto Nacional da Propriedade Industrial expede o presente certificado de Registro de Programas de Computador, válido por 50 anos a partir de 1º de janeiro subsequente à data de Criação: 04 de junho de 2017, em conformidade com o parágrafo 2º, artigo 2º da Lei Nº 9.609, de 19 de Fevereiro de 1998.</p> <p>Título: Formulário para detecção de doença renal</p> <p>Data de Criação: 04 de junho de 2017</p> <p>Titular(es): UNIVERSIDADE FEDERAL DO MARANHÃO</p> <p>Autor(es): ALLAN KARDEC DUAILIBE BARROS FILHO / CARLOS MAGNO SOUSA JUNIOR / ERIKA CRISTINA RIBEIRO DE LIMA CARNEIRO / EWALDO EDER CARVALHO SANTANA / GUILHERME DE OLIVEIRA LIMA / LUANA MONTEIRO ANAISSE AZOUBEL / NILVIANE PIRES SILVA SOUSA / SALLY CRISTINA MOUTINHO MONTEIRO</p> <p>Linguagem: JAVA</p> <p>Campo de Aplicação: SD-01, SD-06, SD-09</p> <p>Tipo Programa: AP-01, FA-01, FA-03</p> <p>Algoritmo Hash: SHA-512</p> <p>Resumo Digital: a14f12dc473f45bbd3a955aa03ffc2f14855865d60411c232b10ff2dfd470095dfe68d7dea8a6b 455dfa648b1e624e04a138a012f0cb3dddf5a457225f363676</p> <p>Expedido em: 08 de fevereiro de 2018</p> <p>Aprovado por Julio Cesar Castelo Branco Reis Moreira</p>

ANEXO C - REGISTRO DE SOFTWARE

 REPÚBLICA FEDERATIVA DO BRASIL MINISTÉRIO DA ECONOMIA INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS INTEGRADOS	
<h2>Certificado de Registro de Programa de Computador</h2>	
Processo Nº: BR512019000228-8	
<p>O Instituto Nacional da Propriedade Industrial expede o presente certificado de registro de programa de computador, válido por 50 anos a partir de 1º de janeiro subsequente à data de 01/11/2018, em conformidade com o §2º, art. 2º da Lei 9.608, de 19 de Fevereiro de 1998.</p>	
<p>Título: Estimador de gordura corporal para adolescentes</p>	
<p>Data de publicação: 01/11/2018</p>	
<p>Titular(es): UNIVERSIDADE FEDERAL DO MARANHÃO</p>	
<p>Autor(es): ALLAN KARDEC DUAILIBE BARROS FILHO; LUIS CLAUDIO DE OLIVEIRA SILVA; EWALDO EDER CARVALHO SANTANA; GUILHERME DE OLIVEIRA LIMA; NILVANE PIRES SILVA SOUSA; LUANA MONTEIRO ANAISSE AZOUBEL; ERIKA CRISTINA RIBEIRO DE LIMA CARNEIRO; CARLOS MAGNO SOUSA JUNIOR; MARCUS VINÍCIUS DE SOUSA LOPES; GABIANA SOUSA ARAUJO.</p>	
<p>Linguagem: JAVA; XML</p>	
<p>Campo de aplicação: SD-01; SD-05; SD-09</p>	
<p>Tipo de programa: AP-01; FA-01; FA-03</p>	
<p>Algoritmo hash: SHA-512</p>	
<p>Resumo digital hash: cdf0d259450f801180a0190ee298fe8b73ade16c3e85419d45a971027707a1e5d770355abede210be19b5ee5dd56dfc09bc9f90ced27c41108a71fedd8f</p>	
<p>Expedido em: 18/02/2019</p>	
<p>Aprovado por: Alexandre Gomes Glancio Diretor Substituto de Patentes, Programas de Computador e Topografias de Circuitos Integrados</p>	

ANEXO C - ARTIGO PUBLICADO

Journal of Health Informatics - Fator de impacto: 1,833



XVI Congresso Brasileiro de Informática em Saúde

APLICATIVO PARA TRIAGEM DE PACIENTES EM RISCO DE DOENÇA RENAL CRÔNICA

APPLICATION FOR SCREENING PATIENTS AT RISK FOR CHRONIC RENAL DISEASE

APLICACIÓN PARA LA SELECCIÓN DE PACIENTES EN RIESGO DE ENFERMEDAD RENAL CRÓNICA

Guilherme Lima¹, Nilviane Pires², Luana Azoubel², Gabiana Sousa³, Erika Carneiro⁴, Renato Barros³, Carlos Sousa Junior¹, Naydson Sousa¹, Allan Kardec Duailibe^{1,2}, Ewaldo Santana¹

¹ Programa de Pós-Graduação em Engenharia Elétrica . Universidade Federal do Maranhão – UFMA. São Luís/MA, Brasil

² Programa de Pós-Graduação em Biotecnologia. Universidade Federal do Maranhão – UFMA. São Luís/MA, Brasil.

³ Laboratório de Processamento da Informação Biológica. Universidade Federal do Maranhão -UFMA. São Luís/MA, Brasil.

⁴ Programa de Pós-Graduação em Ciências Médicas. Universidade Estadual do Rio de Janeiro – UERJ. Rio de Janeiro/ RJ, Brasil.

Resumo: Objetivo: Automatizar o processo de triagem de pacientes com doença renal oculta tendo por base o questionário SCORED. **Método:** O aplicativo (*app*) foi desenvolvido no *software Android Studio*[®]. Após obtenção da versão final, o *app* foi testado em um dispositivo *mobile Android*. E para avaliação do seu desempenho foram utilizados dados de 30 pacientes oriundos de um centro de referencia para prevenção e tratamento da doença renal. Foram calculados os seguintes indicadores de desempenho acurácia, sensibilidade e especificidade. **Resultado:** O aplicativo baseado no questionário SCORED obteve alta sensibilidade (85%). Além disso, cabe destacar que o desenvolvimento do aplicativo gerou agilidade e rigor no processo de triagem/diagnóstico uma vez que elimina o uso de papel e as pontuações são obtidas automaticamente. **Conclusão:** O aplicativo SCORED torna-se uma poderosa



ferramenta de triagem de pacientes em risco de doença renal. Sendo um método com alta sensibilidade, reprodutível e baixo custo que pode ser utilizado nos mais diversos locais, incluindo em lugares remotos e sem fácil acesso a um profissional de saúde.

Descritores: Nefrologia, Aplicativos para Saúde, Sistemas de Apoio.

Abstract: Objective: *To automate the screening process of patients with hidden kidney disease based on the SCORED questionnaire. Method:* *The app (app) was developed in Android Studio® software. After getting the final version, the app was tested on an Android mobile device. To evaluate its performance, data from 30 patients from a referral center for the prevention and treatment of renal disease were used. The following indicators of accuracy, sensitivity and specificity were calculated. Result:* *The application based on the SCORED questionnaire obtained high sensitivity (85%). In addition, it should be noted that application development has generated agility and accuracy in the screening / diagnosis process since it eliminates the use of paper and scores are obtained automatically. Conclusion:* *The SCORED application becomes a powerful screening tool for patients at risk for kidney disease.*

Keywords: *Nephrology, Health Applications, Support Systems.*

Descriptores: *Nefrología, Aplicaciones para la Salud, Sistemas de Apoyo.*

Introdução

A doença renal crônica (DRC) é caracterizada pela alteração da função renal¹. Sendo definida classicamente como anormalidades da estrutura ou função dos rins, presentes por mais de três meses e com implicações para a saúde^{2,3}. É um problema de saúde global que está aumentando em todo mundo⁴.

A DRC geralmente se apresenta assintomática em seus estágios iniciais. E isso muitas vezes faz com que o paciente procure o serviço de saúde tardiamente já necessitando de terapia renal substitutiva⁵. Desta forma, os profissionais de saúde devem sempre estar em alerta a sinais e sintomas principalmente em pacientes que já possuam fatores de risco⁵ para DRC como hipertensão e diabetes, por exemplo.



Com intuito de auxiliar no diagnóstico precoce da DRC na população brasileira em 2012, um questionário para avaliação da doença renal crônica oculta denominado "Screening for Ocult Renal Disease" (SCORED) foi traduzido para o português brasileiro e validado, tal questionário apresentou uma acurácia de 60% na predição de pacientes com DRC. Sendo uma arma poderosa na prevenção e diagnóstico dessa patologia⁶. Porém, algumas modificações ainda podem ser efetuadas no questionário com intuito de aumentar sua abrangência, aplicabilidade e acurácia.

Uma ferramenta relevante para avaliação em saúde e acompanhamento de pacientes é uso de aplicativos. O Brasil possui 198 milhões de "celulares inteligentes" (*smartphones*), e a tendência é que esse número chegue 236 milhões em dois anos⁷. Assim, a tecnologia dos *smartphones* tornou-se comum entre as pessoas, da mesma forma que a disponibilidade e popularidade de aplicativos (*apps*) de saúde aumentaram. Esses aplicativos são utilizados para apoiar o diagnóstico e manejo de diversas patologias como: diabetes, doenças cardíacas, obesidade e hipertensão arterial, por exemplo. Por esse motivo diversos estudos estão focados no desenvolvimento de aplicativos de saúde em *smartphone*⁸.

Logo, o presente estudo tem por objetivo desenvolver um aplicativo de triagem para doença renal crônica baseado no questionário "Screening for Ocult Renal Disease" (SCORED) traduzido. Além de propor algumas modificações em sua estrutura a fim de melhorar sua aplicabilidade e abrangência de uso.

Métodos

Desenvolvimento do aplicativo

O *software* para triagem da DRC foi desenvolvido para dispositivos móveis com o objetivo de atender a um público diversificado visto que grande parte da população possui um dispositivo móvel. O Sistema Operacional *mobile* escolhido foi o *Android*, pois ele possui a maior predominância no mercado de dispositivos *mobile*, chegando a 85% do mercado no início de 2017⁹.

O ambiente de programação utilizado foi o *Android Studio*[®], fornecido gratuitamente pela Google. A plataforma de desenvolvimento é instalada em um computador e os aplicativos desenvolvidos nela podem ser rodados direto em um celular conectado ao computador, ou em um celular virtual gerado pela plataforma o que permite a execução de testes prévios do *app*.



A linguagem de programação utilizada foi o Java que é padrão do *Android Studio*[®] e possui ampla documentação para desenvolvimento de *apps*. Para implementação da *interface* de usuário foi utilizado o XML (*Extensible Markup Language*), também padrão do *Android Studio*[®].

Foi utilizado o processo de desenvolvimento incremental de *software*, que permite uma construção rápida de uma versão inicial, e um processo flexível à adição e alteração de funcionalidades, lançando novas versões em cada modificação até chegar à versão final do *software*¹⁰.

Os requisitos iniciais foram levantados por meio das informações essenciais que o SCORED estabelece, então foi elaborada a versão inicial do *app*. Apresentando a versão inicial a profissionais que trabalham com DRC e através de pesquisa bibliográfica se obteve requisitos adicionais permitindo a construção de novas versões até versão final.

Manter as informações do paciente é essencial para traçar um perfil de saúde do mesmo, o modelo adotado para preservar essas informações foi salvá-las em arquivo de texto no formato txt, facilitando a transição do arquivo no dispositivo móvel para computador, independente do sistema operacional.

Base de dados

Para avaliação dos indicadores de desempenho sensibilidade, especificidade e acurácia foram utilizados dados de 30 pacientes oriundos do Centro de Prevenção a Doença Renal Crônica do Hospital Universitário da Universidade Federal do Maranhão. Onde foram avaliadas as seguintes variáveis: massa corporal, estatura, idade, gênero, circunferência da cintura, perda de proteína pela urina, presença de diabetes, anemia e pressão arterial elevada, histórico de: ataque cardíaco, derrame, insuficiência cardíaca ou problema circulatório nas pernas.

Avaliação da Função Renal

A avaliação da função renal dos indivíduos pertencentes ao grupo de validação (n=30) do aplicativo foi obtida através da análise da taxa de filtração glomerular estimada tendo por base a equação *Chronic Kidney Disease Epidemiology Collaboration* (CKD-EPI)¹¹ para adultos. A estimativa obtida através da CKD-EPI foi utilizada como padrão na análise dos resultados obtidos pela ferramenta de triagem (*app*).



Antropometria

O peso foi medido com balança eletrônica calibrada (*Omron*[®] HBF214 LA, Japão) com resolução de 0,1 kg. A altura foi medida através do estadiômetro transportável vertical com resolução de 0,1 cm (*Sanny*[®], Brasil). A medição da circunferência da cintura foi mensurada, utilizando uma trena antropométrica inelástica com precisão de 0,1 cm (*Seca*[®] 213, Hamburgo, Alemanha), no ponto médio entre a última costela e a crista ilíaca na respiração mínima.

O IMC e RCE foram estimados através das fórmulas (1) e (2), respectivamente¹².

$$IMC = \frac{\text{peso}}{\text{altura}^2} \quad (1)$$

$$RCE = \frac{\text{circunferência da cintura}}{\text{altura}} \quad (2)$$

No cálculo do IMC a unidade de medida usada para o peso é quilogramas (Kg) e a altura é usada em metros (m). No cálculo do RCE ambos os indicadores são usados em centímetros (cm).

Crítérios de Avaliação da Doença Renal Oculta (app)

A análise da doença renal oculta pelo aplicativo teve por base os critérios descritos por Magacho et al. (2012)⁶, conforme ilustrado na Tabela 1.

Tabela 1 - Critérios para avaliação da doença renal oculta⁶

Crítérios	Pontuação
Ter entre 50 e 59 anos	2
Ter entre 60 e 69 anos	3
Ter 70 anos ou mais	4
Ser do gênero feminino	1
Tive/tenho anemia	1
Tenho pressão alta	1
Sou diabético	1
Histórico de infarto ou derrame/AVC/AVE	1
Tenho insuficiência cardíaca congestiva ou insuficiência cardíaca	1
Problema de circulação/ doença circulatória nas pernas	1



Perda de proteína na urina	1
----------------------------	---

Abreviações: AVC- acidente vascular cerebral; AVE- acidente vascular encefálico.

Critérios para triagem:

- Pontuação geral de 4 ou mais pontos: o paciente tem 1 chance em 5 ter doença renal crônica (o paciente, nessa faixa, deve procurar um médico para diagnosticar se realmente possui DRC);
- Pontuação de 0-3 pontos provavelmente não tem doença renal, porém pelo menos uma vez no ano deve refazer este teste.

Cálculo dos indicadores de desempenho

Os indicadores utilizados na avaliação do desempenho do aplicativo foram acurácia, sensibilidade e especificidade. Esses métodos de avaliação utilizam como parâmetros de suas fórmulas a quantidade de verdadeiro positivo – VP, verdadeiro negativo – VN, Falso Positivo – FP e Falso Negativo – FN¹³.

O índice de acurácia mostra quanto o *app* acertou, ou seja, quanto pacientes doentes ele classificou com risco e quanto dos pacientes normais ele detectou como sem risco. Sua equação é definida na equação (3).

$$\text{acurácia} = \frac{VP+VN}{VP+VN+FP+FN} \quad (3)$$

A sensibilidade determina quanto o *app* foi capaz de acertar os pacientes com DRC, a fórmula para sua estimação é dado pela equação (4).

$$\text{sensibilidade} = \frac{VP}{VP+FN} \quad (4)$$

Já a especificidade indica quanto dos pacientes normais foram triados corretamente pelo *app*. A equação 5 mostra como este índice é calculado.

$$\text{especificidade} = \frac{VN}{VN+FP} \quad (5)$$

Aspectos éticos



O presente estudo possui aprovação do Comitê de Ética em Pesquisa da Universidade Federal do Maranhão segundo parecer CAAE 67030517.5.0000.5087.

Resultado

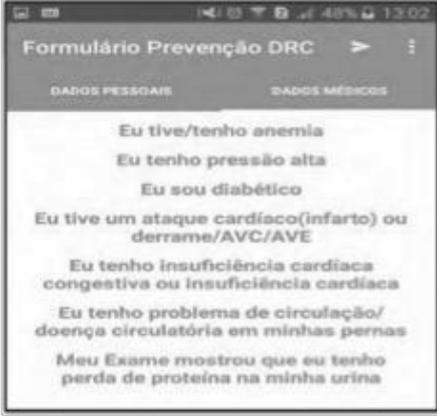
O aplicativo apresenta uma interface simples e ágil que permite a rápida avaliação do paciente. Ao abrir o usuário do *app* é direcionado para as telas de coleta de informações do paciente a ser avaliado. A coleta de dados consiste em duas etapas:

1. Coleta de dados clínicos e pessoais;
2. Coleta de dados do questionário.

Cada uma dessas etapas possui uma tela específica para coletar as informações, essas etapas podem ser executadas em qualquer ordem. A primeira etapa, conforme Figura 1, consiste em obter dados pessoais, medidas antropométricas e contato do paciente.

Figura 1: Tela referente aos dados clínicos e pessoais.

A segunda etapa (Figura 2) consiste em obter dados referentes aos questionário SCORED. Os textos nesta tela são selecionáveis, e caso o paciente apresente uma das situações clínicas citadas basta que o aplicador clique em cima da mesma e a cor de fundo se tornará azul (Figura 3).



Formulário Prevenção DRC

DADOS PESSOAIS DADOS MÉDICOS

Eu tive/tenho anemia

Eu tenho pressão alta

Eu sou diabético

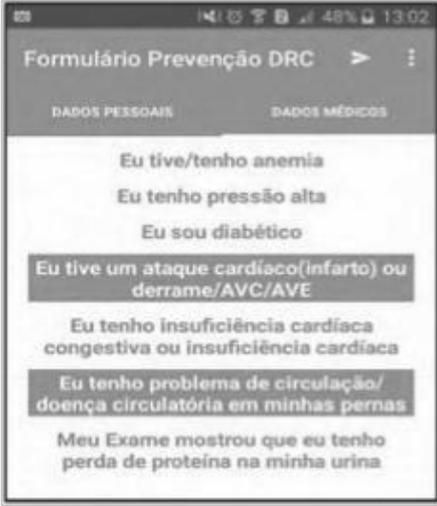
Eu tive um ataque cardíaco (infarto) ou derrame/AVC/AVE

Eu tenho insuficiência cardíaca congestiva ou insuficiência cardíaca

Eu tenho problema de circulação/ doença circulatória em minhas pernas

Meu Exame mostrou que eu tenho perda de proteína na minha urina

Figura 2: Tela relativa as informações do questionário SCORED.



Formulário Prevenção DRC

DADOS PESSOAIS DADOS MÉDICOS

Eu tive/tenho anemia

Eu tenho pressão alta

Eu sou diabético

Eu tive um ataque cardíaco (infarto) ou derrame/AVC/AVE

Eu tenho insuficiência cardíaca congestiva ou insuficiência cardíaca

Eu tenho problema de circulação/ doença circulatória em minhas pernas

Meu Exame mostrou que eu tenho perda de proteína na minha urina

Figura 3: Exemplo de seleção de características referentes aos dados do questionário.

Após preencher as informações basta o aplicador clicar na seta que fica no canto superior esquerdo da tela. E o sistema irá analisar as informações preenchidas, verificar se há campos vazios, e caso exista será dado um aviso, e o aplicador será direcionado ao campo que deve ser preenchido.



Caso não tenha nenhuma informação obrigatória pendente o sistema irá salvar as informações do paciente, calcular o Índice de Massa Corporal (IMC) e a Relação Cintura Estatura (RCE), além da pontuação geral para determinar se o paciente tem risco de DRC. As informações da pontuação geral, IMC e RCE são exibidas na tela de resultado (Figura 4).

Total:	IMC:	RCE:
6	29,76	0,38

Você tem 1 chance em 5 de ter doença renal crônica. Na sua próxima visita a um médico, um simples exame de sangue deve ser pedido. Somente um profissional da saúde pode determinar com certeza se você tem doença renal.

Figura 4: Tela de diagnóstico para paciente com risco de DRC.

Contudo, se o sistema detectar que o paciente não apresenta risco de doença renal o diagnóstico será exibida uma mensagem e um alerta que o mesmo deve refazer esta pesquisa um vez por ano, conforme Figura 5.

Total:	IMC:	RCE:
2	30,92	0,16

Você, provavelmente, não tem uma doença renal agora, mas, pelo menos uma vez por ano, você deve fazer esta pesquisa.

Figura 5: Tela de diagnóstico para paciente sem risco de DRC

Ao finalizar o uso do aplicativo basta que o usuário o feche, e para cadastrar novo paciente se deve clicar no botão voltar do dispositivo. O arquivo com as informações do paciente fica salvo no aparelho, e essas informações podem ser passadas para o computador através do cabo USB.



Para análise do desempenho do aplicativo utilizou-se uma amostra de 30 pacientes com média de idade, IMC, CC e RCE de $63 \pm 11,04$ anos, $28,52 \pm 4,09$ kg/m², $96,97 \pm 10,73$ cm e $0,63 \pm 0,07$, respectivamente.

A acurácia obtida foi de 50%, sensibilidade 85% e especificidade 24%, conforme demonstrado na Tabela 2.

Tabela 2 – Tabela de resultados da avaliação do aplicativo

Acurácia (%)	Sensibilidade (%)	Especificidade (%)	VP (n)	VN (n)	FP (n)	FN (n)
50	85	24	10	4	14	2

Abreviação: VP- verdadeiro positivo; VN- verdadeiro negativo; FP-Falso positivo; FN- falso negativo.

Discussão

A triagem obtida através do aplicativo apresentou alta sensibilidade, o que é um fator de suma importância uma vez que métodos de triagem com alta sensibilidade permitem a identificação mais precisa de grupos de risco, além da implementação de medidas preventivas de forma mais rápida e precisa evitando desta forma desfechos negativos¹⁴ como morte precoce.

Entretanto, os indicadores acurácia e especificidade ainda podem ter seus valores elevados. Uma alternativa seria a inserção de variáveis de baixo custo como os indicadores antropométricos. Tais indicadores já são utilizados para o rastreio de diversas comorbidades incluindo a DRC e apresentam vantagens como baixo custo, reprodutibilidade e fácil aplicação tornando a identificação do risco para DRC ainda mais prática.

Assim, destaca-se que é de fundamental importância incluir a análise da obesidade nestes critérios, visto que esta é fator de risco para desenvolvimento de diabetes tipo II e doenças cardiovasculares, e tais patologias estão relacionadas ao aparecimento de DRC¹⁴. Em recente estudo realizado nos Estados Unidos, verificou-se que o número de renais crônicos com obesidade chega a 44,1%¹⁵.

Logo, reconhecer a necessidade do paciente é um fator importante para uma maior disseminação das tecnologias de saúde. E para que estes usuários se envolvam ainda mais com o cuidado da própria saúde e quando possível possam receber



assistência médica ágil e até mesmo remota¹⁶, possibilitando desta maneira melhor qualidade de vida.

O uso de questionários digitais tem se apresentado como uma ferramenta promissora, visto que apresentam uma série de vantagens como agilidade, disponibilidade, fácil entendimento e baixo custo. O fato de estarem disponíveis nos *smartphones* amplia sua abrangência e aplicabilidade já que poderão ser utilizados em lugares públicos, consultórios médicos, hospitais e campanhas de saúde¹⁷.

A utilização dessas tecnologias digitais tem auxiliado os profissionais da saúde na prevenção e tratamento de doenças crônicas, possibilitando uma forma singular de promoção à saúde, pois alcançam uma grande parcela da população¹⁷. Somando-se ao fato que a aplicação de questionários digitais facilitam o diagnóstico precoce, a intervenção e acompanhamento médico, uma vez que eliminam o uso de papel. E permitem que a triagem seja feita de forma automática sem a necessidade de obtê-la de forma manual (através da soma das pontuações de cada característica do questionário SCORED).

Ressaltasse que com o aumento alarmante da obesidade e sua associação com desenvolvimento de diversas patologias como a DCR, é de fundamental importância que seu rastreo seja incluindo nos métodos de triagem e/ou diagnóstico. Logo, para a inclusão desta variável e dos indicadores antropométricos deverá ser efetuada uma reestruturação do questionário SCORED, com posterior validação, onde deve ser atribuída uma pontuação a cada uma dessas variáveis. O aplicativo possui registro junto ao Instituto Nacional da Propriedade Industrial conforme parecer BR 51 2018 000138-6.

Conclusão

A criação do *app* baseado no formulário SCORED agiliza o processo de diagnóstico, uma vez que sua implementação permitirá o uso em lugares remotos, inclusive sem acesso à internet, e a auto-aplicação do questionário pelo paciente facilitará o acesso dos profissionais de saúde a informações sobre o paciente. Aliado a essa questão, a implementação do *app* eliminará a necessidade do uso de papel durante aplicação do questionário. E em um mundo, onde se busca a sustentabilidade a exclusão de mais uma barreira, como o uso ineficaz do papel, ajudará ainda mais na incorporação desse conceito em nosso cotidiano.



Agradecimentos

Os autores agradecem a FAPEMA pela concessão de bolsa (Processo n. BM-02811/18). E ao Centro de Prevenção à Doença Renal Crônica HUUFMA pela liberação para coleta de dados.

Referências

- [1] Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância de Doenças e Agravos não Transmissíveis e Promoção da Saúde. (2015). "Vigitel Brasil 2014 : vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico". Brasília: Ministério da Saúde.
- [2] National Kidney Foundation. (2013). "KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease" [Internet]. *Kidney Inter Suppl*, 3(1): 1-150, jan. Disponível em: http://www.kdigo.org/clinical_practice_guidelines/pdf/CKD/KDIGO_2012_CKD_GL.pdf. Acesso em: 16 fev. 2018.
- [3] Draibe SA. (2014). "Panorama da Doença Renal Crônica no Brasil e no mundo". UNASUS/UFMA - São Luís.
- [4] Baumgarten M and Gehr T. (2011). "Chronic Kidney Disease: Detection and Evaluation". *American Family Physician*, 84(10).
- [5] Holstein JM, et. al. (2016). "A APLICAÇÃO DO QUESTIONÁRIO SCORED COMO PREDITOR DE DOENÇA RENAL NO MUNICÍPIO DE URUGUAIANA, RS". *Anais do Salão Internacional de Ensino, Pesquisa e Extensão*, 8(3).
- [6] Magacho EJC, et. al. (2012). "Tradução, adaptação cultural e validação do questionário Rastreamento da Doença Renal Oculta (Screening For Occult Renal Disease - SCORED) para o português brasileiro". *Jornal Brasileiro de Nefrologia*, 34(3): 251-258.
- [7] Pesquisa anual do uso de TI. FUNDAÇÃO GETULIO VARGAS (FGV) [Internet]. 2017. [2018 fev. 26]. Disponível em: <http://easp.fgv.br/ensinoeconhecimento/centros/cia/pesquisa>.
- [8] Al-Mardini M, et al. (2014). "Classifying obstructive sleep apnea using smartphones". In: *Journal of Biomedical Informatics*, 52: 251-259.
- [9] Douglas Ciriaco. Android cresce, iOS diminui e Windows Phone quase some, aponta relatório. *techmundo* [Internet]. 2017 [cited 2018 jun. 28]. Disponível em: <http://www.webcitation.org/70iJXqJdD>.



- [10] Pressman, RS. Engenharia de Software: Uma Abordagem Profissional. 7 ed. Porto Alegre: AMGH, 2011.
- [11] "Passo a passo para a implantação da estimativa da taxa de filtração glomerular (eTFG)". SBPC/ML [Internet]. 2015. Disponível em: http://www.sbpc.org.br/upload/conteudo/padronizacao_eTFG_4nov2015.pdf.
- [12] Ashwell M, Hsieh SD. (2005). "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". *International journal of food sciences and nutrition*, 56(5), 303-307.
- [13] Ribeiro ACC. "Rastreamento não invasivo para diabetes tipo 2" [Tese de Doutorado]. Universidade Federal do Maranhão; 2015.
- [14] Zang S, Tjortjis C, Zeng X, Qiao H, Buchan I, Keane J. "Comparing data mining methods with logistic regression in childhood obesity prediction". *Information Systems Frontiers*, 2009; 11:449-460.
- [15] Evangelista LS, Cho WK, Kim Y. (2018). Obesity and chronic kidney disease: A population-based study among South Koreans. *Plos One*, 13(2).
- [16] Chang AR, Zafar W, Grams ME. (2018). "Kidney Function in Obesity-Challenges in Indexing and Estimation". *Advances in Chronic Kidney Disease*, 25(1): 31-40.
- [17] Neves NTAT, Araújo, YB, Costa CM, Cardoso ELS, Ferreira EMV. "TENDÊNCIAS DE ESTUDOS SOBRE APLICATIVOS MÓVEIS PARA SAÚDE: REVISÃO INTEGRATIVA". *Journal of Health Informatics*, 2016; 8 suppl I: 499-508.
- [18] Baldo C, et al. (2015). "Diabetes Food Control – Um aplicativo móvel para avaliação do consumo alimentar de pacientes diabéticos". *RECIIS*, 9(3).

Contato

Guilherme de Oliveira Lima

Universidade Federal do Maranhão.

Laboratório de Processamento da Informação Biológica – PIB.

Telefone: +55 (98) 3272-8372

E-mail: guilhermeolima2010@gmail.com

ANEXO D - ARTIGO PUBLICADO

Arquivos Brasileiros de Cardiologia, fator de impacto: 2,0

Short Editorial



Analysis of Sensitivity and Specificity of Cutoff Points for Resting Heart Rate in 6,794 Brazilian Adolescents: A Cross-Sectional Study

Luana Anaisse Azoubel,¹ Erika Ribeiro Carneiro,¹ Nilviane Pires,¹ Allan Kardec Barros,³ Carlos José Dias²

Universidade Federal do Maranhão, Centro de Prevenção de Doenças Renais do Hospital Universitário do Maranhão, 1 São Luís, MA - Brazil
Universidade Federal do Maranhão, Laboratório de Adaptações Cardiorrenais Ao Exercício Físico (LACE), 2 Pinheiro, MA - Brazil

Universidade Federal do Maranhão, Departamento de Engenharia Elétrica, Laboratório de Processamento de Informação Biológica, 3 São Luís, MA - Brazil

Short Editorial related to the article: Sensitivity and Specificity of Cutoff Points of Resting Heart Rate from 6,794 Brazilian Adolescents:

A CrossSectional Study

Resting heart rate (RHR) is an accessible measure that may reflect the balance between the sympathetic and parasympathetic nervous system.¹ Elevated RHR levels are associated with cardiovascular events in adults² and cardiovascular risk factors, such as overweight,³ abdominal obesity and high blood pressure have been associated with high RHR.⁴ Therefore, obtaining a cutoff point for the RHR in children and adolescents is essential for better screening and subsequent clinical follow-up.

Furthermore, the early identification and screening of cardiovascular changes and cardiac autonomic modulation can be reflected by simple measures such as the RHR, since sedentary adolescents with abdominal obesity have lower heart rate variability (HRV) and, consequently, greater sympathetic stimulation, maintaining the RHR high and increasing the risk of developing heart damage throughout life.⁵

One method of verifying autonomic modulation is through the HRV, which when decreased indicates autonomic imbalance and has been associated with mortality and various cardiovascular risk indicators (general and abdominal obesity, high blood pressure and sedentary lifestyle) in adolescents.⁶⁻⁸ A previous study by Farah et al.,⁹ carried out with male adolescents, demonstrated that cutoff points for HRV have moderate-to-high sensitivity for detecting cardiovascular risk factors.⁹ This reinforces the relevance of studies that propose establishing cutoff points for biological signal variables, such as the RHR and its different clinical analyses, aiming to diagnose possible cardiovascular alterations in children and adolescents.

This is what Farah et al.¹⁰ proposed in this study, by establishing cutoff points for RHR in Brazilian adolescents and analyzing associations between cutoff points and cardiovascular risk factors. A total of 6,794 adolescents (aged 10 to 19 years), of both genders, were evaluated and blood pressure and RHR were measured using an oscillometric device. Body mass index (BMI) and waist circumference were also evaluated. The ROC curve was

used to analyze sensitivity and specificity, and associations of high RHR with cardiovascular risk factors were analyzed by binary logistic regression.¹⁰

The main findings of this study, in addition to determining the RHR cutoff points, was to establish the association of these cutoff points with cardiovascular risk factors, such as abdominal obesity, overweight and high blood pressure in boys and girls.

A study by Christofaro et al.¹¹ found an association between heart rate and systolic and diastolic blood pressure in adolescents, and it was pointed out that, for boys, there was an influence of abdominal obesity and BMI in those with a higher heart rate, but this did not occur in girls; thus, the authors suggested that patterns of body fat distribution and hormonal variations may influence these findings.¹¹ In fact, the android body fat distribution pattern is more associated with cardiovascular risk factors and female hormones (estrogens) provide some cardiovascular protection in girls.¹²

Furthermore, as expected, there is a decline in RHR with increasing age, probably due to the improvement in the baroreflex sensitivity and neural function with sexual maturation. Indeed, there is a progressive increase in vagotonic cardiac activity in relation to sympathetic activity during maturation, resulting in lower RHR in late adolescence.¹³

This study also has two other important highlights: the number of participants was considerable (important for epidemiological purposes) and it was carried out in three different regions of the country (northeast, southeast and south), possibly with different ethnicities, although it was not reported, therefore being representative from the perspective of genetic variability. Despite these results, the study has limitations: it was not possible to establish a cutoff point for girls aged 10-14 years and the reason was not explained. External factors that influence heart rate (smoking, level of physical activity, consumption of alcoholic beverages and caffeine) could not be identified, and it is important to highlight these limitations for future studies.¹⁴

For epidemiological screening purposes, establishing cutoff points that allow the identification of increased cardiovascular risk is of utmost importance, since it can improve public health strategies for a population that is presumably at higher risk. Also, regarding the young population, these strategies can be implemented even in schools (encouraging the practice of sports and healthy eating) with the reinforcement of a physiological substrate, the RHR, to ratify the importance of these strategies. Therefore, since it is a simple and low-cost variable to measure, its use can improve and increase the clinical practicality for health professionals in determining diagnosis and prognosis of young Brazilians.

Keywords

Heart Rate/physiology; Adolescent; Obesity; Adiposity; Blood Pressure; Epidemiology; Cross-sectional studies/methods.

Mailing Address: Luana Anaisse Azoubel

Hospital Universitário da Universidade Federal do Maranhão - Rua Barão do Itapary, 227, Postal Code 65020-070, São Luís, MA - Brazil
E-mail: laazoubel@gmail.com

DOI: <https://doi.org/10.36660/abc.20210512>

References

1. Valentini M, Parati G. Variables influencing heart rate. *Prog Cardiovasc Dis.* 2009; 52(1): 11-9.
2. Aune D, Sen A, G-Hartagh B, Janszky I, Romundstad PR, Tonstad S, et al. Resting heart rate and the risk of cardiovascular disease, total cancer, and all-cause mortality—a systematic review and dose-response meta-analysis of prospective studies. *Nutr Metab Cardiovasc Dis.* 2017; 27(6): 504-17.
3. Jardim PCBV. Overweight, the cardiovascular risk of the century. *Arq Bras Cardiol.* 2019; 113(2): 185-7.
4. Fernandes RA, Freitas J I, Codogno JS, Christofaro DC, Monteiro HL, Lopes DM. Resting heart rate is associated with blood pressure in male children and adolescents. *J Pediatr.* 2011; 158(4): 634-7.
5. Farah BQ, Andrade-Lima A, Germano-Soares AH, Christofaro DC, Barros MV, Prado WL, et al. Physical activity and heart rate variability in adolescents with abdominal obesity. *Pediatr Cardiol.* 2018; 39(3): 466-72.
6. Farah BQ, Prado WLd, Tenório TR. Heart rate variability and its relationship with central and general obesity in obese normotensive adolescents. *Einstein (SP).* 2013; 11(3): 285-90.
7. Vanderlei LCM, Patre CM, Freitas Júnior IF, Freitas JI, Godoy ME. Geometric indexes of heart rate variability in obese and eutrophic children. *Arq Bras Cardiol.* 2010; 95(1): 35-40.
8. Farah BQ. Variabilidade da Frequência Cardíaca como Indicador de Risco Cardiovascular em Jovens. *Arq Bras Cardiol.* 2020; 115(1):59-60.
9. Farah BQ, Christofaro DCD, Cavalcante BR, Andrade Lima A, Germano Soares FN, Ritti RM. Cutoffs of short-term heart rate variability parameters in Brazilian adolescents male. *Pediatr Cardiol.* 2018; 39(7): 1397-403.
10. Farah B, Christofaro D, Andrade Lima A, Soares AG, Tebar W, et al. William & Barros. Sensitivity and specificity of cutoff points of resting heart rate from 6794 Brazilian adolescents: a cross-sectional study. *Arq Bras Cardiol.* 2021; 117(1):74-81.
11. Christofaro DCD, Casonatto J, Vanderlei LCM, Cucato GC, Dias RMR. Relationship between Resting Heart Rate, Blood Pressure and Pulse Pressure in Adolescents. *Arq Bras Cardiol.* 2017;108(5):405-410. doi:10.5935/abc.20170050
12. O'Neil A, Scovelle AJ, Milner AJ, Kavanagh A. Gender/Sex as a Social Determinant of Cardiovascular Risk. *Circulation.* 2018; 137(8):854-64. doi:10.1161/CIRCULATIONAHA.117.028595
13. Chen S-R, Chiu H-W, Lee Y-J, Sheen TC, Jong C. Impact of pubertal development and physical activity on heart rate variability in overweight and obese children in Taiwan. *J Sch Nurs.* 2012;28(4):284-90.
14. Doom JR, Mason SM, Saqlia SE, Clark CJ. Pathways between childhood/ adolescent adversity, adolescent socioeconomic status, and long-term cardiovascular disease risk in young adulthood. *Soc Sci Med.* 2017;188:166-75. doi:10.1016/j.socscimed.2017.06.044



This is an open-access article distributed under the terms of the Creative Commons Attribution License

ANEXO E - ARTIGO PUBLICADO

Journal on Advances in Signal Processing, fator de impacto: 1,63

Queiroz et al. *EURASIP Journal on Advances in Signal Processing* (2019) 2019:16
<https://doi.org/10.1186/s13634-019-0613-9>

EURASIP Journal on Advances
in Signal Processing

RESEARCH

Open Access

Support system for classification of beat-to-beat arrhythmia based on variability and morphology of electrocardiogram



Jonathan Araujo Queiroz^{*} , Luana Monteiro Anaisse Azoubel and Allan Kardec Barros

Abstract

Background: Several authors use the R-R interval, which is the temporal difference between the largest waves (R waves) of the electrocardiogram (ECG), to propose a support system for the diagnosis of arrhythmias. However, R-R interval analysis does not measure ECG waveform deformations such as P wave deformations for atrial fibrillation.

Objective: In this study, we propose an arbitrary analysis the any segment of the heartbeat. This analysis is a generalization of a previous work that measures the wave deformations of the ECG signal.

Methods: We proposed to investigate the voltage (mV) variation occurring at each heartbeat interval using statistical moments. Unlike the R-R interval in which each heartbeat is associated with a single real number, the proposed method associates each heartbeat to a set of points, that is, a vector. The heartbeats were obtained in the following databases: MIT-BIH Normal Sinus Rhythm, MIT-BIH Atrial Fibrillation (AF), and MIT-BIH Arrhythmia; and the classifiers used to evaluate the proposed method were linear discriminant analysis, k-nearest neighbors, and support vector machine. The experiments were conducted using 80% of the patients for training (16 healthy patients, 41 patients with arrhythmia, and 20 patients with AF) and 20% of the patients for testing (2 healthy patients, 6 patients with arrhythmia, and 3 patients with AF).

Results: The proposed method proved to be efficient in solving global (accuracy is up to 99.78% in the arrhythmia classification) and local (accuracy of 100% in the AF classification) heartbeat problems.

Conclusion: The results obtained by the proposed method can be used to support decision-making in clinical practices.

Keywords: Heartbeat, R-R interval, Morphological information, Statistical moments, Arrhythmias, Atrial fibrillation

1 Introduction

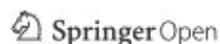
World Health Organization indicates that, each year, 17.3 million people die worldwide from cardiovascular disease [1]. In this context, the interest in noninvasive analysis techniques for the diagnosis of arrhythmias, such as the electrocardiogram (ECG), is increasing.

The ECG is an examination that records the variation of the electrical potentials of the cardiac muscle. ECG is composed of the P wave that corresponds to the electrical activity of the atria, by the Q, R, and S waves composing the QRS complex that corresponds to depolarization

of the ventricles, by the T wave that registers the repolarization of the ventricles [2]. From the QRS complex, specifically the R waves (larger peaks of the ECG signals), it is possible to obtain R-R intervals which are the time difference between two R waves.

Conventionally, the R-R interval is used to data extraction from the ECG signal in order to diagnose different types of arrhythmia [3–8]. However, the analysis of the R-R intervals is not able to measure changes on other ECG waves, such as the distortions on P wave for atrial fibrillation (AF) [9–12]. Thereby, some studies segment the ECG signal [13–17]. Nevertheless, the diagnosis is still not ideal. The present research proposes an arbitrary analysis to any segment of the ECG signal. This analysis is based on ECG variability and morphology. The morphological

*Correspondence: queirozj@gmail.com
 Department of Electrical Engineering, Federal University of Maranhão, Av. dos Portugueses, 1966, Bacanga, 65080-805 São Luís, Maranhão, Brazil



© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

information of the ECG signal is obtained by the voltage (mV) variation investigation occurring to each heartbeat interval. This new method of data extraction will allow a beat-to-beat analysis. Unlike the R-R interval in which each heartbeat is associated with a single real number, the proposed method associates each heartbeat to a set of points, that is to say, to a vector.

In this study, we propose to investigate the morphological information occurring at each heartbeat interval using statistical moments. The statistical moments used were variance, skewness, and kurtosis. The improvement in performance is due to the information obtained from the voltage variability and the morphology of the ECG signal, i.e., in addition to the frequency modulation information used in the heart rate variability (HRV) calculation, we added the amplitude modulation. Thus, to classify different types of arrhythmias, we use two modulation information of the ECG signal (frequency and amplitude), not just one as in the calculation of HRV. Unlike the R-R interval, in which each heartbeat is associated with a single real number, the proposed method associates each heartbeat with a set of points, that is, a vector.

The ECG signals were obtained in the databases MIT-BIH Normal Sinus Rhythm, MIT-BIH Atrial Fibrillation, and MIT-BIH Arrhythmia. The classifiers used to evaluate the proposed method were linear discriminant analysis, k-nearest neighbors, and support vector machine.

2 Materials and methods

The proposed method is based on the observation that arrhythmias episodes change the morphology of the ECG

signal. A block diagram is shown in Fig. 1, and each of the processing blocks is described in the following.

2.1 Datasets

The ECG signals were obtained from the databases (DB): MIT-BIH Normal Sinus Rhythm (NSR) database contains 18 ECG recordings of approximately 24 h duration. Subjects included in this database had no significant arrhythmias; they include 5 men, aged 26 to 45, and 13 women, aged 20 to 50 [18]; MIT-BIH Arrhythmia database contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects [19], and MIT-BIH AF contains 319 episodes of atrial fibrillation. The individual recordings have approximately a 10-h duration of 25 individuals [20].

Database records contain the rhythm types: atrial bigeminy, atrial fibrillation, atrial flutter, ventricular bigeminy, 2° heart block, idioventricular rhythm, normal sinus rhythm, nodal (A-V junctional) rhythm, paced rhythm, pre-excitation (WPW), sinus bradycardia, supraventricular tachyarrhythmia, ventricular trigeminy, ventricular flutter, and ventricular tachycardia.

2.1.1 Pre-processing

In the preprocessing step, the goal is to reduce contamination of different types of noise and artifacts in the ECG signal. Therefore, to perform this work, the following types of noise have been removed: a signal in the frequency of 60 Hz and its bandwidth below 1 Hz; baseline wander, a low-frequency (0.15 up to 0.3 Hz) noise that results from the patient inhaling and compels a baseline shifting

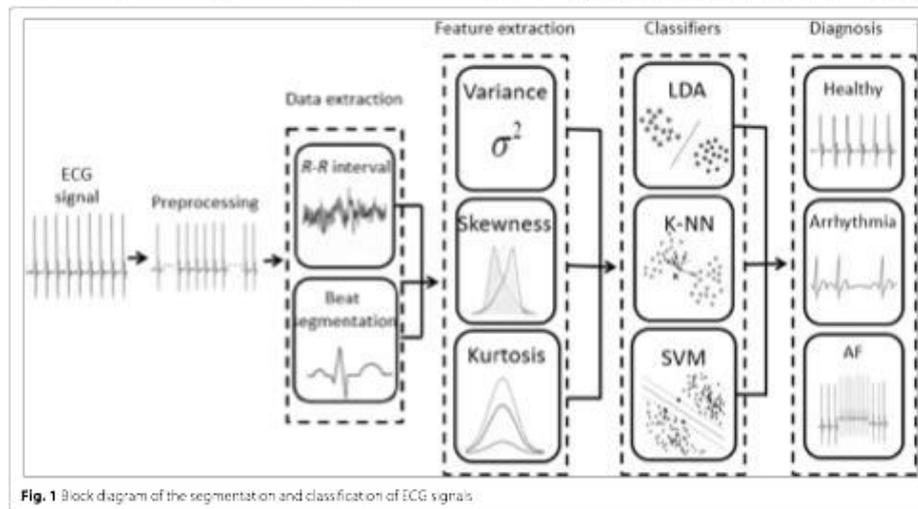


Fig. 1 Block diagram of the segmentation and classification of ECG signals

of the ECG signals; electrode contact noise, noise that results from a deficiency in the contiguity between the electrode and skin, which adequately cuts off the measurement system from the subject; electrode motion artifacts, artifacts that result from variations in the electrode-skin impedance with electrode motion; muscle contractions, noise that results from the contraction of other muscles apart from the heart; electrosurgical noise, noise produced from other medical apparatus in the patient care circumstance at frequencies between 100 and 1 MHz; and instrumentation noise, noise produced by the electronic equipment utilized in the ECG measurements [21].

From the records of the MIT-BIH NSR, MIT-BIH Arrhythmia, and MIT-BIH AF bases, as described in Section 2, 50,000 healthy heartbeat and 50,000 heartbeat of people with AF were withdrawn. Less than 1%, at the beginning and at the end, of the ECG signals were excluded due to measurement error. The ECG signal was normalized, and the sampling frequency was set to 128 Hz with 12-bit resolution in a range of ± 10 mV. Two or more cardiologists independently annotated each record; disagreements were resolved to obtain the computer-readable reference annotations for each beat (approximately 110,000 annotations in all) included with the database.

2.2 Data extraction of the heartbeat

The data extraction of the ECG signal proposed in this study is carried out by analyzing the voltage variation on each heartbeat and is given by [22]

$$B = (b_{start}, b_2, \dots, b_{end}), \quad (1)$$

where B is a heartbeat, b_{start} and b_{end} are given by

$$b_{start} = P_R - F_s \lambda, \quad (2)$$

and

$$b_{end} = P_R + F_s \theta, \quad (3)$$

where P_R is the position of the R peak (P_R are found in annotation files in MIT-BIH database), F_s is the sampling frequency, and λ and θ are the proportion weights of the heartbeat, being $\lambda + \theta \leq 1$. The parameters λ and θ are heuristically assigned and function as sliding windows on the heartbeat.

2.3 Method's generalization

The method presented in the previous Section 2.2 specifically analyzes heartbeat. Nevertheless, an arbitrary analysis to any segment of the heartbeat is defined as

$$X = (x_{start}, x_2, \dots, x_{end}), \quad (4)$$

where X is any segment of a heartbeat. If the segment of interest in the heartbeat begins before the peak of the R wave, Eq. 2 must be used, but if the segment of interest

starts after the peak of the R wave, Eq. 3 should be used. Both equations are heuristically adjusted. The end of an arbitrary segment in the ECG signal, from a specific part in a given heartbeat or even a succession of heartbeats, is given by

$$x_{end} = t_x F_s, \quad (5)$$

where t_x is the arbitrary period in the heartbeat or the ECG signal.

Another point of interest within the ECG is the peak of the P , QRS , and T waves. Therefore, the peak of the waves is defined as

$$P_X = \max(X). \quad (6)$$

Finally, a heartbeat can be defined as

$$B = P \cup PQ \cup QRS \cup ST \cup T = \quad (7)$$

$$= (p_{start}, p_2, \dots, p_P, \dots, p_{end}) \cup (pq_{start}, pq_2, \dots, pq_{end}) \quad (8)$$

$$\cup (qrs_{start}, qrs_2, \dots, P_{QRS}, \dots, qrs_{end}) \cup (st_{start}, st_2, \dots, st_{end}) \cup (t_2, \dots, P_T, \dots, t_{end}). \quad (10)$$

2.4 Feature extraction of the heartbeat

Modulation of the ECG signal can be performed in time (or frequency) and in amplitude (or energy). For frequency modulation, HRV is used, and for amplitude modulation, presented in Section 2.3, the voltage variation of the ECG signal was used. Thus, to classify different types of arrhythmias, we use two modulation information of the ECG signal, unlike several authors that only use the time modulation (HRV) [23–28]. The use of the two modulations allows a greater characterization of the ECG signal, improving the quality of the classification of arrhythmia. The variance, skewness, and kurtosis were used in this study to extract characteristics of both the ECG signal modulations.

- Variance

$$\sigma_X^2 = E(X^2) - (E(X))^2; \quad (11)$$

- Skewness

$$\gamma_X = E[(X - E(X))\sigma^{-1}]^3; \quad (12)$$

- Kurtosis

$$\kappa_X = E[(X - E(X))\sigma^{-1}]^4. \quad (13)$$

Proposed method will be evaluated in a generalist (arrhythmia classification) and specialist (AF classification) manner.

2.5 Performance evaluation

The new method of ECG data extraction will be evaluated based on the ECG window for arrhythmia classification.

Metrics for evaluation are specificity (SPEC—how efficient is the method for diagnosing healthy patients), sensibility (SENS—how efficient is the method for diagnosing patients with arrhythmias), and accuracy (ACC—how efficient is the method regarding the diagnosis).

The sensitivity and specificity are defined, respectively, given by

$$\text{SPEC} = \frac{TN}{TN + FP} \times 100. \quad (14)$$

and

$$\text{SENS} = \frac{TP}{TP + FN} \times 100, \quad (15)$$

And the accuracy is given by

$$\text{ACC} = \frac{TP + TN}{TP + TN + FN + FP} \times 100 \quad (16)$$

where TP is the true positive, TN is the true negative, FP is the false positive, and FN is the false negative.

3 Results

The experiments were conducted using 150,000 heartbeats (50,000 healthy heartbeats, 50,000 heartbeats of people with arrhythmia, and 50,000 heartbeats of people with AF). In the experiments, we used 80% of the patients for training (16 healthy patients, 41 patients with arrhythmia, and 20 patients with AF) and 20% of the patients for testing (2 healthy patients, 6 patients with arrhythmia, and 3 patients with AF). The classification of heartbeats was performed based on the following: the linear discriminant analysis (LDA) was used in its linear version; k-nearest neighbors (k-NN) ranged in up to 2–5 nearest neighbors, and support vector machine (SVM) used the polynomial kernel function. Heartbeat obtained according to Section 2.1 and the parameters used were shown in Table 1.

Table 2 shows the average accuracy of the classifiers (LDA, k-NN, SVM) for 4 different ECG windows (heartbeat, P wave, QRS complex, T wave).

Figure 2 illustrates the heartbeat, P wave, QRS complex, T wave, and PQ segment processed by the proposed

Table 1 Parameters Used in the ECG window

	Heartbeats (B)	Atrium (P)	AV node (PQ)	Ventricular depolarization (QRS)	Ventricular repolarization (T)
Time (ms)	1000	110	160	180	130
λ	0.6	0.066	0.096	0.108	0.078
θ	0.4	0.044	0.064	0.072	0.052
Start	0	0	21	26	66
End	80	20	25	60	80

λ is the proportion weight used from the start of the heartbeat, and θ is the proportion weight used from the end of the heartbeat

Table 2 ECG window by the method proposed in 5 different cases

Cases	Classifier	Average accuracy %			
		Heartbeat	P	QRS	T
Healthy	LDA	99.99	99.95	99.98	99.95
	k-NN	99.99	99.96	99.98	99.95
	SVM	99.99	99.96	99.98	99.96
Arrhythmia	LDA	99.97	99.01	99.93	99.03
	k-NN	99.97	99.03	99.94	99.04
	SVM	99.97	99.08	99.94	99.04
Atrial fibrillation	LDA	99.97	98.07	99.95	99.98
	k-NN	99.97	99.07	99.98	99.98
	SVM	99.99	99.03	99.98	99.07

LDA linear discriminant analysis, k-NN k-nearest neighbors, SVM support vector machine

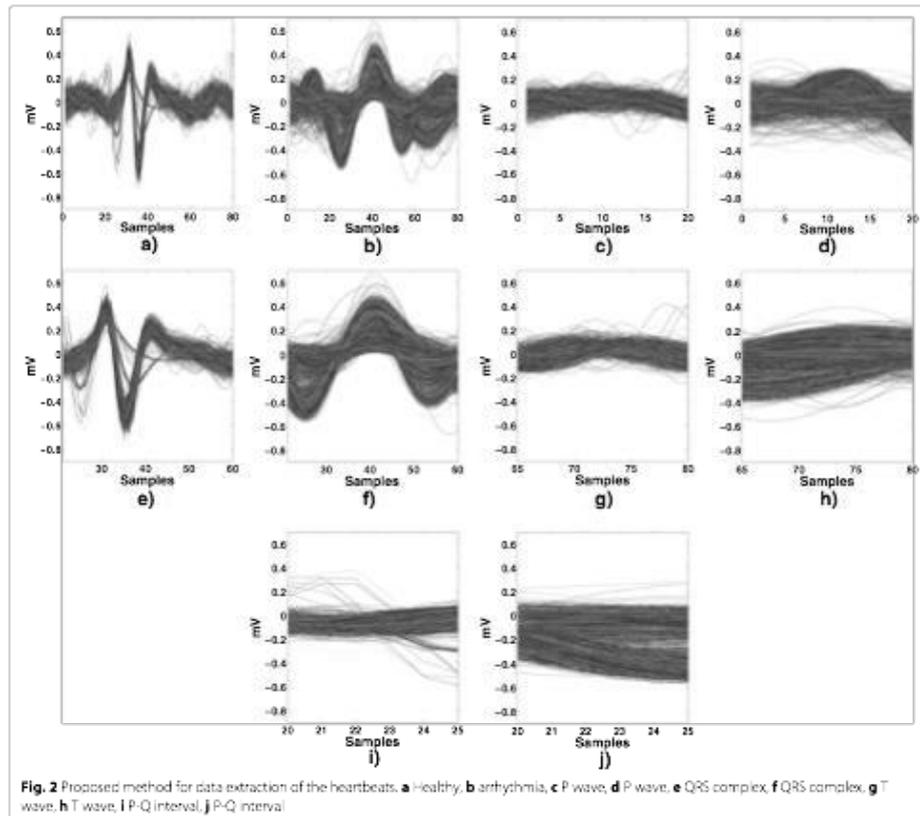
method according to the parameters described in the Table 1

Table 3 shows the average accuracy of the classifiers using variance, skewness, and kurtosis in the diagnosis of arrhythmia and AF. The results in this study are evaluated using 10-fold cross-validation, that is, definitive classification is the mean of the metrics (specificity, sensitivity, and accuracy) for ten different sets of training and testing.

Figure 3 illustrates the dispersion to features extracted by the proposed method for each heartbeat.

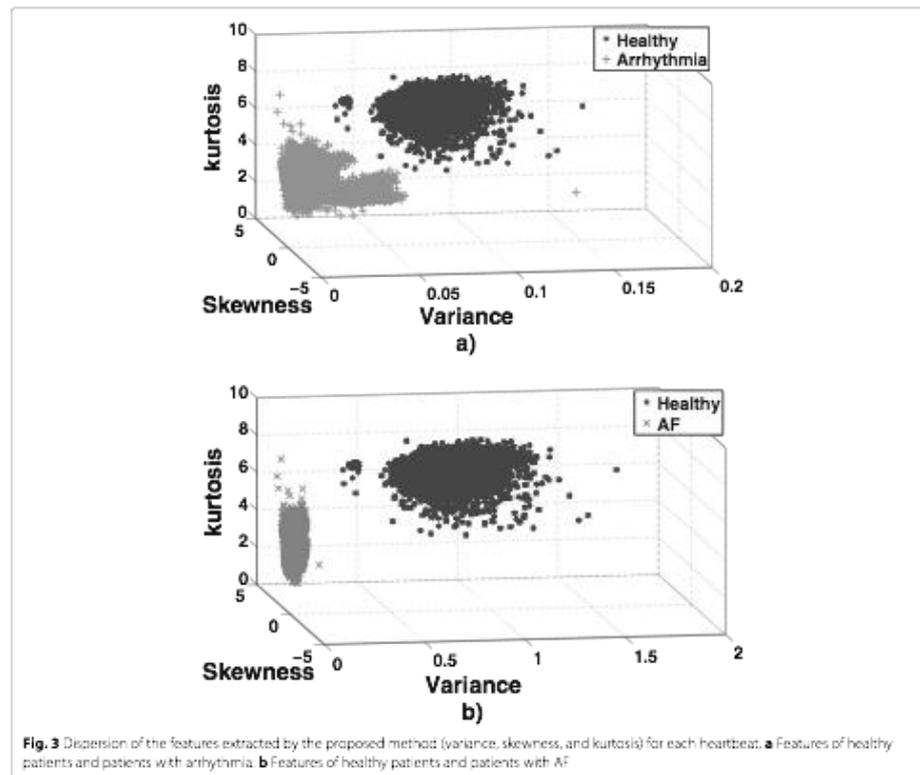
4 Discussion

ECG window was efficient in healthy heartbeat (Fig. 2a) in all evaluated cases (heartbeat, P wave, QRS complex, T wave, PQ segment), and this behavior is a consequence of signal uniformity; indeed, there is little or no difference in signal morphology, differently from the heartbeat with arrhythmia (Fig. 2b). P wave with arrhythmia (Fig. 2d) present deformations compared to healthy P wave (Fig. 2c), a consequence of the arrhythmia's associated with atrium (e.g., atrial fibrillation, atrial flutter) [29–32]. Ventricular diseases, such as heart block, cause irregularities in the electrical activity of the ventricles, which is evident in QRS complexes with arrhythmia (Fig. 2f), a characteristic not observed in healthy patients (Fig. 2e) [33–36]. The morphology of the T wave is completely apparent in healthy heartbeats; however, in heartbeats with arrhythmia, the T wave is deformed [37–40]. The PQ segment is completely visible in healthy heartbeats (Fig. 2i); nevertheless, due to increased heart rate, the PQ segments are reduced in heartbeats with arrhythmia (Fig. 2j), in some cases a consequence of the AV blocked [41].

**Table 3** Average accuracy of the classifiers (LDA, k-NN, SVM) using the proposed data extraction

	Average accuracy %							
	R-R	σ_x^2	γ_x	κ_x	$\sigma_x^2 + \gamma_x$	$\sigma_x^2 + \kappa_x$	$\gamma_x + \kappa_x$	$\sigma_x^2 + \gamma_x + \kappa_x$
Arrhythmia								
LDA	84.40	92.07	91.43	93.02	96.73	99.45	99.28	99.78
k-NN	84.40	92.07	91.43	93.02	96.73	99.45	99.28	99.78
SVM	84.40	92.07	91.43	93.02	96.73	99.45	99.28	99.78
Atrial fibrillation								
LDA	85.70	94.73	93.45	95.52	97.72	99.97	99.78	100
k-NN	85.70	94.73	93.45	95.52	97.72	99.97	99.78	100
SVM	85.70	94.73	93.45	95.52	97.72	99.97	99.78	100

LDA linear discriminant analysis, k-NN k-nearest neighbors, SVM support vector machine, σ_x^2 variance, γ_x skewness, κ_x kurtosis



The proposed method proved to be efficient in solving global (accuracy is up to 99.78% in the arrhythmia classification) and local (accuracy of 100% in the AF classification) heartbeat problems. The improvement in performance is due to the information obtained from the voltage variability and the morphology of the ECG signal, i.e., in addition to the frequency modulation information used in the HRV calculation, we added the amplitude modulation. The variation of k values (2–5 nearest neighbors) in the classification of the heartbeats was lower than 0.07%. The accuracy rate of the moments (variance, skewness, and kurtosis) when used separately is up to 9.82% and together is up to 15.38% when compared to the R-R interval (Table 3). The invariance in the average accuracy of the classifiers used (LDA, k-NN, SVM) is due to separability between healthy and with arrhythmia patients (Fig. 3a) and also between healthy and with AF patients (Fig. 3b).

The separability between healthy and those with arrhythmia patients (Fig. 3a), also between healthy and those with AF (Fig. 3b), is due to the information obtained from the voltage variability and the morphology of the ECG signal, i.e., in addition to the frequency modulation information used in the heart rate variability (HRV), we added the amplitude modulation. Thus, to classify different types of arrhythmias, we use two modulation information of the ECG signal (frequency and amplitude). In addition, the value in the right part of the image (cross red) corresponds to a single beat that is located at the beginning of the ECG signal. Of course, the 1% rejection window used in this work was not large enough to exclude this beat.

Table 4 compares the proposed methodology with the performance in the literature for arrhythmia and AF classification.

Table 4 Sensitivity, specificity, and accuracy of the proposed methodology compared to the literature for arrhythmia and AF classification, evaluated for MIT-BIH Normal Sinus Rhythm, MIT-BIH Arrhythmia, and MIT-BIH Atrial Fibrillation databases

Author (year)	Feature ^a	Classifier ^a	SENS %	SPEC %	ACC %
Arrhythmia					
This work (2018)	Voltage variation	LDA	99.64	99.91	99.78
		k-NN	99.64	99.91	99.78
		SVM	99.64	99.91	99.78
Mihandoust et al. (2018) [3]	Sparse decomposition	SVM	91.47	85.88	95.11
Raj et al. (2018) [42]	Spectral analysis	SVM, k-NN	91.47	85.88	95.11
Jović et al. (2017) [6]	AlphEn, HRV	Random forest	91.10	97.01	91.20
Kim et al. (2016) [7]	HRV from 5s	SVM	–	–	91.69
Ehaj et al. (2016) [8]	PCA, DWT, ICA, HOS	NN	98.90	98.90	98.90
Martis et al. (2012) [23]	DWT, HOS	NN	98.61	98.41	94.52
Atrial Fibrillation					
This work (2018)	P wave voltage variation	LDA	100	100	100
		k-NN	100	100	100
		SVM	100	100	100
Andersen et al. (2019) [24]	30 R-R	CNN, RNN	99.82	87.94	89.30
Xia et al. (2018) [25]	STFT, SWT	CNN	98.79	97.87	98.63
Kennedy et al. (2016) [26]	CoSEn+CV+RMSSD+MAD	Random forest	92.80	98.30	–
Orchard et al. (2016) [43]	P wave absence	Proposed algorithm	95.00	99.00	–
Petrenas et al. (2015) [27]	R-R	Threshold	97.10	98.30	–
Zhou et al. (2014) [28]	SD+SE	Threshold	97.53	98.26	98.16

The symbol (–) represent the values not specified in the works

AlphEn: alphabetic entropy, HRV: heart rate variability, PCA: principal component analysis, DWT: discrete wavelet transform, ICA: independent component analysis, HOS: higher order spectra, CoSEn: coefficient of sample entropy, CV: coefficient of variance, RMSSD: root mean square of the successive differences, MAD: median absolute deviation, 3D: symbolic dynamics, SE: Shannon entropy, RR: R-R intervals, STFT: short-term Fourier transform, SWT: stationary wavelet transform, LDA: linear discriminant analysis, k-NN: k-nearest neighbors, SVM: support vector machine, NN: neural network, CNN: convolutional neural network, RNN: recurrent neural network

Table 4 shows that even when compared with linear techniques such as SVM or nonlinear techniques like neural network, this study showed superior results in the arrhythmia classification (accuracy of 99.78%) and AF classification (accuracy of 100%). The improvement in performance is due to the information obtained from the voltage variability and the morphology of the ECG signal, i.e., in addition to the frequency modulation information used in the HRV calculation, we added the amplitude modulation. Thus, to classify different types of arrhythmias, we use two modulation information of the ECG signal (frequency and amplitude), not just one as in the calculation of HRV.

4.1 Limitation

The proposed data extraction is measured by statistical moments. The statistical moments are dependent on the average of the samples; actually, a large number of outliers can complicate the analysis of the data. In addition, the temporal information of the ECG signal can not be

obtained directly, that is, it is not trivial to verify changes in the heart rate by the proposed method.

5 Conclusions

The main implication of this study is the complete analysis of the heartbeat. The critical factor that improved performance was the information obtained from the voltage variability and ECG morphology rather than the classifiers (LDA, k-NN, SVM). Unlike the R-R interval, in which each heartbeat is associated with a single real number, the proposed method associates each heartbeat with a set of points, that is, a vector. This factor provides an information gain so that techniques for extracting characteristics based on signal statistics are able to obtain the presented results. Therefore, it is enough to know the duration time of the region of interest (Table 1). Furthermore, the simplicity of the proposed method allows for application in imbedded systems similar to the 24-h Holter. Thus, this system will be able not only to record the electrical activities of the heart and its variations, but will also provide

a prognosis for various arrhythmias. Our next challenge is to implement and evaluate the proposed method in real time.

Abbreviations

AlphaM: Alphabet entropy; CV: Coefficient of variance; ACC: Accuracy; AF: Atrial fibrillation; CoSEn: Coefficient of sample entropy; DWT: Discrete wavelet transform; ECG: Electrocardiogram; HOS: Higher order spectra; HRV: Heart rate variability; ICA: Independent component analysis; k-NN: k-Nearest neighbors; LDA: Linear discriminant analysis; MAD: Median absolute deviation; mV: voltage; NSR: Normal Sinus Rhythm; PCA: Principal component analysis; RMSSD: Root mean square of the successive differences; SD: Symbolic dynamics; SE: Shannon entropy; SENS: Sensibility; SPEC: Specificity; SVM: Support vector machine

Acknowledgements

We would like to thank PPGEE-UFPA for infrastructure development.

Funding

We would like to thank FAPEMA and CNPq for financial support.

Availability of data and materials

The database used in this study is public and can be found in <https://physionet.org/>

Authors' contributions

All authors contributed to the design and analysis of the paper and participated in the writing of the document. JAQ drafted the manuscript, carried out the basic literature review, and elaborated the method of the presented approach. LMAA developed the methods of the data evaluation. AKB reviewed the manuscript and evaluation of all the proposed methods. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors reviewed the manuscript critically. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 26 July 2018 Accepted: 4 February 2019

Published online: 04 March 2019

References

1. S. Mendis, P. Puska, B. Norving, World Health Organization, World Heart Federation, et al., *Global atlas on cardiovascular disease prevention and control* (S. Mendis, ed.) (World Health Organization, Geneva, 2011). <http://www.who.int/nmh/handle/10665/44707>
2. S. A. Khansari, *Guyton and Hall Textbook of Medical Physiology* (Department of Neurosurgery, Skull Base Laboratory, UCLA School of Medicine, Los Angeles, 2013). https://doi.org/10.4103/03065627.11327_17
3. S. Raj, K. C. Ray, Sparse representation of eeg signals for automated recognition of cardiac arrhythmias. *Expert Syst. Appl.* **105**, 49–64 (2018). <https://doi.org/10.1016/j.eswa.2018.03.038>
4. D. Ai, J. Yang, Z. Wang, J. Fan, C. A. Y. Wang, Fast multi-scale feature fusion for eeg heartbeat classification. *EURASIP J. Adv. Signal Process.* **2015**(1), 46 (2015). <https://doi.org/10.1186/s13634-015-0231-0>
5. H. Sheikhzadeh, R. L. Brennan, S. Sa, Real-time cardiac arrhythmia detection using wola filterbank analysis of eeg signals. *EURASIP J. Adv. Signal Process.* **2007**(1), 076256 (2007). <https://doi.org/10.1155/2007/76256>
6. A. Jovic, F. Jovic, Classification of cardiac arrhythmias based on alphabet entropy of heart rate variability time series. *Biomed. Signal Process. Control.* **31**, 217–230 (2017). <https://doi.org/10.1016/j.bspc.2016.08.010>
7. Y. J. Kim, J. Heo, K. S. Park, S. Kim, Proposition of novel classification approach and features for improved real-time arrhythmia monitoring. *Comput. Biol. Med.* **75**, 190–202 (2016). <https://doi.org/10.1016/j.combiomed.2016.06.009>
8. F. A. Elhaj, N. Salim, A. R. Harris, T. I. Swee, T. Ahmed, Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals. *Comput. Methods Prog. Biomed.* **127**, 52–63 (2016). <https://doi.org/10.1016/j.cmpb.2015.12.024>
9. A. Martinez, R. Arzand, J. J. Riera, Morphological variability of the p-wave for premature emission of paroxysmal atrial fibrillation events. *Physiol. Meas.* **35**(1), 1 (2014)
10. A. Petrenas, V. Marozas, A. Solistenko, R. Kubilius, J. Skarbaliene, J. Ozter, L. Sormmo, Electrocardiogram modeling during paroxysmal atrial fibrillation: application to the detection of brief episodes. *Physiol. Meas.* **38**(11), 2058 (2013)
11. A. Mashi, M. Mansour, J. N. Ruskin, E. K. Heitz, Impact of catheter ablation on p-wave parameters on 12-lead electrocardiogram in patients with atrial fibrillation. *J. Electrocardiol.* **47**(3), 725–733 (2014). <https://doi.org/10.1016/j.jelectrocard.2014.04.010>
12. I. Romero, J. Penders, C. Krietsels, P-wave analysis for atrial fibrillation detection in ambulatory recordings. *J. Electrocardiol.* **43**(6), 647 (2010). <https://doi.org/10.1016/j.jelectrocard.2010.10.022>
13. G. Ganeshaipala, J. Guttag, Real-time reconstruction of quasiperiodic multi-parameter physiological signals. *EURASIP J. Adv. Signal Process.* **2012**(1), 173 (2012). <https://doi.org/10.1186/1687-6180-2012-173>
14. I. Berast, I. Romero, Comparative study of algorithms for eeg segmentation. *Biomed. Signal Process. Control.* **34**, 166–173 (2017). <https://doi.org/10.1016/j.bspc.2017.01.013>
15. P. Fonseca, R. M. Aarts, J. Fousseir, X. Long, A novel low-complexity post-processing algorithm for precise qrs localization. *SpringerPlus* **3**(1), 376 (2014). <https://doi.org/10.1186/2199-1801-3-376>
16. H. Khamsi, R. Weiss, Y. Xu, C. W. Chang, N. H. Lovel, S. J. Redmond, Qrs detection algorithm for telehealth electrocardiogram recordings. *IEEE Trans. Biomed. Eng.* **63**(7), 1377–1386 (2016). <https://doi.org/10.1109/TBME.2016.2549060>
17. R. He, K. Wang, Q. Li, Y. Yuan, N. Zhao, Y. Liu, H. Zhang, A novel method for the detection of r-peaks in eeg based on k-nearest neighbors and particle swarm optimization. *EURASIP J. Adv. Signal Process.* **2017**(1), 82 (2017). <https://doi.org/10.1186/s13634-017-0519-3>
18. A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, H. E. Stanley, PhysioBank, physioToolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* **101**(23), 215–220 (2000)
19. A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, H. E. Stanley, PhysioBank, physioToolkit, and physionet. *Circulation* **101**(23), 215–220 (2000). <https://doi.org/10.1161/01CIR.101.23.e215>. <http://arxiv.org/abs/http://circphjournal.org/content/101/23/e215.full.pdf>
20. G. B. Moody, R. G. Mark, A new method for detecting atrial fibrillation using R-R intervals. *Comput. Cardiol.* **10**, 227–230 (1983)
21. S. K. Berkaya, A. K. Uysal, E. S. Gunal, S. Ergin, S. Gunal, M. B. Guimezoglu, A survey on eeg analysis. *Biomed. Signal Process. Control.* **43**, 216–235 (2018). <https://doi.org/10.1016/j.bspc.2018.03.001>
22. J. A. Queiroz, A. Junior, F. Lucena, A. K. Barros, Diagnostic decision support systems for atrial fibrillation based on a novel electrocardiogram approach. *J. Electrocardiol.* (2017). <https://doi.org/10.1016/j.jelectrocard.2017.10.014>
23. R. J. Martins, U. R. Acharya, C. M. Lim, K. M. Mandana, A. K. Ray, C. Chakraborty, Application of higher order cumulant features for cardiac health diagnosis using eeg signals. *Int. J. Neural Syst.* **23**(4), 1330014-1–1330014-19 (2013)
24. R. S. Andersen, A. Peimankar, S. Puthusserthy, A deep learning approach for real-time detection of atrial fibrillation. *Expert Syst. Appl.* **115**, 466–473 (2019). <https://doi.org/10.1016/j.eswa.2018.08.011>
25. Y. Xia, N. Wu, K. Wang, H. Zhang, Detecting atrial fibrillation by deep convolutional neural networks. *Comput. Biol. Med.* **93**, 84–92 (2018). <https://doi.org/10.1016/j.combiomed.2017.12.007>
26. A. Kennedy, D. D. Fray, D. Guldenring, R. B. Bond, K. Moran, J. McLaughlin, Automated detection of atrial fibrillation using r-r intervals and multivariate-based classification. *J. Electrocardiol.* **49**(6), 871–876 (2018). <https://doi.org/10.1016/j.jelectrocard.2016.07.033>
27. A. Petrenas, M. Valdovas, L. Sormmo, Low-complexity detection of atrial fibrillation in continuous long-term monitoring. *Comput. Biol. Med.* **1–8** (2015). <https://doi.org/10.1016/j.combiomed.2015.01.019>

28. X. Zhou, H. Ding, B. Ung, E. Pickwell-MacPherson, Y. Zhang, et al. Automatic online detection of atrial fibrillation based on symbolic dynamics and shannon entropy. *Biomed. Eng. Online* **13**(1), 18 (2014)
29. R. Linares, J. Igual. Exploiting periodicity to extract the atrial activity in atrial arrhythmias. *EURASIP J. Adv. Signal Process.* **2011**(1), 134 (2011). <https://doi.org/10.1186/1687-6180-2011-134>
30. M. Mešk, M. Diertoni, M. Marini, F. Ravelli. Characterization of rate and regularity of ventricular response during atrial tachyarrhythmias: insight on atrial and nodal determinants. *Physiol. Meas.* **30**(5), 800 (2017)
31. A. Kennedy, D. D. Finlay, D. Guldenring, R. R. Bond, J. McLaughlin. Detecting the elusive p-wave: A new ecg lead to improve the recording of atrial activity. *IEEE Trans. Biomed. Eng.* **63**(2), 243–249 (2016). <https://doi.org/10.1109/TBME.2015.2450212>
32. Y. Huo, F. Holmqvist, J. Carlson, T. Gaspar, G. Hindricks, C. Piskowski, A. Bollmann, P. G. Platonov. Variability of p-wave morphology predicts the outcome of circumferential pulmonary vein isolation in patients with recurrent atrial fibrillation. *J. Electrocardiol.* **40**(2), 218–225 (2015). <https://doi.org/10.1016/j.jelectrocard.2014.11.011>
33. V. Vancura, D. Wichterle, M. Brabec, J. Bytešník, K. Lefferová, J. Kautzner. The relationship between right ventricular pacing voltage and qrs complex duration. *Physiol. Meas.* **30**(5), 517 (2009)
34. D. C. Sigg, P. A. Iakoz, B. He. *Cardiac Electrophysiology Methods and Models*. Springer, New York, 2010. ISBN 978-1-4419-6657-5 e-ISBN 978-1-4419-6658-2 <https://doi.org/10.1007/978-1-4419-6658-2>
35. Q. A. Rahman, L. G. Tereshchenko, M. Kongkatong, T. Abraham, M. R. Abraham, H. Shaddy. Utilizing ecg-based heartbeat classification for hypertrophic cardiomyopathy identification. *IEEE Trans. NanoBioscience.* **14**(5), 505–512 (2015). <https://doi.org/10.1109/TNB.2015.2426213>
36. S. Lee, D. Park, K. H. Park. Qrs complex detection based on primitive. *J. Commun. Netw.* **19**(5), 442–450 (2017). <https://doi.org/10.1109/JCN.2017.000076>
37. S. A. Immanuel, A. Sadiq, M. Baumert, J. P. Couderc, W. Zareba, A. P. Hill, J. J. Vandenberg. T-wave morphology can distinguish healthy controls from lqts patients. *Physiol. Meas.* **37**(5), 1456 (2016)
38. A. Vimet, B. Dubé, R. Nadeau, O. Mankidine, V. Jacquemet. Estimation of the q-r relation: trade-off between goodness-of-fit and extrapolation accuracy. *Physiol. Meas.* **38**(3), 397 (2017)
39. G. T. Limes, B. L. de Oliveira, O. Skavhaug, M. M. Malecker. Simple t-wave metrics may better predict early ischemia as compared to st segment. *IEEE Trans. Biomed. Eng.* **64**(6), 1305–1309 (2017). <https://doi.org/10.1109/TBME.2016.2600198>
40. J. Ramirez, M. Orini, J. D. Tucker, E. Pueyo, P. Laguna. Variability of ventricular repolarization dispersion quantified by time-warping the morphology of the t-waves. *IEEE Trans. Biomed. Eng.* **64**(7), 1619–1630 (2017). <https://doi.org/10.1109/TBME.2016.2614899>
41. B. M. Gover, P. Brugada. *Clinical Handbook of Cardiac Electrophysiology*. Springer International Publishing, Brussels, 2016. <https://doi.org/10.1007/978-3-319-40818-7>
42. S. Mithrandros, M. C. Amirani. Cyclic spectral analysis of electrocardiogram signals based on garch model. *Biomed. Signal Proc. Control.* **31**, 79–88 (2017). ISBN 9783319408187 (online/9783319408163 (print)) <https://doi.org/10.1016/j.bspc.2016.07.012>
43. J. Orchard, N. Lowmes, S. B. Freedman, L. Ladak, W. Lee, N. Zwan, D. Peiris, Y. Kamaladasa, J. Li, L. Neubeck. Screening for atrial fibrillation during influenza vaccinations by primary care nurses using a smartphone electrocardiograph (ecg): A feasibility study. *Eur. J. Prev. Cardiol.* **23**(2, suppl), 13–20 (2016). <https://doi.org/10.1177/2047487316670255>. PMID: 27892421. <https://doi.org/10.1177/2047487316670255>

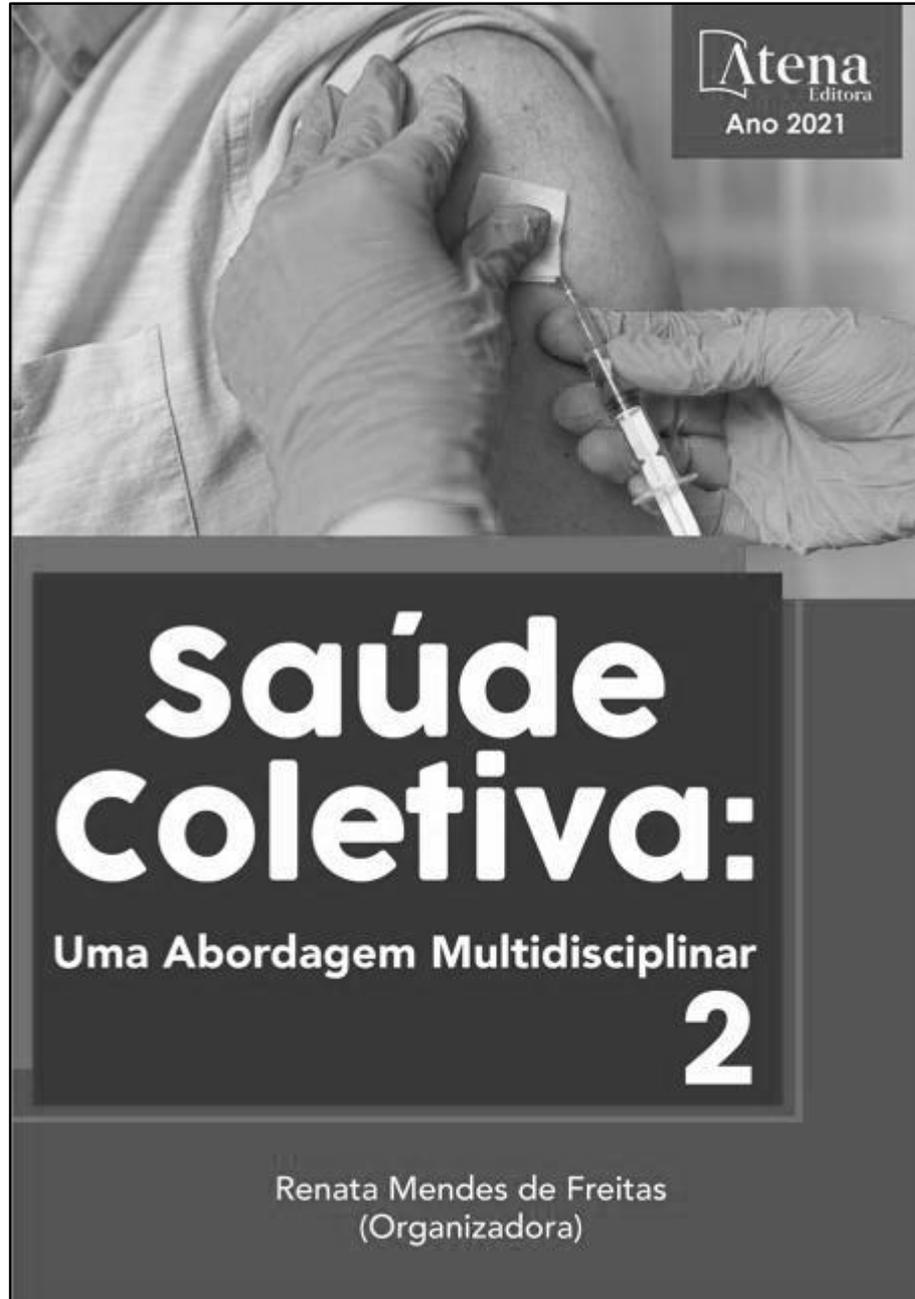
Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com

ANEXO F - CAPÍTULO DE LIVRO PUBLICADO

DOI: 10.22533/at.ed.94321220410



CAPÍTULO 10

CARACTERIZAÇÃO DO FENÓTIPO DA CINTURA HIPERTRIGLICERIDÊMICA EM PACIENTES RENAIIS ATENDIDOS EM UM HOSPITAL DE REFERÊNCIA

Data de aceite: 01/04/2021

Data de submissão: 04/02/2021

Terezinha de Jesus Vale Cantanhede
Curso de Farmácia, Faculdade Florence
São Luís - Maranhão

Cindy Lima Pereira
Universidade Federal do Maranhão
São Luís – Maranhão
<http://lattes.cnpq.br/4656862542152750>

Giselle Cutrim de Oliveira Santos
Universidade Federal do Maranhão
São Luís – Maranhão
<http://lattes.cnpq.br/1244895533719807>

Erika Cristina Ribeiro de Lima Carneiro
Hospital Universitário Presidente Dutra
São Luís – Maranhão
<http://lattes.cnpq.br/8296273562446833>

Luana Monteiro Anaisse Azoubel
Hospital Universitário Presidente Dutra
São Luís – Maranhão
<http://lattes.cnpq.br/9583257218358657>

Carlos Magno Sousa Junior
Universidade Federal do Maranhão
São Luís – Maranhão
<http://lattes.cnpq.br/9561853644051629>

Naruna Aritana Costa Melo
Universidade Estadual do Maranhão
Caxias – Maranhão
<http://lattes.cnpq.br/3534475456450489>

Talita Souza da Silva
Universidade Federal do Maranhão
São Luís – Maranhão
<http://lattes.cnpq.br/5033002356747250>

Maria Claudene Barros
Universidade Estadual do Maranhão
Caxias – Maranhão
<http://lattes.cnpq.br/5604314745118032>

Ewaldo Eder Carvalho Santana
Universidade Estadual do Maranhão
São Luís – Maranhão
<http://lattes.cnpq.br/0660692009750374>

Allan Kardec Duailibe Barros Filho
Universidade Federal do Maranhão
São Luís – Maranhão
<http://lattes.cnpq.br/0492330410079141>

Nílviane Pires Silva Sousa
Universidade Federal do Maranhão
São Luís – Maranhão
<http://lattes.cnpq.br/7098173750289255>

RESUMO: No Brasil, cerca de 60% das mortes são ocasionadas pelas doenças crônicas e há cerca de 2 milhões de brasileiros portadores da Doença Renal Crônica (DRC), em que 60% permanecem sem diagnóstico. Assim, destaca-se um método alternativo, associado a diversas comorbidades, o Fenótipo da Cintura Hipertrigliceridêmica (FCH), baseado na presença de Circunferência da cintura (CC) e os Triglicérides (TG) alterados. O FCH pode ser uma alternativa para triagem de comorbidades em pacientes renais. Este trabalho objetiva

avaliar a prevalência do FCH em pacientes atendidos em um hospital de referência ao tratamento e prevenção da DRC. Trata-se de um estudo transversal realizado com 121 pacientes de ambos os gêneros, acompanhados em um hospital de referência em São Luís-MA. Foram avaliadas variáveis antropométricas, bioquímicas e estilo de vida. O presente estudo possui aprovação do comitê de ética em pesquisa da UFMA, segundo o parecer 67030517.5.0000.5087. O grupo com o FCH apresentou maior média nas variáveis IMC, CC, PAD, LDLC e HDLC; ($p < 0,05$) que o grupo sem o fenótipo. Em relação, as frequências o grupo com FCH apresentou 31,3% ($n=25$) de excesso de peso, 39,5% ($n=15$) de CT elevado, 50% ($n=6$) glicemia de jejum elevada, 35,4% ($n=17$) HDLC alterado; ($p < 0,05$), quando comparado ao grupo sem o fenótipo. Acredita-se que FCH possa ser um parâmetro alternativo para avaliação da DRC e fatores de risco envolvido na progressão e tal ferramenta apresenta a vantagem de ser de baixo custo e fácil aplicação. Sugere-se para trabalhos futuros a validação dessa ferramenta na população estudada.

PALAVRAS - CHAVE: Cintura Hipertriglicéridêmica. Obesidade. Fatores de Risco. Doença Renal Crônica.

CHARACTERIZATION OF THE HYPERTRIGLICERIDEMIC WAIST PHENOTYPE IN RENAL PATIENTS ATTENDED AT A REFERRAL HOSPITAL

ABSTRACT: In Brazil, about 60% of deaths are caused by chronic diseases and there are about 2 million Brazilians with chronic renal disease (CRD), in which 60% remain undiagnosed. Thus, an alternative method stands out, associated with several comorbidities, the Hypertriglyceridemic Waist Phenotype (HWP), based on the presence of waist circumference (WC) and altered triglycerides (TG). HWP can be an alternative for screening comorbidities in renal patients. This study aims to evaluate the prevalence of HWP in patients seen at a hospital that is a reference in the treatment and prevention of CRD. This is a cross-sectional study conducted with 121 patients of both genders, followed up at a referral hospital in São Luís-MA. Anthropometric, biochemical and lifestyle variables were assessed. This study is approved by the UFMA Research Ethics Committee, according to the opinion 67030517.5.0000.5087. The group with the HWP showed the highest mean in the variables BMI, WC, PAD, LDLC and HDLC; ($p < 0,05$) than the group without the phenotype. In relation, the frequencies the group with HWP presented 31.3% ($n = 25$) of excess weight, 39.5% ($n = 15$) of elevated TC, 50% ($n = 6$) high fasting glucose, 35, 4% ($n = 17$) HDLC changed; ($p < 0,05$), when compared to the group without the phenotype. It is believed that HWP can be an alternative parameter for assessing CRD and risk factors involved in progression and such a tool has the advantage of being low cost and easy to apply. It is suggested for future work to validate this tool in the studied population.

KEYWORDS: Hypertriglyceridemic waist; Obesity; Risk factors; Chronic Renal Disease.

1 | INTRODUÇÃO

As doenças crônicas ultimamente vêm crescendo muito, em torno de 17 milhões de casos por ano, levando a numerosas mortes em todo mundo. No Brasil, uma faixa de 60% das mortes é ocasionada pelas doenças crônicas, estima-se que em média 2 milhões

de brasileiros são portadores da doença renal crônica (DRC), dos quais 60% não são diagnosticadas (MADEIRO *et al*, 2010).

A doença renal crônica é caracterizada pela perda progressiva e geralmente irreversível da filtração glomerular, diminuição da função ou lesões celular no rim, podendo ser leve, moderado, grave e até mesmo terminal. A doença renal crônica tem se tornado um grande problema de saúde pública e médica, já que vários casos são registrados de forma crescente anualmente, a DRC está associada a elevadas taxas de mortalidade, o que a torna um grande desafio para a comunidade científica (BONFIM *et al*, 2010).

Segundo Marinho *et al* (2017), no Brasil, a incidência e a prevalência da falência da função renal são crescentes e com um prognóstico ruim. Além dos custos altíssimos associados ao tratamento, cerca de 10% da população possui algum problema renal, enquanto 17% desconhece o problema, por ser geralmente silenciosa, não dando sinais ou sintomas, sobretudo no início.

Por ser uma doença de evolução lenta, regressa e de longa duração, exigindo um tratamento duradouro, pois a doença provoca uma série de fatores estressantes, mudanças no estilo de vida, redução da energia física, modificação da aparência pessoal e novas responsabilidades, ou seja, novas condições de vida, é de grande importância seu diagnóstico precoce (MACIEL *et al*, 2015).

Pacientes com DRC ou em risco da doença apresentam diversas comorbidades ou fatores de risco. No Brasil 20% da população são portadores de hipertensão arterial, 8% diabetes mellitus, 18% são tabagistas, enquanto 50% da população apresentam excesso de peso (GOUVEIA *et al*, 2017).

Independentemente do tratamento da doença renal, a taxa de mortalidade continua sendo maior, por conta de um diagnóstico tardio, por conta disso, adotar medidas de prevenção da doença, tais como identificação dos fatores e dos grupos de riscos, algo que facilite o diagnóstico precoce, tem conquistado uma atenção maior e mais investimentos (GOUVEIA *et al*, 2017).

Métodos de rastreamentos convencionais são muito importantes, tanto quanto encontrar um método alternativo que seja eficaz, de baixo custo e que esteja ao alcance da população, nessa situação o Fenótipo da Cintura Hipertrigliceridêmica (FCH), Circunferência da cintura (CC) e os Triglicérides (TG), demonstram ser uma ferramenta de rastreio para detecção desses pacientes com diabetes, doenças cardiovasculares, e doença renal crônica (RAMEZANKHANI *et al*, 2016).

Segundo Lemieux *et al* (2000) o Fenótipo Cintura Hipertrigliceridêmica (FCH) é caracterizada pela presença dos níveis séricos de triglicérides elevados e da circunferência da cintura aumentada. O FCH tem sido proposto como ferramenta de prognóstico da tríade metabólica aterogênica (hiperinsulinemia, níveis elevados de apolipoproteína B e concentrações aumentadas de partículas pequenas e densas da lipoproteína de baixa densidade (LDL-C), como uma possibilidade de diagnóstico da síndrome metabólica, e

como indicador de risco cardiovascular e metabólico, associado com obesidade visceral (LEMIEUX *et al.*, 2010).

Alguns estudos transversais sugerem que o Fenótipo Cintura Hipertrigliceridêmica pode ser um indicador simples e de grande confiança para prever o risco de doença renal crônica, e alguns fatores conhecidos da doença, tais como idade avançada, hipertensão, diabetes mellitus e obesidade (LI *et al.*, 2014).

Nesse contexto, é de suma importância o estudo dos fatores de risco associados à DRC, uma vez que essa situação traz um risco bem maior de mortalidade ocasionados por problemas cardiovasculares, Síndrome Metabólica e Diabetes Mellitus, podendo também ocasionar a redução da qualidade de vida dos indivíduos que a possuem. Portanto, medidas preventivas ou corretivas devem ser pensadas, para que o estado atual de saúde desses pacientes não sofra agravos e que assim se possa evitar o desenvolvimento de futuras morbidades.

Logo, o presente estudo tem como principal objetivo, avaliar a prevalência do FCH em pacientes atendidos em um hospital de referência ao tratamento e prevenção da DRC.

2.1 MÉTODO

Trata-se de uma pesquisa de campo, exploratória, de caráter transversal e quantitativo. A pesquisa foi realizada em centro de prevenção à doença renal na cidade de São Luís/Maranhão, entre os meses de junho a agosto de 2019. A amostra foi formada por 121 pacientes de ambos os sexos, com idade entre 20 a 65 anos, os pacientes foram escolhidos de forma não probabilística.

Os seguintes critérios foram adotados para escolha dos participantes, critérios de inclusão: ter entre 20 a 65 anos de idade; ter capacidade de responder verbalmente as perguntas do questionário. Foram excluídos da amostra: participantes que apresentem incapacidade física que impossibilite ou comprometa a coleta; mulheres grávidas ou lactantes; apresentaram doença que comprometesse a cognição ou prejudicasse sua tomada de decisão, como Alzheimer ou problemas psicológicos.

Os dados sociodemográficos e de estilo de vida foram coletados através de questionário semiestruturado, esses dados foram autodeclarados pelos participantes do estudo e incluem: gênero, prática de exercício físico e uso de cigarro.

As medidas foram realizadas por um único pesquisador, com o auxílio de instrumentos calibrados e, para análise dos dados, obtiveram-se medidas em duplicata. O peso foi aferido em balança eletrônica calibrada (Omron® HBF 214 LA, Japão) com resolução de 0,1 kg. A altura foi medida através de estadiômetro transportável vertical com resolução de 0,1 cm (Sanny®, Brasil). A circunferência da cintura (CC) foi medida através de trena antropométrica inelástica com precisão de 0,1 cm (Seca® 213, Hamburg, Alemanha), a partir do ponto médio entre a crista ilíaca e o último rebordo costal, na respiração mínima.

Para classificação da CC, utilizou-se como ponto de corte os valores publicados pela Associação Brasileira para o Estudo da Obesidade e da Síndrome Metabólica – ABESO (2016), tendo por base valores propostos pela IDF, no qual o ponto de corte adotado para homens foi de ≥ 90 cm e para mulheres foi de ≥ 80 cm. Para análise do estado nutricional foi utilizado a fórmula do índice de massa corporal (IMC): $\text{peso} / \text{altura}^2$, onde os pacientes foram divididos em dois grupos: com excesso de peso (sobrepeso e obeso) e eutrófico (IMC normal), conforme a classificação da Organização Mundial da Saúde (OMS) (ABESO, 2016).

As análises bioquímicas foram realizadas pelo Laboratório de Análises Clínicas do Hospital Universitário Presidente Dutra – HUUFMA, no aparelho automatizado Cobas 6000 Roche®, a metodologia utilizada na dosagem seguiu o roteiro preconizado pelo manual do fabricante. A coleta de sangue foi realizada por meio do sistema de multiplas coletas a vácuo, após o período de 12 horas em jejum. O ponto de corte para glicemia em jejum (GJ) seguiu a determinação da Diretriz da Sociedade Brasileira de Diabetes 2019/2020 - SBD (2019).

O *Low Density Lipoprotein* (LDLC) foi calculado pela fórmula de Friedewald considerando valor de triglicérides <400 mg/dL, caso essa prerrogativa não fosse atendida utilizou-se a fórmula de Martin, conforme preconizado pela Diretriz Brasileira de Dislipidemias e Prevenção da Aterosclerose (SBC, 2017). A Tabela 1 traz a classificação do Colesterol total, *High Density Lipoprotein* (HDLC) e Triglicérides (TG) com base no preconizado pela Sociedade Brasileira de Cardiologia - SBC (2017). Quanto ao LDLC, foram avaliados somente seus valores médios na população estudada, devido necessidade de avaliação do risco cardiovascular para posterior sugestão de alvos terapêuticos.

Lípides	Com jejum (mg/dL)	Categoria Referencial
Colesterol Total	<190	Desejável
HDLC	> 40	Desejável
Triglicérides	< 150	Desejável

Abreviação: HDLC - *High Density Lipoprotein*.

Tabela 1. Valores de referência para o perfil lipídico de acordo com a Diretriz Brasileira de Dislipidemias e Prevenção da Aterosclerose (SBC, 2017).

Os indivíduos foram classificados como portadores do FCH se apresentassem simultaneamente a CC e TG elevados.

- **Sem FCH:** CC adequado e TG adequado; CC inadequado e TG adequado; CC adequado e TG inadequado;
- **Com FCH:** CC inadequado e TG inadequado.

A CH que se baseou nos pontos de corte de CC e TG já descritos na metodologia (MADEIRO *et al.*, 2010).

A pressão arterial sistólica (PAS) e a pressão arterial diastólica (PAD) foram aferidas com auxílio de aparelho monitor de pressão arterial de braço (OMRON[®], modelo HEM 7130). O procedimento para aferição foi iniciado após repouso do participante com duração de 3 a 5 minutos. Realizou-se a aferição em triplicata, com intervalo de um minuto entre as medições, na qual o participante estava sentado, pernas descruzadas, pés apoiados no chão, dorso recostado na cadeira e relaxado, conforme procedimento descrito na VII Diretriz Brasileira de Hipertensão Arterial (SBC, 2016). Para o arquivo de dados e a análise estatística, foi utilizado o *software SPSS (Statistical Package for the Social Sciences, Inc., Chicago, IL, USA) versão 19.0*. Os dados tratados por meio de procedimentos descritivos (média ou mediana e desvio padrão). O teste *Kolmogorov-Smirnov* foi usado para verificar a normalidade dos dados. Para as comparações entre os grupos utilizou-se o teste *t de Student* para amostras independentes no caso de distribuição normal das variáveis, ou o teste de *Mann-Whitney U*, para as variáveis que não apresentarem distribuição normal dos dados. Para avaliação da prevalência das variáveis categóricas entre os grupos foi utilizado o teste Qui-quadrado. Os resultados serão considerados estatisticamente significativos para $p < 0,05$.

O presente estudo possui aprovação do Comitê de Ética e Pesquisa da Universidade Federal do Maranhão CAAE: 67030517.5.0000.5087 Os voluntários foram incluídos após serem informados, em detalhes, sobre o estudo e preencherem corretamente o termo de Consentimento Livre e Esclarecido.

3 | RESULTADOS E DISCUSSÃO

De um total de 121 pacientes selecionados, verificou-se que 78,5% (n=95) não apresentaram o fenótipo cintura hipertrigliceridêmica (FCH). A média de idade dos pacientes com fenótipo foi de $42,7 \pm 12,1$ anos, a tabela 2 que apresenta a caracterização da amostra avaliada por grupo. As variáveis IMC, CC, RCE, PAD, LDLC, HDLC e glicemia de jejum apresentaram as maiores médias no grupo com o FCH (Tabela 2).

Variáveis	Sem o fenótipo (n=95)	Com o fenótipo (n=26)	p-valor
Idade (anos)	39,9 ± 14,4	42,7 ± 12,1	0,316
IMC (kg/m ²)	26,2 ± 5,1	30,8 ± 4,7	<0,001
CC (cm)	81,6 ± 12,1	96,1 ± 8,1	<0,001
RCE	0,52 ± 0,01	0,60 ± 0,05	<0,001
PAS (mmHg)	118,5 ± 21,5	126 ± 17,7	0,080
PAD (mmHg)	75,8 ± 12,7	81,6 ± 9,3	0,019
LDL-C (mg/dL)	96,1 ± 31,0	129 ± 38,2	0,001
HDL-C (mg/dL)	47,2 ± 15	38,2 ± 10	0,002
Colesterol total (mg/dL)	162,8 ± 40,8	206 ± 41,6	0,873
Glicemia de jejum (mg/dL)	81,5 ± 13,1	98 ± 34,5	<0,001

Abreviações: IMC- índice de massa corporal; CC- circunferência da cintura; RCE- relação cintura estatura; PAS-pressão arterial sistólica; PAD – Pressão arterial diastólica; GJ- Glicemia de jejum; HDLC - *High Density Lipoprotein*; LDLC- *Low Density Lipoprotein*. Valores são apresentados como média ± desvio padrão (teste t-amostras independentes).

Tabela 2. Caracterização da amostra estratificada através da presença ou não do fenótipo

Fonte: Os autores, 2020.

Já Tabela 3, descreve a frequência dos fatores de risco cardiometabólico por grupo com e sem o FCH, onde se observa que somente as variáveis: estado nutricional, glicemia de jejum, colesterol total e HDLC apresentaram diferença significativa.

Variáveis	Com o fenótipo (n=26)	Sem o fenótipo (n=95)	p-valor
Gênero^a			
Mulheres	24 (06)	76 (19)	0,460
Homens	20,8 (20)	79,2 (76)	
Cor^a			
Caucasiano	21,4 (06)	78,6 (22)	0,610
Não caucasiano	21,5 (20)	78,5 (73)	
Exercício^a			
Não	20,3 (15)	79,7 (59)	0,424
Sim	23,4 (11)	76,6 (36)	
Fumo^a			
Não	22 (26)	78 (92)	0,481
Sim	-	100 (03)	
Estado Nutricional^a			
Eutrófico	2,4 (01)	97,6 (40)	<0,001
Excesso de peso	31,3 (25)	68,8 (55)	
Colesterol Total^a			
Normal	13,3 (11)	86,7 (72)	0,002
Elevado	39,5 (15)	60,5 (23)	
Glicemia em jejum^a			
Normal	18,3 (20)	81,7 (89)	0,021
Aterada	50 (06)	50 (06)	
HDL-C^a			
Normal	12,3 (09)	87,7 (64)	0,003
Aterado	35,4 (17)	64,6 (31)	

^aValores são apresentados como percentual (frequência). ^ateste qui-quadrado. Abreviatura: *High Density Lipoprotein* (HDL).

Tabela 3. Prevalência dos fatores de risco cardiometabólico estratificado por grupo.

Fonte: Os autores, 2020.

O fenótipo da Cintura Hipertrigliceridêmica, foi proposto com a intenção de identificar pacientes que possuem a tríade metabólica aterogênica (hiperinsulinemia, níveis elevados de apo B e de partículas pequenas e densas de colesterol LDL) que geralmente é utilizada para prever os riscos cardiovasculares e doenças crônicas como a doença renal (SANTOS,

2016). Segundo Cabral *et al* (2012), em uma pesquisa feita no Peru, constatou-se que os pacientes com a presença do fenótipo apresentam 3 vezes mais chances de desencadear Doenças Cardiovasculares e Doença Arterial Coronariana (DAC).

A doença renal crônica (DRC) é um grande problema de saúde pública com alta taxa de morbidade e mortalidade em todo o mundo. A prevalência global estimada de DRC é de 13,4% (11,7% –15,1%), afetando 26–30 milhões de adultos nos Estados Unidos, por exemplo (JUSZCZAK *et al*, 2020). Dessa forma, é de grande auxílio para os profissionais de saúde o uso de indicadores de baixo custo e fácil aplicação que avaliem a presença de DRC ou contribuam para sua prevenção, onde se destaca a Cintura Hipertrigliceridêmica (MADEIRO *et al*, 2010).

O FCH é tido como um marcador que prediz a Doença Renal Crônica, no presente estudo a prevalência do FCH por gênero foi de 24% no sexo feminino e 20,8% no sexo masculino. Em seu estudo, Freitas (2016), também encontrou uma maior prevalência do fenótipo no sexo feminino de 56,9%. Oliveira *et al* (2014), observou uma maior prevalência desse fenótipo também em pacientes do sexo feminino. Logo, estudos futuros devem avaliar a influência do gênero na presença do FCH.

A apesar de não apresenta diferença significativa, cabe chamar atenção, que a amostra de pacientes com FCH, teve a média de idade maior que o grupo sem FCH. A idade avançada é um fator que coopera para desencadear a doença renal e pacientes com idade mais avançada são predominantes quanto à presença de FCH, devido a fatores como: alterações fisiológicas, mudança na composição corporal, perda de massa magra, acúmulo de gordura na região abdominal, o que aumenta também a taxa de doenças crônicas nesses pacientes (RIM *et al*, 2015).

A DRC está associada a distúrbios fisiológicos e metabólicos de prevalência mundial elevada, como hipertensão, obesidade, resistência à insulina, doenças cardiovasculares e envelhecimento, que também são fatores de risco para a patogênese e progressão da DRC (JUSZCZAK *et al*, 2020). Nesse contexto, Mota *et al* (2016) observou uma prevalência do FCH em 30% dos homens hipertensos da sua amostra, em contrapartida de 26,3% de mulheres hipertensas com FCH. A hipertensão arterial sistêmica expõe o paciente a uma grande chance de apresentar o FCH, devido as alterações metabólicas envolvidas no desenvolvimento dessa patologia.

Além das associações com a tríade metabólica, o FCH está associado a diversos fatores de risco cardiometabólico como: diabetes, hiperglicemia, obesidade, elevação nos níveis de colesterol total e diminuição no colesterol HDL. Apesar de ser um parâmetro recente e até desconhecido, há evidências que o FCH pode ser tão discriminante quanto à síndrome metabólica (SM) para alterações metabólicas como a diabetes e resistência à insulina (FREITAS *et al*, 2018). Tal ponto é de grande relevância já que a DM, é uma das doenças de base responsável por cerca 70% dos casos de DRC (JUNIO *et al*, 2017).

Um outro ponto importante relatado por Andrade *et al* (2017) é que a falta da prática

2016). Segundo Cabral *et al* (2012), em uma pesquisa feita no Peru, constatou-se que os pacientes com a presença do fenótipo apresentam 3 vezes mais chances de desencadear Doenças Cardiovasculares e Doença Arterial Coronariana (DAC).

A doença renal crônica (DRC) é um grande problema de saúde pública com alta taxa de morbidade e mortalidade em todo o mundo. A prevalência global estimada de DRC é de 13,4% (11,7% –15,1%), afetando 26–30 milhões de adultos nos Estados Unidos, por exemplo (JUSZCZAK *et al*, 2020). Dessa forma, é de grande auxílio para os profissionais de saúde o uso de indicadores de baixo custo e fácil aplicação que avaliem a presença de DRC ou contribuam para sua prevenção, onde se destaca a Cintura Hipertrigliceridêmica (MADEIRO *et al*, 2010).

O FCH é tido como um marcador que prediz a Doença Renal Crônica, no presente estudo a prevalência do FCH por gênero foi de 24% no sexo feminino e 20,8% no sexo masculino. Em seu estudo, Freitas (2016), também encontrou uma maior prevalência do fenótipo no sexo feminino de 56,9%. Oliveira *et al* (2014), observou uma maior prevalência desse fenótipo também em pacientes do sexo feminino. Logo, estudos futuros devem avaliar a influência do gênero na presença do FCH.

A apesar de não apresenta diferença significativa, cabe chamar atenção, que a amostra de pacientes com FCH, teve a média de idade maior que o grupo sem FCH. A idade avançada é um fator que coopera para desencadear a doença renal e pacientes com idade mais avançada são predominantes quanto à presença de FCH, devido a fatores como: alterações fisiológicas, mudança na composição corporal, perda de massa magra, acúmulo de gordura na região abdominal, o que aumenta também a taxa de doenças crônicas nesses pacientes (RIM *et al*, 2015).

A DRC está associada a distúrbios fisiológicos e metabólicos de prevalência mundial elevada, como hipertensão, obesidade, resistência à insulina, doenças cardiovasculares e envelhecimento, que também são fatores de risco para a patogênese e progressão da DRC (JUSZCZAK *et al*, 2020). Nesse contexto, Mota *et al* (2016) observou uma prevalência do FCH em 30% dos homens hipertensos da sua amostra, em contrapartida de 26,3% de mulheres hipertensas com FCH. A hipertensão arterial sistêmica expõe o paciente a uma grande chance de apresentar o FCH, devido as alterações metabólicas envolvidas no desenvolvimento dessa patologia.

Além das associações com a tríade metabólica, o FCH está associado a diversos fatores de risco cardiometabólico como: diabetes, hiperglicemia, obesidade, elevação nos níveis de colesterol total e diminuição no colesterol HDL. Apesar de ser um parâmetro recente e até desconhecido, há evidências que o FCH pode ser tão discriminante quanto à síndrome metabólica (SM) para alterações metabólicas como a diabetes e resistência à insulina (FREITAS *et al*, 2018). Tal ponto é de grande relevância já que a DM, é uma das doenças de base responsável por cerca 70% dos casos de DRC (JUNIO *et al*, 2017).

Um outro ponto importante relatado por Andrade *et al* (2017) é que a falta da prática

de atividade física está associada a uma alta prevalência do fenótipo, o estudo também constatou que o risco de obesidade está associado ao sedentarismo, além disso verificou que os pacientes que apresentaram o FCH, eram sedentários e ainda faziam uso frequente do cigarro.

Já a presença de excesso de peso, avaliado através do Índice de Massa Corporal (IMC), está devidamente associado ao FCH. Amini *et al* (2011), ao estudarem uma amostra de pacientes no Irã, constataram que os pacientes com a presença do fenótipo, possuem maiores índices de IMC, isso porque indivíduos com o fenótipo, possuem obesidade global, gordura visceral aumentada, e isso também se justifica pelo fato do IMC ter uma forte relação com a CC alterada.

Deste modo, o acúmulo exagerado de gordura corporal leva a um desequilíbrio dos lipídeos sanguíneos, que podem resultar em acúmulo no fígado, músculo e no próprio tecido adiposo, além disso, pode gerar consequências como: elevação da pressão arterial, alterações metabólicas, risco cardiovascular elevado, que estão associados a alterações renais hemodinâmicas, que ocasiona desordens metabólicas e bioquímicas, que pode predispor a doença renal, mesmo com a função renal estando normal nos exames convencionais (OLIVEIRA *et al*, 2014).

A doença renal está relacionada a vários fatores tais como, as doenças cardiovasculares, hipertensão e diabetes melítus, pois o acúmulo desses fatores podem aumentar a probabilidade de outras lesões, ou até mesmo progredir a doença, para isso o rastreamento pelo FCH é uma ferramenta de suma importância, pois tem potencial para triagem de pacientes com risco da doença, ajudando então a pensar em medidas de controle e tratamento desses pacientes (BREGMAN, 2006; MARTINS, 2017), visto que é um indicador simples e de baixo custo, que inclusive pode ser utilizado na rotina do farmacêutico como meio de detecção e controle da DRC principalmente em pacientes assintomáticos.

4 | CONCLUSÃO

O FCH pode ser uma ferramenta útil para triagem de pacientes com doença renal, bem como, de outros fatores de risco associados à sua progressão. Ele torna-se uma boa alternativa, por ser uma ferramenta de baixo custo, fácil de usar, reproduzível, podendo ser usada na prática clínica preventiva, ou até mesmo nos serviços públicos de atenção primária a saúde até mesmo em lugares remotos devido aos sistemas *point of care* que já dosam com sangue capilar o triglicérideo. Adicionalmente, a análise desse parâmetro também pode auxiliar na implementação de políticas públicas para o enfrentamento das doenças crônicas não transmissíveis. Para estudos futuros deixa-se como sugestão a validação dessa ferramenta nessa população.

ANEXO G - ARTIGO PUBLICADO

Journal of Environmental Research and Public Health, fator de impacto: 3,39



Article

Development of a Computational Model to Predict Excess Body Fat in Adolescents through Low Cost Variables

Carlos Magno Sousa ¹, Ewaldo Santana ^{1,2,*}, Marcus Vinicius Lopes ¹, Guilherme Lima ¹, Luana Azoubel ³, Érika Carneiro ³, Allan Kardec Barros ¹ and Nilviane Pires ¹

¹ Department of Electrical Engineering, Biological Information Processing Lab, Federal University of Maranhão, São Luis 65085680, MA, Brazil

² Laboratory of Signals Acquisition and Processing, LAPS, State University of Maranhão, Campus Paulo VI, São Luis 65700000, MA, Brazil

³ Centro de Prevenção de Doenças Renais, University Hospital of Maranhão, São Luis 65080805, MA, Brazil

* Correspondence: ewaldoeder@gmail.com; Tel.: +55-98-3272-8372

Received: 9 July 2019; Accepted: 14 August 2019; Published: 17 August 2019



Abstract: Background: Excess body fat has been growing alarmingly among adolescents, especially in low income and middle income countries where access to health services is scarce. Currently, the main method for assessing overweight in adolescents is the body mass index, but its use is criticized for its low sensitivity and high specificity, which may lead to a late diagnosis of comorbidities associated with excess body fat, such as cardiovascular diseases. Thus, the aim of this study was to develop a computational model using linear regression to predict obesity in adolescents and compare it with commonly used anthropometric methods. To improve the performance of our model, we estimated the percentage of fat and then classified the nutritional status of these adolescents. Methods: The model was developed using easily measurable socio-demographic and clinical variables from a database of 772 adolescents of both genders, aged 10–19 years. The predictive performance was evaluated by the following metrics: accuracy, sensitivity, specificity, and area under ROC curve. The performance of the method was compared to the anthropometric parameters: body mass index and waist-to-height ratio. Results: Our model showed a high correlation ($R = 0.80$) with the body fat percentage value obtained through bioimpedance. In addition, regarding discrimination, our model obtained better results compared to BMI and WHtR: AUROC = 0.80, 0.64, and 0.55, respectively. It also presented a high sensitivity of 92% and low false negative rate (6%), while BMI and WHtR showed low sensitivity (27% and 9.9%) and a high false negative rate (65% and 53%), respectively. Conclusions: The computational model of this study obtained a better performance in the evaluation of excess body fat in adolescents, compared to the usual anthropometric indicators presenting itself as a low cost alternative for screening obesity in adolescents living in Brazilian regions where financial resources are scarce.

Keywords: obesity; adolescent; screening

1. Introduction

The prevalence of excess body fat has been growing alarmingly worldwide, this increase is also observed among teenagers [1,2]. In certain developed, the prevalence of obesity among adolescents has achieved high levels [3,4], as in the USA, that 17% of teenagers are considered obese [4]. Statistics are also alarming in developing countries like Brazil, since the rate of obesity also grows rapidly, that nearly 20% of teenagers are obese [1]. Consequently, the assessment of nutritional status plays a relevant role in the fluctuations in the body composition of the individual, as well as in the survival

rate under pathological conditions, since it allows the early diagnosis of comorbidities associated with overweight such as cardiovascular disease and endocrine disorders [5,6].

There are several techniques for assessing body composition, among which bioimpedance (BIA) stands out, which is a fast, portable, and non-invasive method that uses the principle of electrical impedance. Moreover, it has little technical difficulty and high correlation with dual energy X-ray absorptiometry (DXA), and its use is indicated in epidemiological studies and clinical practice [7–11]. However, the use of anthropometric parameters is still more viable due to their low cost when compared to BIA.

The most commonly used anthropometric indicator in the assessment of adolescent nutritional status is the body mass index (BMI) due to its low cost, simplicity, and high reproducibility. However, BMI has controversies regarding its efficiency, since it has low sensitivity in predicting excess body fat [12–16]. Individuals who have a high percentage of lean mass will have their total weight affected, and on BMI assessment, this individual will be mistakenly classified as obese.

Thus the identification of individuals with high body fat percentage subclinical is an important measure to identify individuals who need earlier interventions and especially for adolescents. Early detection of excess body fat through low cost and high sensitivity methods allows the implementation of appropriate therapeutic approaches to mitigate the development of comorbidities associated with excess body fat [17]. Therefore, we propose a statistical method to predict obesity in adolescents, using low cost and easily verifiable variables (such as gender, age, height, and body mass).

2. Materials and Methods

2.1. Construction of The Database

The database is composed of 722 adolescents. The study included adolescents of both genders, aged 10 to 19 years old, duly enrolled in public schools in São Luis, MA. Exclusion criteria were: pregnancy, lactation, use of birth control pills, being on menstrual period, eating disorders, malnutrition, dehydration, body water retention, refusal to participate in the study, and absenteeism in collections. A single researcher performed each measurement with the same calibrated instrument. Based on the protocol by Lohman [18], the measurements were performed in duplicate, in a single moment (transversal study), and for data analysis, the mean values of the collected measurements were calculated. The variables evaluated were socio-demographic (age and gender) and anthropometric (body mass, height, body circumferences, and body fat percentage). The present study is approved by the Ethics Committee in Research with Human Beings of the Federal University of Maranhão according to legal opinion CAAE: 83206118.1.0000.5087.

2.2. Sample Calculation

The sample size was calculated by proportion estimation [19], based on the prevalence of overweight in adolescents of 20.5% [20], which suggested a prevalence of outcome of 26.9% [21], tolerable error of 5% (type I error), and test power of 90% (Type II error), with 10 % added for possible losses or refusals. A minimum sample of 513 teenagers was reached.

2.3. Data Collect

Body mass was measured with a calibrated electronic scale (Omron® HBF 214 LA, Japan) with a precision of 0.1 kg. Height was determined through a portable vertical stadiometer with precision of 0.1 cm (Sanny®, Brazil). For the measurement of circumferences, an anthropometric tape measure with 0.1 cm precision was used (Seca® 213, Hamburg, Germany). Waist circumference was measured at the midpoint between the last rib and the iliac crest at minimum respiration [22]. Hip circumference was measured at the widest point as described by Huang et al. [23]. Arm circumference was evaluated at midpoint between olecranon and acromial process on the upper left-arm with the subject in standing

position [24], and the calf was measured with the participant seated, knee bent at a ninety-degree angle to the floor and this was considered the point of largest calf circumference [25].

The BMI was calculated according to the index of weight (kg)/height² (m), and the cut-off point used for classification of nutritional status was the one recommended by the World Health Organization for gender and age, where >3 percentile and <85 percentile was classified as eutrophic and >85 percentile was overweight [26,27]. The waist-to-height ratio (WHtR) was calculated using the index of waist circumference (cm)/height (cm). For WHtR evaluation, the cutoff point used was 0.5 for both genders and age [28]. Age was calculated as the difference between the date of birth and the date of measurement, and the gender and ethnicity was self-declared by the participant.

The percentage of body fat was obtained through the (calibrated) tetrapolar bioimpedance method (Maltron 906BF[®], England). All procedures established were followed. It should be noted that the participants were fasting 2 hours before the evaluation, did not drink alcohol or perform vigorous exercises in the 24 hours prior to the exam, and urinated at least 30 minutes before the test. Measurements were taken with the individual in supine position and without any metallic objects on their body [29–31]. For the classification of the adolescents' nutritional status through the body fat percentage, the following cutoff points were used [32,33]: For males, 10.1–20% was considered normal body fat percentage (BFP) and ≥20.1% as high BFP; and for females, 15.1–25% was considered as normal BFP and ≥25.1 as high BFP.

For BFP measurement through DXA, the GE Healthcare Lunar Prodigy device was used, and the scans were analyzed using software version 14.10 (GE Healthcare). This analysis was performed in a subgroup of 12 adolescents also from the public school system of São Luís, MA, in order to show the correlation between the value of BFP obtained by bioimpedance and by DXA.

2.4. Predictor Variables

All the characteristics used as input from the model (Table 1) were chosen based on low cost indicators and easy application. They also had to be described in the literature for the assessment of nutritional status and health of adolescents. The variables—age, body mass, height, and gender—are indexes recommended by the World Health Organization and have been measured in several studies for the nutritional analysis of this population [1,27,28,34–36].

Table 1. Entry attributes in our model.

Attribute	Abbreviation	Unit
Body Mass	BM	Kg
Height	Ht	M
Gender	-	-
Age	-	years
Arm Circumference	AC	cm
Waist Circumference	WC	cm
Calf Circumference	CC	cm
Hip Circumference	HC	cm

Abbreviations: kg—kilogram; m—meters; cm—centimeter.

Those variables are also used as a criterion for the evaluation and classification of BMI and waist circumference (for example, References [17,22,37]). In contrast to the anthropometric indicators, waist circumference is already a consolidated index for the analysis of central fat and cardiovascular risk [37–39]. In addition, the circumference of the arm, hip, and calf are used to assess the nutritional status and population health [22,23,25,40–46]. Other factors were considered, such as low cost, reproducibility, and accessibility to the entry attributes, especially if the method is used in remote places.

2.5. Statistical Model

In many areas of knowledge, such as engineering and health, many problems involve investigating the relationship between two or more variables [47–50]. Multiple linear regression (MLR) is a statistical technique widely used in the literature to verify the relationship between a dependent variable and several independent variables [51]. Therefore, to build the computational model to predict body fat percentage, the concept of multiple linear regression was applied, and MATLAB[®] was used to build the model.

The MLR is based on least squares [51], which minimizes the error between the actual results (BFP obtained by the BIA) of the model and the expected results of the training set. The multiple linear regression aims to find an estimate of the real output by means of an equation, according to Equation (1):

$$y = x_1\beta_1 + \dots + x_n\beta_n + \varepsilon \quad (1)$$

where y represents the dependent variable, x_i the independent variables, β_i indicates the regression coefficients, and ε is the error term.

Our model has eight predictive (independent) variables described in Table 1, and the model output (dependent variable) will be the estimated body fat percentage value. After obtaining the BFP value through the model, the participant's nutritional status was assessed using the Lohman study cutoff point [32,33].

For generalization, and in order to avoid overfitting in the proposed method, the K-fold cross validation test was used, which consists of dividing the data into training and testing sets, where the data is equally divided into equal or nearly equal k segments. In these partitioned folds, both the training and the test are performed through k iterations. In each iteration, we leave a fold to test and train the model in the remaining $k-1$ folds [52]. Based on Lopes et al. [53], Afzal et al. [54], Song et al. [50], and Chang et al. [55], our dataset was randomly divided into five subsets ($k = 5$). Thus, from the 772 adolescents, 618 were randomly chosen to compose the training set and 154 were used as the test set (i.e., a proportion of 80% of the data for training and 20% for testing).

2.6. Performance Analysis

The performance of the method was evaluated in regard to the sensitivity (Sens—percentage of the cases that are correctly identified as true), specificity (Spe—percentage of the cases that are correctly identified as false), and accuracy (Accu—percentage of the cases that are correctly identified among all subjects). To obtain these measurements, the values of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) were calculated [56]. The area under receiver operating characteristic (AUROC) and confidence intervals were determined. The cutoff points used for BMI, WHtR, and BFP performance analysis are described in Section 2.3.

2.7. Statistical Analysis

For the statistical analysis, the SPSS software (Statistical Package for the Social Sciences, Inc., Chicago, IL, USA) version 25.0 was used. Data was treated with descriptive procedures (median and interquartile range). To compare groups with normal and altered BFP, the Mann–Whitney U test was used. The Chi-squared test was performed to verify the frequency of categorical variables. Pearson's correlation was used to evaluate the degree of correlation between the estimated value by the model and the real one (obtained by BIA), as well as for the analysis of the correlation between BIA value obtained by the BIA and DXA value. The results were considered statistically significant if the p -value was < 0.05 .

3. Results

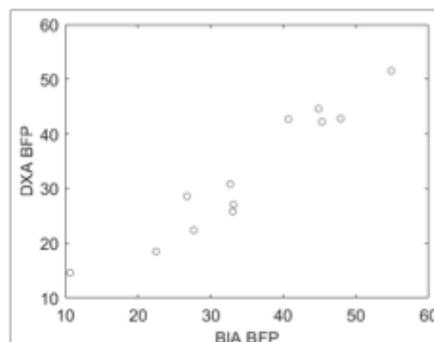
Table 2 presents the general characteristics of the sample composed of 772 adolescents aged 10 to 19 years.

Table 2. Socio-demographic, anthropometric, and hemodynamic characteristics from the database, stratified through body fat percentage (BFP).

Variables	Normal BFP * (n = 233)	Elevated BFP * (n = 539)	All # (n = 772)
Ht (m)	1.67 (1.59–1.75)	1.62 (1.57–1.68)	1.64±0.09
BM (kg)	53.6 (45.7–61.5)	56.9(50.1–64.3)	57.18±11.67
Age (years)	17 (15–17)	16(15–17)	15.66±1.72
HC (cm)	86.5 (82–91)	92 (86.5–97)	90.31±8.6
WC (cm)	66 (62–71)	69 (65–75)	69±8.4
AC (cm)	23 (21–25.5)	25 (23–28)	24.9±3.90
CC (cm)	32 (30–34.7)	34 (31.5–36)	33.2±3.9
BFP (%)	15.7(12.7–19.8)	33.7 (28.4–39.1)	28.4±10.23
Gender §			
Female	99	402	501
Male	134	137	271
Ethnicity §			
Caucasian	53	124	177
Non-Caucasian	180	415	595

Abbreviations: Ht—height; BM—body mass; HC—hip circumference; WC—waist circumference; AC—arm circumference; CC—calf circumference; kg—kilogram; m—meters; cm—centimeter; %—percentage; BFP—body fat percentage; * Values are presented as median (interquartile range: 25–75%); # Results are presented using mean ± standard deviation; § Values shown in frequency.

The BFP obtained by BIA showed a high correlation ($R = 0.96$) with DXA BFP in the validation subgroup ($n = 12$), with a confidence interval of 0.85–0.98. Figure 1 presents the relationship between the DXA value and the BIA BFP value, demonstrating the validity of BIA for BFP estimation in this population.

**Figure 1.** Relation between the DXA BFP and BIA BFP. BFP—body fat percentage; DXA—dual energy X-ray absorptiometry; BIA—Bioimpedance.

When comparing the model in relation to the body fat percentage (BIA), there was a significant association between the BFP and the chosen input attributes (Table 3).

Table 3. Regression coefficient (Beta) and significance (*p*) for each analyzed variable of the model, in relation to the body fat percentage.

Attribute	Beta	<i>p</i>
Body Mass	−0.12	0.018
Height	−6.84	0.005
Gender	−11.5	<0.001
Age	−1.32	<0.001
Arm Circumference	0.37	0.002
Waist Circumference	0.31	<0.001
Calf Circumference	0.40	0.001
Hip Circumference	0.30	<0.001

The proposed method presented high correlation with the BFP value obtained through bioimpedance ($R = 0.80$), with a confidence interval of 0.73–0.85. Figure 2 presents the relation between the real value (BIA) and the value estimated by our method.

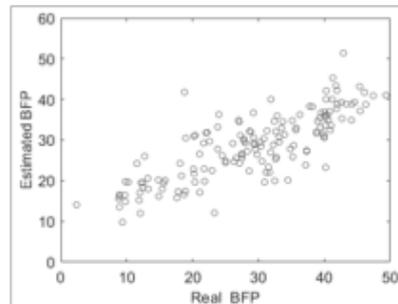


Figure 2. Relation between the real value (BIA) and value estimated by the proposed method. Abbreviations: BFP—body fat percentage.

In Table 4, it is observed that BMI and WHtR present a low AUROC discriminatory power when compared to the proposed method. The WHtR showed low performance compared to the other methods evaluated, presenting a confidence interval less than 0.5 and low sensitivity. Similarly to BMI, the WHtR failed to diagnose more than 50% of the sample of adolescents with excess body fat.

Table 4. Analysis of the performance in the test set of the proposed methods relative to the anthropometric indicators BMI and WHtR.

Indicators	AUROC * (CI 95 %)	Accu (%)	Sens (%)	Spe (%)	TP (%)	TN (%)	FP (%)	FN (%)
MP	0.80 (0.70–0.90)	85.1	92	67.4	66	19	9	6
BMI	0.64 (0.51–0.77)	47.4	27	100	19	28	0	53
WHtR	0.55 (0.36–0.74)	35.1	9.9	100	7	28	0	65

Abbreviations: CI 95 %—confidence interval; * Area under the ROC curve demonstrating discriminatory power for body fat (lower limit of CI 95 % > 0.50); MP—proposed method; WHtR—waist-to-height ratio; BMI—body mass index; Accu—accuracy; Sens—sensitivity; Spe—specificity; TP—true positives; FP—false positives; TN—true negatives; FN—false negatives; %—percentage.

To verify the performance of our method, during the test phase and relative to the obesity indicators commonly used in clinical practices and epidemiological studies, the performance indicators AUROC,

precision, sensitivity, and specificity were analyzed. Our model presented excellent discriminatory power, with respect to AUROC, as shown in Table 4 and represented in Figure 3.

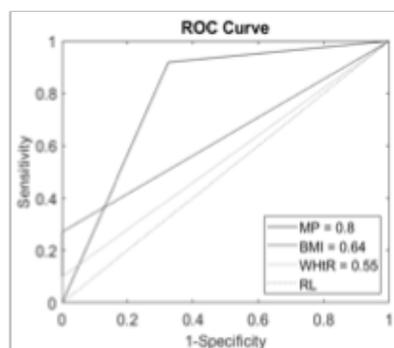


Figure 3. Area under Receiver Operating Characteristic demonstrating the abilities of discrimination of the model and of the anthropometric indicators in the prediction of high body fat in adolescents in the dataset. Abbreviations: MP—proposed method; WHtR—waist-to-height ratio; BMI—body mass index; RL—reference line.

Besides, our model showed a better performance than the BMI and WHtR indicators, having high sensitivity with respect to these indicators (Table 4). The development and implementation of a sensible model is of great importance, as a screening method must present high sensitivity, especially if it is used in the analysis of high body fat in adolescents.

Figure 3 is a graphical representation of AUROC, which is generated by plotting a routine (true positive rate) without axis in relation to a specificity (false positive rate) on the x-axis. Thus, for a diagnostic test to be ingested, it is necessary to have a curve without the upper left triangle above the reference line. When higher, it will be better than the next model [57].

Adolescence can be divided into three stages [58]: early adolescence (10–14 years of age), late adolescence (15–19 years of age), and young adults (20–24 years of age). Our test set was divided into two groups, taking into account the gender: 10–14 years of age (precocious adolescence) and 15–19 years of age (late adolescence). For all groups, the measures of accuracy, sensitivity, and specificity were calculated in order to evaluate our method against the anthropometric indicators. The values of accuracy, sensitivity, and specificity are set forth in Table 5. Our method continued to present better sensitivity than the BMI and WHtR anthropometric indices (Table 5).

Table 5. Analysis of the performance of the proposed methods in the data set relative to the anthropometric indicators BMI and WHtR, stratified by gender and age.

(%)	Male						Female					
	10–14 (n = 13)			15–19 (n = 44)			10–14 (n = 17)			15–19 (n = 80)		
	MP	BMI	WHtR									
Accu	84	46	46	82	73	64	76	53	23	89	32	20
Sens	75	12	12	80	40	20	100	38	0	96	23	8
Spe	100	100	100	83	100	100	0	100	100	40	100	100
TP	46	8	8	36	18	9	76	29	0	84	20	7
TN	39	39	39	46	55	55	0	24	24	5	13	13
FP	0	0	0	9	0	0	24	0	0	7	0	0
FN	15	53	53	9	27	36	0	47	76	4	67	80

Abbreviations: MP—proposed method; WHtR—waist-to-height ratio; BMI—body mass index; Accu—accuracy; Sens—sensitivity; Spe—specificity; TP—true positives; FP—false positives; TN—true negatives; FN—false negatives; %—percentage.

4. Discussion

Adolescence is a stage of physical and intellectual changes [2,59] where there is intense body growth that interferes with the accumulation and distribution of body fat [44]. It is therefore seen as one of the most critical periods for the development of obesity [17]. In the assessment of nutritional status, one of the most relevant indicators is the percentage of body fat (PBF), which is an independent risk factor for insulin resistance and a strong predictor of morbidity, as well as being a key parameter for the preventive and therapeutic intervention of pediatric obesity [60].

Measurements of body composition derived from BIA are valuable for the analysis of the nutritional status of pediatric patients [11,61]. It is one of the most reliable and affordable methods for assessing body fat and has a high correlation ($r = 0.96$ – 0.92) with dual energy X-ray absorptiometry [9], corroborating our results ($r = 0.96$). Despite the advantages, BIA still has a high cost when compared to the use of anthropometric indicators and clinical variables.

Thus, we built a method to assess nutritional status in adolescents by estimating body fat percentage using low cost and easy application parameters. The method used for estimation was the MLR, which was already successfully applied to solve several problems such as clinical data analysis [47] to verify the association between autonomic cardiac function and clinical variables [49]; to investigate the effects of food contamination on gastrointestinal tract morbidity [50]; and soil density measurement [48].

In the assessment of excess body fat, the proposed method obtained better performance (accuracy, sensitivity, and AUROC) than the anthropometric indicators BMI and WHtR, which are usually used to assess the nutritional status of adolescents. It is an attractive method with low cost and easy application when compared to other methods of body composition analysis such as bioimpedance. The best performance of our method can be explained by the use of predictor variables, such as body circumferences (Table 1), which are widespread measures in the literature for the assessment of obesity in this population and are associated with the presence of visceral fat and cardiovascular risk factors [38,62–64], as well as the use of indicators such as gender, age, height, and body mass together, which are widely used in the analysis of the nutritional profile and health status of the juvenile population [1,17,34,65].

The body mass index, despite being the most used method in clinical practice and epidemiological studies to assess excess body fat [21,66], is widely criticized for not correlating with body composition and being a poor predictor of body fat [9,12,15,60,67]. In a recent review, it was observed that BMI has high specificity (approximately 92%) and low sensitivity (approximately 50%) to detect obesity based on body fat percentage [68,69]. Therefore, more than half of individuals could be mistakenly classified with normal BFP by BMI [68], corroborating the results of the present study.

This fact is concerning because the low sensitivity of BMI indicates that excess adiposity is being underdiagnosed in several individuals. Moreover, because the first step in dealing with a risk factor is the precise identification of the pathophysiological problem, late diagnosis of excess adiposity will delay the treatment of associated comorbidities, as well as the implementation of intervention and control measures [68,70,71], especially in the juvenile population [71].

Although used in several studies with adolescents, WHtR did not show efficient discriminatory power in the evaluation of adolescents with high body fat in the present study. In a systematic review, Lo et al. [72] observed that WHtR did not perform better in predicting cardiometabolic risk factors than anthropometric indicators BMI and WC (waist circumference) in children and adolescents.

Thus, the use of high sensitivity methods is a major challenge for the early diagnosis of obese adolescents. The model proposed here has predictive and practical advantages in situations with limited resources, such as areas without access to equipment for body composition analysis (bioimpedance, for example). Thus, this is an alternative for screening adolescents for excess body fat. In addition, such a method may guide health professionals in decision-making and potentially expedite tests such as lipid profile and insulin resistance, which are associated with high body fat levels and cardiovascular risk.

A limitation of this study was to use as a test set approximately 20 % of the sample, since a larger sample could provide more robust results. However, in this study, the main goal was to present a new method for screening obesity in adolescents. The authors believe that an external validation study should be performed in other regions of Brazil and other countries due to variations in ethnic descent.

5. Conclusions

The computational model of this study obtained a better performance in the evaluation of excess body fat in adolescents compared to the usual anthropometric indicators, thus presenting itself as a low cost alternative for screening obesity in adolescents living in Brazilian regions where financial resources are scarce.

Author Contributions: Conceptualization, N.P. and C.M.S.; methodology, E.S.; validation, E.S., A.K.B., and M.V.L.; formal analysis, E.S.; investigation, N.P.; data curation, G.L.; writing—original draft preparation, C.M.S.; writing—review and editing, L.A. and É.C.

Funding: This research received no external funding.

Acknowledgments: FAPEMA and Departamento de Saúde Pública da Universidade Federal do Maranhão, for assigning DXA to the body fat percentage assessment of the validation group.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Katia, V.C.; Moyses, S.; Maria, C.C.K.; Gabriela, A.A.; Laura, A.B.; Carlos, H.K.; Maurício, T.L.V.; Glória, V.V.; Valeska, C.F.; Adriano, D.; et al. The study of cardiovascular risk in adolescents—ERICA: Rationale, design and sample characteristics of a national survey examining cardiovascular risk factor profile in Brazilian adolescents. *BMC Public Health* **2015**, *15*, 94.
2. Dahl, R.E.; Allen, N.B.; Wilbrecht, L.; Suleiman, A.B. Importance of investing in adolescence from a developmental science perspective. *Nature* **2018**, *554*, 441–450. [CrossRef] [PubMed]
3. Hagani, N.; Moran, M.R.; Caspi, O.; Plaut, P.; Endevelt, R. The Relationships between Adolescents' Obesity and the Built Environment: Are They City Dependent? *Int. J. Environ. Res. Public Health* **2019**, *16*, 1579. [CrossRef]
4. Reinehr, T. Long-term effects of adolescent obesity: Time to act. *Nat. Rev. Endocrinol.* **2018**, *14*, 183. [CrossRef] [PubMed]
5. Andreoli, A.; Garaci, F.; Cafarelli, F.P.; Guglielmi, G. Body composition in clinical practice. *Eur. J. Radiol.* **2016**, *85*, 1461–1468. [CrossRef] [PubMed]
6. Tewari, N.; Awad, S.; Macdonald, I.A.; Lobo, D.N. A comparison of three methods to assess body composition. *Nutrition* **2018**, *47*, 1–5. [CrossRef] [PubMed]
7. Böhm, A.; Heitmann, B.L. The use of bioelectrical impedance analysis for body composition in epidemiological studies. *Eur. J. Clin. Nutr.* **2013**, *67*, S79. [CrossRef]
8. Dehghan, M.; Merchant, A.T. Is bioelectrical impedance accurate for use in large epidemiological studies? *Nutr. J.* **2018**, *7*, 26. [CrossRef]
9. Ramírez-Vélez, R.; Correa-Bautista, J.; Martínez-Torres, J.; González-Ruiz, K.; González-Jiménez, E.; Valle, J.; Ramírez-Vélez, R. Performance of Two Bioelectrical Impedance Analyses in the Diagnosis of Overweight and Obesity in Children and Adolescents: The FUPRECOL Study. *Nutrients* **2016**, *8*, 575. [CrossRef]
10. Tovar-Galvez, M.I.; González-Jiménez, E.; Martí-García, C.; Schmidt-RioValle, J. Composición corporal en escolares: Comparación entre métodos antropométricos simples e impedancia bioeléctrica. *Endocrinol. Diabetes Nutr.* **2017**, *64*, 424–431. [CrossRef]
11. Von Hurst, P.R.; Walsh, D.C.I.; Conlon, C.A.; Ingram, M.; Kruger, R.; Stonehouse, W. Validity and reliability of bioelectrical impedance analysis to estimate body fat percentage against air displacement plethysmography and dual-energy X-ray absorptiometry. *Nutr. Diet.* **2016**, *73*, 197–204. [CrossRef]
12. Guo, B.; Wu, Q.; Gong, J.; Xiao, Z.; Tang, Y.; Shang, J. Relationships between the lean mass index and bone mass and reference values of muscular status in healthy Chinese children and adolescents. *J. Bone Miner. Metab.* **2015**. [CrossRef] [PubMed]

13. Laurson, K.R.; Welk, G.J.; Eisenmann, J.C. Diagnostic Performance of BMI Percentiles to Identify Adolescents With Metabolic Syndrome. *Pediatrics* **2014**, *133*, 330–338. [CrossRef] [PubMed]
14. de Oliveira, P.M.; Da Silva, F.A.; Oliveira, R.M.S. Association between fat mass index and fat-free mass index values and cardiovascular risk in adolescents. *Revista Paulista de Pediatria (English Edition)* **2016**, *34*, 30–37. [CrossRef]
15. de Cássia Ribeiro-Silva, R.; Florence, T.; da Conceição-Machado, M.E.P.; Fernandes, G.B.; Couto, R.D. Indicadores antropométricos na predição de síndrome metabólica em crianças e adolescentes: um estudo de base populacional. *Rev. Bras. Saúde Matern. Infant.* **2014**, *14*. [CrossRef]
16. Weber, D.R.; Leonard, M.B.; Zemel, B.S. Body composition analysis in the pediatric population. *Pediatr. Endocrinol. Rev.* **2012**, *10*, 130–139.
17. ANS. Agência Nacional De Saúde Suplementar Diretoria De Normas E Habilitação Dos Produtos Gerência De Monitoramento Assistencial Coordenadoria De Informações Assistenciais. 2017. Available online: http://www.ans.gov.br/images/Manual_de_Diretrizes_para_o_Enfrentamento_da_Obesidade_na_Saude_Suplementar_Brasileira.pdf (accessed on 19 June 2019).
18. Lohman, T.G. Applicability of body composition techniques and constants for children and youths. *Exerc. Sport Sci. Rev.* **1986**, *14*, 325–358. [CrossRef]
19. Lwanga, S.; Lemeshow, S. *Sample Size Determination in Health Studies: A Practical Manual*; World Health Organization: Geneva, Switzerland, 1991. [CrossRef]
20. Oliveira MLDE. Estimativa Dos Custos Da Obesidade Para O Sistema Único de Saúde do Brasil. Univ Brasília. 2013. Available online: <http://repositorio.unb.br/handle/10482/13323> (accessed on 19 June 2019).
21. Farias, E.D.S. Excess weight and associated factors in adolescents. *Rev. Nutr.* **2012**, *25*, 229–236. [CrossRef]
22. Brasil A. Associação Brasileira para o Estudo da Obesidade e da Síndrome Metabólica. Obesidade Mata. 2016. Available online: <http://www.abeso.org.br/uploads/downloads/92/57fccc603e5da.pdf> (accessed on 29 June 2019).
23. Huang, B.X.; Zhu, M.F.; Wu, T.; Zhou, J.Y.; Liu, Y.; Chen, X.L.; Zhou, R.F.; Wang, L.J.; Chen, Y.M.; Zhu, H.L. Neck circumference, along with other anthropometric indices, has an independent and additional contribution in predicting fatty liver disease. *PLoS ONE* **2015**, *10*, e0118071. [CrossRef]
24. Rerksuppaphol, S.; Rerksuppaphol, L. Mid-upper-arm circumference and arm-to-height ratio to identify obesity in school-age children. *Clin. Med. Res.* **2017**, *15*, 53–58. [CrossRef]
25. Salmaso, F.V.; Vigário, P.D.S.; Mendonça, L.M.C.D.; Madeira, M.; Netto, L.V.; Guimarães, M.R.M.; Farias, M.L.F.D. Análise de idosos ambulatoriais quanto ao estado nutricional, sarcopenia, função renal e densidade óssea. *Anq. Bras. Endocrinol. Metab.* **2014**, *58*, 226–331. [CrossRef]
26. Brasil. Ministério da Saúde. *Orientações para a Coleta e Análise de Dados Antropométricos em Serviços de Saúde: Norma Técnica do Sistema de Vigilância Alimentar e Nutricional - SISVAN*; Secretaria de Atenção à Saúde, Departamento de Atenção Básica: Brasília, Brazil, 2011. Available online: http://bvsms.saude.gov.br/bvs/publicacoes/orientacoes_coleta_analise_dados_antropometricos.pdf (accessed on 29 June 2019).
27. Onis, M.D. WHO Child Growth Standards based on length/height, weight and age. *Acta Paediatr.* **2007**, *95*. [CrossRef]
28. Ashwell, M.; Hsieh, S.D. 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.* **2005**, *56*, 303–307. [CrossRef] [PubMed]
29. Houtkooper, L.B.; Going, S.B.; Lohman, T.G.; Roche, A.F.; Van Loan, M. Bioelectrical impedance estimation of fat-free body mass in children and youth: A cross-validation study. *J. Appl. Physiol.* **1992**, *72*, 366. [CrossRef] [PubMed]
30. Kyle, U.G.; Bosaeus, I.; De Lorenzo, A.D.; Deurenberg, P.; Elia, M.; Gómez, J.M. Bioelectrical impedance analysis—Part I: Review of principles and methods. *Clin. Nutr.* **2004**, *23*, 1226–1243. [CrossRef]
31. Neves, E.B.; Ripka, W.L.; Ulbricht, L.; Stadnik, A.M.W. Comparison of the fat percentage obtained by bioimpedance, ultrasound and skinfolds in young adults. *Rev. Bras. Med. Esporte* **2013**, *19*, 323–327. [CrossRef]
32. Lohman, T.G. Advances in body composition assessment. *Med. Sci. Sport Exerc.* **1993**, *25*, 762. [CrossRef]
33. Ripka, W.L. Modelos Matemáticos Para Estimativa da Gordura Corporal de Adolescentes Utilizando Dobras Cutâneas, a Partir da Absorciometria de Raios-X de Dupla Energia. Ph.D. Thesis, Universidade Tecnológica Federal do Paraná, Curitiba, Brazil, April 2017.

34. Jésus, P.; Guerchet, M.; Pilleron, S.; Fayemendy, P.; Maxime Mouanga, A.; Mbelesso, P. Undernutrition and obesity among elderly people living in two cities of developing countries: Prevalence and associated factors in the EDAC study. *Clin. Nutr. ESPEN* **2017**, *21*, 40–50. [CrossRef]
35. Sigulem, D.M.; Devincenzi, M.U.; Lessa, A.C. Diagnosis of the nutritional status of children and adolescents. *J. Pediatr.* **2000**, *76*, S275–S284. [CrossRef]
36. Onis, M.D.; Onyango, A.W.; Borghi, E.; Siyam, A.; Nishida, C.; Siekmann, J. Development of a WHO growth reference for school-aged children and adolescents. *Bull. World Health Organ.* **2007**, *85*, 660–667. [CrossRef]
37. Taylor, R.W.; Jones, I.E.; Williams, S.M.; Goulding, A. Evaluation of waist circumference, waist-to-hip ratio, and the conicity index as screening tools for high trunk fat mass, as measured by dual-energy X-ray absorptiometry, in children aged 3–19 y13. *Am. J. Clin. Nutr.* **2000**, *72*, 490–495. [CrossRef] [PubMed]
38. Savva, S.C.; Tornaritis, M.; Savva, M.E.; Kourides, Y.; Panagi, A.; Silikiotou, N. Waist circumference and waist-to-height ratio are better predictors of cardiovascular disease risk factors in children than body mass index. *Int. J. Obes.* **2000**, *24*, 1453–1458. [CrossRef]
39. Schröder, H.; Ribas, L.; Koenig, C.; Fantikova, A.; Gomez, S.F.; Fito, M. Prevalence of abdominal obesity in Spanish children and adolescents. do we need waist circumference measurements in pediatric practice? *PLoS ONE* **2014**, *9*, e87549. [CrossRef]
40. Alencar, M.D.S.S.; Melo, M.T.S.M.; Coê, R.; Meneses, A.V.; Sá, L.D.; Nunes, I.F.D.O.C. Loss of muscle and fat mass after institutionalization: Attention to older. *Geriatr. Gerontol. Aging* **2015**, *9*, 150–155. [CrossRef]
41. de Faria, E.R.; Gortijo, C.A.; Franceschini, S.D.C.C.; Peluzio, M.D.C.G.; Priore, S.E. Body composition and risk for metabolic alterations in female adolescents. *Rev. Paul. Pediatr.* **2014**, *32*, 207–215. [CrossRef] [PubMed]
42. Kroll, C.; Mastroeni, S.S.B.S.; Czarnobay, S.A.; Ekwara, J.P.; Veugelers, P.J.; Mastroeni, M.F. The accuracy of neck circumference for assessing overweight and obesity: a systematic review and meta-analysis. *Ann. Hum. Biol.* **2017**, *44*, 667–677. [CrossRef]
43. Li, C.; Ford, E.S.; Mokdad, A.H.; Cook, S. Recent Trends in Waist Circumference and Waist-Height Ratio Among US Children and Adolescents. *Pediatrics* **2006**, *118*, 1390–1398. [CrossRef]
44. Pereira, P.F.; de Faria, E.R.; de Faria, E.R.; Hermsdorff, H.H.M.; Peluzio, M.D.C.G.; Franceschini, S.D.C.C.; Priore, S.E. Anthropometric indices to identify metabolic syndrome and hypertriglyceridemic waist phenotype: a comparison between the three stages of adolescence. *Rev. Paul. Pediatr.* **2015**, *33*, 194–203. [CrossRef]
45. Ripka, W.L.; Ulbricht, L.; Gewehr, P.M. Body composition and prediction equations using skinfold thickness for body fat percentage in Southern Brazilian adolescents. *PLoS ONE* **2017**, *12*, e0184854. [CrossRef]
46. Talma, H.; Van Dommelen, P.; Schweizer, J.J.; Bakker, B.; Kist-Van Holthe, J.E.; Chinapaw, J.M.M. Is mid-upper arm circumference in Dutch children useful in identifying obesity? *Arch. Dis. Child.* **2019**, *104*, 159–165. [CrossRef]
47. Coelho-Barros, E.A.; Simões, P.A.; Achcar, J.A.; Martinez, E.Z.; Shimano, A.C. Methods of Estimation in Multiple Linear Regression: Application to Clinical Data. *Rev. Colomb. Estad.* **2008**, *31*, 111–129.
48. Carvalho Junior, W.D.; Filho, B.C.; Chagas, C.D.S.; Bhering, S.B.; Pereira, N.R.; Pinheiro, H.S.K. Multiple linear regression and Random Forest model to estimate soil bulk density in mountainous regions. *Pesqui. Agropecu. Bras.* **2016**, *51*, 1428–1437. [CrossRef]
49. Santos, R.R.; Rosa, E.C.; Rosa, T.; Ferreira, E.A.; Gris, E.F.; de Andrade, R.V. Sedentary Behavior: A Key Component in the Interaction between an Integrated Lifestyle Approach and Cardiac Autonomic Function in Active Young Men. *Int. J. Environ. Res. Public Health* **2019**, *16*, 2156. [CrossRef]
50. Song, Q.; Zheng, Y.J.; Yang, J. Effects of Food Contamination on Gastrointestinal Morbidity: Comparison of Different Machine-Learning Methods. *Int. J. Environ. Res. Public Health* **2019**, *16*, 838. [CrossRef]
51. Akan, R.; Keskin, S.N.; Uzundurukan, S. Multiple Regression Model for the Prediction of Unconfined Compressive Strength of Jet Grout Columns. *Procedia Earth Planet. Sci.* **2015**, *75*, 299–303. [CrossRef]
52. Yadav, S.; Shukla, S. Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. In Proceedings of the 6th International Advanced Computing Conference, IACC 2016, Taipei, Taiwan, 12–16 September 2016.
53. Lopes, M.V.; Barradas Filho, A.O.; Barros, A.K.; Viegas, I.M.A.; Silva, L.C.O.; Marques, E.P. Attesting compliance of biodiesel quality using composition data and classification methods. *Neural Comput. Appl.* **2019**, *31*, 539–551. [CrossRef]

54. Afzal, Z.; Schuemie, M.J.; Van Blijderveen, J.C.; Sen, E.F.; Sturkenboom, M.C.; Kors, J.A. Improving sensitivity of machine learning methods for automated case identification from free-text electronic medical records. *BMC Med. Inf. Decis. Mak.* **2013**, *13*, 30. [CrossRef]
55. Chang, F.; Chen, H.C.; Liu, H.C. Double K-Folds in SVM. In Proceedings of the 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, IMIS 2015, Santa Catarina, Brazil, 8–10 July 2015.
56. Rovini, E.; Marenmani, C.; Moschetti, A.; Esposito, D.; Cavallo, F. Comparative Motor Pre-clinical Assessment in Parkinson's Disease Using Supervised Machine Learning Approaches. *Ann. Biomed. Eng.* **2018**, *46*, 2057–2068. [CrossRef]
57. Hoo, Z.H.; Candlish, J.; Teare, D. What is an ROC curve? *Emerg. Med. J.* **2017**, *34*. [CrossRef]
58. Das, J.K.; Salam, R.A.; Thornburg, K.L.; Prentice, A.M.; Campisi, S.; Lassi, Z.S. Nutrition in adolescents: physiology, metabolism, and nutritional needs. *Ann. NY Acad. Sci.* **2017**, *1393*, 21–33. [CrossRef]
59. Moraes, M.M.D.; Veiga, G.V.D. Acurácia da gordura corporal e do perímetro da cintura para prever alterações metabólicas de risco cardiovascular em adolescentes. *Arq. Bras. Endocrinol. Metabol.* **2014**, *58*, 341–351. [CrossRef]
60. Vanderwall, C.; Randall Clark, R.; Eickhoff, J.; Carrel, A.L. BMI is a poor predictor of adiposity in young overweight and obese children. *BMC Pediatr.* **2017**, *17*, 135. [CrossRef]
61. Kyle, U.G.; Earthman, C.P.; Pichard, C.; Coss-Bu, J.A. Body composition during growth in children: limitations and perspectives of bioelectrical impedance analysis. *Eur. J. Clin. Nutr.* **2015**, *69*, 1298. [CrossRef]
62. Brambilla, P.; Bedogni, G.; Moreno, L.A.; Goran, M.I.; Gutin, B.; Fox, K.R. Crossvalidation of anthropometry against magnetic resonance imaging for the assessment of visceral and subcutaneous adipose tissue in children. *Int. J. Obes.* **2006**, *30*, 23–30. [CrossRef]
63. Rerksuppaphol, S.; Rerksuppaphol, L. Waist circumference, waist-to-height ratio and body mass index of Thai children: Secular changes and updated reference standards. *J. Clin. Diagn. Res.* **2014**, *8*, 5–9. [CrossRef]
64. Tebar, W.R.; Ritti-Dias, R.M.; Farah, B.Q.; Zanuto, E.F.; Vanderlei, L.C.M.; Christofaro, D.G.D. High blood pressure and its relationship to adiposity in a school-aged population: Body mass index vs waist circumference. *Hypertens. Res.* **2018**, *41*, 135. [CrossRef]
65. Sousa, N.P.S.; Salvador, E.P.; Barros, A.K.; Polisel, C.G.; de Carvalho, W.R.G. Anthropometric predictors of abdominal adiposity in adolescents. *J. Exerc. Physiol. Online* **2016**, *19*, 66–76.
66. Lavrador, M.S.F.; Abbas, P.T.; Escrivão, M.A.M.S.; Taddei, J.A.A.C. Riscos cardiovasculares em adolescentes com diferentes graus de obesidade. *Arquivos Brasileiros de Cardiologia* **2011**, *96*, 205–211. [CrossRef]
67. De Lorenzo, A.; Soldati, L.; Sarlo, F.; Calvani, M.; Di Lorenzo, N.; Di Renzo, L. New obesity classification criteria as a tool for bariatric surgery indication. *World J. Gastroenterol.* **2016**, *22*, 681. [CrossRef]
68. Oliveros, E.; Somers, V.K.; Sochor, O.; Goel, K.; Lopez-Jimenez, F. The Concept of Normal Weight Obesity. *Prog. Cardiovasc Dis.* **2014**, *56*, 426–433. [CrossRef]
69. Frayon, S.; Cavaloc, Y.; Wattelez, G.; Cherrier, S.; Lerrant, Y.; Ashwell, M. Potential for waist-to-height ratio to detect overfat adolescents from a Pacific Island, even those within the normal BMI range. *Obes. Res. Clin. Pract.* **2018**, *12*, 351–357. [CrossRef]
70. Javed, A.; Jumean, M.; Murad, M.H.; Okorodudu, D.; Kumar, S.; Somers, V.K. Diagnostic performance of body mass index to identify obesity as defined by body adiposity in children and adolescents: A systematic review and meta-analysis. *Pediatr. Obes.* **2015**, *10*, 234–244. [CrossRef]
71. Zhang, S.; Tjørtjns, C.; Zeng, X.; Qiao, H.; Buchan, I.; Keane, J. Comparing data mining methods with logistic regression in childhood obesity prediction. *Inf. Syst. Front.* **2009**, *11*, 449–460. [CrossRef]
72. Lo, K.; Wong, M.; Khalechelvam, P.; Tam, W. Waist-to-height ratio, body mass index and waist circumference for screening paediatric cardio-metabolic risk factors: A meta-analysis. *Obes. Rev.* **2016**, *17*, 1342–1343. [CrossRef]



ANEXO H - ARTIGO SUBMETIDO

Comprovante de submissão Journal of Biotechnology, fator de impacto: 3,307

21/12/2021 16:45 Gmail - Submission Confirmation

 Luana Azoubel <luazoubel@gmail.com>

Submission Confirmation
1 mensagem

em@editorialmanager.com <em@editorialmanager.com> 21 de dezembro de 2021 16:44
Responder a: support@elsevier.com
Para: Luana Azoubel <luazoubel@gmail.com>

Re: PREDICTION OF SERUM CREATININE IN HEALTH AND CHRONIC KIDNEY PATIENTS BY MULTIPLE LINEAR REGRESSION USING NON-INVASIVE VARIABLES
by Luana Azoubel; Érika Cristina Ribeiro de Lima Carneiro; Carlos Sousa Magno Junior; Nilviane Pires Silva Sousa; Ronaldo dos Santos Silva Junior; Guilherme Oliveira Lima; Giselle Andrade Santos Silva; Naruna Aritana Costa Melo; Ewaldo Eder Carvalho Santana; Allan Kardec Barros
Research Paper

Dear Ms Luana Azoubel,

Your submission entitled "PREDICTION OF SERUM CREATININE IN HEALTH AND CHRONIC KIDNEY PATIENTS BY MULTIPLE LINEAR REGRESSION USING NON-INVASIVE VARIABLES" has been received by Journal of Biotechnology

You may check on the progress of your paper by logging on to the Editorial Manager as an author. The URL is <https://www.editorialmanager.com/jbiotec/>.

Your username is: luana.azoubel

If you need to retrieve password details, please go to: <https://www.editorialmanager.com/jbiotec/l.asp?i=278065&l=X1KDOLXW>

Your manuscript will be given a reference number once an Editor has been assigned.

Thank you for submitting your work to this journal.

For further assistance, please visit our customer support site at <http://help.elsevier.com/app/answers/list/p/7923>. Here you can search for solutions on a range of topics, find answers to frequently asked questions and learn more about EM via interactive tutorials. You will also find our 24/7 support contact details should you need any further assistance from one of our customer support representatives.

Kind regards,

Editorial Manager
Journal of Biotechnology
#AU_JBIOTEC#

To ensure this email reaches the intended recipient, please do not delete the above code

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Use the following URL: <https://www.editorialmanager.com/jbiotec/login.asp?a=r>). Please contact the publication office if you have any questions.

<https://mail.google.com/mail/u/0/?ik=3478bd9e71&view=pt&search=all&permthid=thread-F%3A1719786143935973822&siml=msg-f%3A171978614393597...> 1/1

ANEXO I - ARTIGO SUBMETIDO

Comprovante de submissão Nutrição Hospitalar, fator de impacto: 1,057

17/12/2021 10:21 Gmail - Nutr hosp Article created

 Luana Azoubel <luazoubel@gmail.com>

Nutr hosp Article created
2 mensagens

Revista Nutricion Hospitalaria <noreply@nutricionhospitalaria.org> 3 de novembro de 2021 16:03
Para: LUANA <luazoubel@gmail.com>

Dear Dr AZOUBEL:

We confirm that your article "PREDICTIVE POWER OF ANTHROPOMETRIC MEASURES AND LOW GLOMERULAR FILTRATION RATE IN KIDNEY TRANSPLANT RECIPIENTS" (03937-01) has been loaded successfully to the Nutrición Hospitalaria platform.

From this moment on, we will begin the evaluation on it. You can check its status at https://www.nutricionhospitalaria.org/admin/m_articulos/5563/show

Best regards,

Editorial Department
Nutrición Hospitalaria

Revista Nutricion Hospitalaria <noreply@nutricionhospitalaria.org> 3 de novembro de 2021 16:05
Para: LUANA <luazoubel@gmail.com>

[Texto das mensagens anteriores oculto]

<https://mail.google.com/mail/u/0/?ik=3478bd9c71&view=pt&search=all&permthid=thread-f%3A1715434886440179083&simpl=msg-f%3A171543488644017...> 1/1

APÊNDICE

APÊNDICE A - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)



UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
LABORATÓRIO DE PROCESSAMENTO DA INFORMAÇÃO BIOLÓGICA



TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Nome do Estudo: Predição da Doença Renal Crônica Utilizando Redes Neurais Artificiais.

Prezado (a) sr (a).

Você está sendo convidado (a) para participar de um estudo que está sendo desenvolvido por pesquisadores da Universidade Federal do Maranhão – UFMA, sob coordenação do professor Allan Kardec Barros, o qual tem por objetivo desenvolver um software capaz de prever os estágios da doença renal crônica tendo como base os fatores de risco envolvidos na progressão da doença.

Para isso serão feitas as medidas de peso, altura, circunferência do quadril, braço, panturrilha, cintura e pescoço, pregas cutâneas, exame de bioimpedância elétrica, aplicação de questionário e coleta de sangue para determinação de exames laboratoriais marcadores da função renal (creatinina, ureia e albumina). *Todos os resultados serão informados aos participantes (se assim o desejarem).* O perfil hemodinâmico (aferição da pressão arterial, HOLTER 24 horas e MAPA 24 horas), perfil bioquímico (hemograma, ferro sérico, ferritina, capacidade de ligação de ferro à transferrina, creatinina sérica, ureia sérica, cistatina C, ácido úrico, taxa de filtração glomerular, depuração de creatinina, filtração glomerular estimada, potássio, cálcio, vitamina D, paratormônio, fosfatase alcalina, osteocalcina, proteína C reativa, sumário de urina, albumina, colesterol total e suas frações, glicemia de jejum, teste de tolerância à glicose, hemoglobina glicada, insulina, *homeostatic model assessment- HOMA*) e eletrocardiográfico (ECG) serão obtidos através dos dados contidos nos prontuários dos participantes atendidos no setor de hemodiálise e diálise peritoneal do Hospital Universitário da Universidade Federal do Maranhão.

A coleta de sangue para dosagem de creatinina sérica, ureia e albumina será realizada por pesquisadores treinados, com material estéril e descartável. Após a coleta de sangue pede-se que o participante pressionar o local da coleta por no mínimo 3 minutos, evite flexionar (dobrar) o braço, não massageie o local da coleta e não carregue peso no braço que foi feita a punção sanguínea (como carregar a bolsa no braço em que

Rubricas: _____ : _____
Participante Pesquisador



UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
LABORATÓRIO DE PROCESSAMENTO DA INFORMAÇÃO BIOLÓGICA



foi coletado sangue). A punção sanguínea pode causar um leve desconforto (como uma picada de inseto, p. ex.: formiga) e muito raramente o aparecimento de hematomas (local ficar roxo) que desaparecem em até 48 horas. Caso ocorra hematoma orienta-se ao paciente: colocar compressas de gelo por 15 min. a cada hora nas primeiras seis horas. Se necessário pode-se prolongar este tempo por mais 2 ou 3 horas. Em seguida podem ser colocadas compressas mornas para ajudar a eliminar o hematoma mais rapidamente. A formação de hematoma (braço roxo) não causa alterações no estado de saúde do participante (não leva a qualquer tipo de doença).

Em caso de dúvidas, você poderá entrar em contato com a equipe de pesquisadores pelo telefone (98) 3272-8273, durante o horário comercial: 08-12h e 14-18 h. Ou em caso de emergência pelo telefone: (98) 99151-1526. Você poderá entrar em contato também com Comitê de Ética em Pesquisa – CEP do Hospital Universitário da Universidade Federal do Maranhão para tirar dúvidas a respeito de questões éticas. Os Comitês de Ética em Pesquisa são colegiados interdisciplinares e independentes, de relevância pública, de caráter consultivo, deliberativo e educativo, criados para garantir a proteção dos participantes da pesquisa em sua integridade e dignidade e para contribuir no desenvolvimento da pesquisa dentro de padrões éticos. O CEP do Hospital Universitário da Universidade Federal do Maranhão está localizado na Rua Barão de Itapary, 227, quarto andar do Hospital Universitário da Universidade Federal do Maranhão, Centro, São Luis-MA. CEP- 65.020-070. Telefone (98) 2109 1250.

Sua participação neste estudo é muito importante, pois irá contribuir para a detecção precoce da doença renal crônica ajudando assim a melhorar a qualidade de vida de diversos pacientes com esta patologia, bem como seus familiares.

Consentimento em participar: A sua participação neste estudo é totalmente voluntária, ou seja, é você que escolhe se quer ou não participar, podendo desistir a qualquer momento, sem precisar se justificar (explicar). Em caso de desistência não haverá qualquer punição ou alteração em seu tratamento (caso o participante seja paciente do setor de hemodiálise e diálise peritoneal do Hospital Universitário da Universidade Federal do Maranhão).

Antes que você decida, releia, pois é importante que você entenda exatamente qual será a sua participação no estudo e o motivo pelo qual ele será desenvolvido. Se você decidir participar, você deverá assinar nos locais adequados deste documento e

Rubricas: _____ ; _____
Participante Pesquisador



UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
LABORATÓRIO DE PROCESSAMENTO DA INFORMAÇÃO BIOLÓGICA



rubricar todas as páginas, isso irá representar o seu "Livre Consentimento em Participar deste Estudo". O pesquisador responsável ou membros da equipe de pesquisa também irão assinar o TCLE e rubricar todas as páginas. Este termo é elaborado em duas vias: uma via ficará em posse do pesquisador responsável e a outra via retida com o participante.

Consentimento em participar:

Eu, _____

li e entendi o Termo de Consentimento Livre e Esclarecido. Entendi que a minha participação neste estudo é totalmente voluntária (por vontade própria) e que posso, a qualquer momento, recusar a participar. Tive tempo suficiente para ler e pensar a respeito da nossa participação no presente estudo. Tive também a oportunidade de esclarecer nossas dúvidas com a Prof. Allan Kardex Barros ou sua equipe. Entendi, ainda, que todos os resultados deste estudo serão utilizados para fins científicos, *sendo garantindo o sigilo de todas as informações repassadas pelo participante e os dados não serão divulgados com qualquer informação que possa identificar o participante*. A minha assinatura abaixo significa que recebi uma via deste termo de consentimento.

CONSENTIMENTO PÓS-INFORMAÇÃO

Eu, _____, fui esclarecido (a) sobre a pesquisa Predição da Doença Renal Crônica Utilizando Redes Neurais Artificiais e concordo que os dados fornecidos sejam utilizados na realização da mesma.

São Luís, ___ de _____ 201__.

Assinatura do participante: _____

Assinatura do pesquisador responsável: _____

Rubricas: _____ ; _____
Participante Pesquisador

APENDICE B – QUESTIONÁRIO SEMIESTRUTURADO

	UNIVERSIDADE FEDERAL DO MARANHÃO <small>FUNÇÃO INSTITUCIONAL LEI Nº 13.103 DE 21/02/1966 São Luís - Maranhão</small>	
	<small>Centro de Ciências Exatas e Tecnologias Departamento de Engenharia Elétrica Laboratório de Processamento de Informação Biológica</small>	
 PROTOCOLO – ANAMNESE CLÍNICA 		
Data da Admissão: ___/___/___ Data da Intervenção: ___/___/___		
 DADOS PESSOAIS		
Nome: _____ Sexo: () Masc () Fem		
Cor: () Branca () Parda () Negra () Outras (indígena, amarela...) Idade: _____ Procedência: _____		
Data de nascimento: ___/___/___ Naturalidade: _____ Cidade que reside: _____		
Bairro: _____ Moradia: () Zona Rural () Zona Urbana Telefone: _____		
WhatsApp: _____ Gravidez: () Sim () Não Nº de pessoas na residência: _____ Renda		
(salário mínimo): () < 1 () 1 a 2 () 2 a 3 () > 3 Aposentado: () sim () não		
Profissão: _____ Grau de Instrução: _____		
Estado Civil: _____ Nº do prontuário: _____		
 QUEIXA PRINCIPAL (QP)		
Motivo da procura ao centro de atendimento (nas palavras do paciente) _____		

O que você sabe sobre a doença? _____		

 HISTÓRIA DA DOENÇA ATUAL (HDA)		
Todo histórico da QP		

 HISTÓRIA PATOLÓGICA PREGRESSA (HPP)		
É portador de diabetes? () Não () Sim () Não sabe		
Consegue controlar o diabetes com alimentação? () Não () Sim		
Outras patologias: _____		



UNIVERSIDADE FEDERAL DO MARANHÃO
FUNDAÇÃO INSTITUIA em 1958 de Lei n.º 1.100 de 20/03/58
 Mãe e Mãe
 Centro de Ciências Exatas e Tecnologias
 Departamento de Engenharia Elétrica
 Laboratório de Processamento da Informação Biológica



Procedimentos cirúrgicos: _____

Internações: _____

Reações alérgicas a medicações/ produtos/ alimento? () Não () Sim () Não sabe. Se sim, Qual?

Sínus e sintomas presentes: _____

Qual é sua frequência de evacuações? _____

Consistência das fezes: () Ressecadas (endurecidas) () Pastosas () Semi-pastosas () Líquidas () Líquidas com restos alimentares. Presença de raias de sangue: () Não () Sim Fezes gordurosas: () Não () Sim

Uso de laxantes? () Não () Sim. Qual a frequência do uso? _____. Em quais situações faz uso? _____

Frequência da diurese? _____

Qual a quantidade em 24h? _____

Qual a coloração? () Amarelo palha () Amarelo citrino () Amarelo âmbar () Outra. Ardência para urinar (disúria)? () Não () Sim. Se sim: () Ao início da urina () Ao final da urina

HISTÓRIA FAMILIAR (HF) – Doença

ESTILO DE VIDA

Realiza exercícios físicos: () Não () Sim. Qual? _____. Qnts vezes na semana?

_____ Duração em Minutos: _____ Orientado por profissional? () Não () Sim. Qual? _____

Tabagismo: () Não () Sim. Quantos cigarros por dia? _____

Etilismo: () Não () Sim. Tipo de bebida: _____ Número de copos: _____ Qnts vezes na semana: _____

Você possui outro vício? () Não () Sim. Qual? _____


UNIVERSIDADE FEDERAL DO MARANHÃO
FUNDAÇÃO instituída nos termos da Lei n.º 1.162 de 20/07/68
Missão - Inovação
Centro de Ciências Exatas e Tecnologias
Departamento de Engenharia Elétrica
Laboratório de Processamento da Informação Biológica


Quantos copos (americano) de líquidos você costuma tomar por dia? () < 4

() De 4 a 8 () > 8. Quais tipos de líquidos? _____

EXAME FÍSICO/CLÍNICO
Edema:

Mem. Sup.: () Não () Sim, () Bilateral () MSD () MSE () + () ++ () +++

Perfusão periférica: () Boa () Reduzida. Obs: _____

Mem. Inf.: () Não () Sim, () Bilateral () MID () MIE () + () ++ () +++

Perfusão periférica: () Boa () Reduzida. Obs: _____

HIPÓTESE DIAGNÓSTICA (HD) – Definida pelo médico

EXAMES BIOQUÍMICOS (Perfil lipídico completo, ureia, creatinina, insulina, glicemia e hemoglobina glicada se houver, validade do exame 3 meses)

ANAMNESE ALIMENTAR

Total de refeições diárias: _____ Local da maioria das refeições: _____

Alimentos mais consumidos: _____

Alimentos menos consumidos: _____

Uso de edulcorantes: _____

Consumo de embutidos: _____


UNIVERSIDADE FEDERAL DO MARANHÃO
FUNDAÇÃO instituída nos termos da Lei n.º 13.103 de 20/04/1966
 600 anos de fundação

Centro de Ciências Exatas e Tecnologias

Departamento de Engenharia Elétrica

Laboratório de Processamento de Informação Biológica



PARÂMETRO	Medida 1	Medida 2	Medida
Peso hab.			
Peso atual			
Peso Seco			
Peso Ideal			
Estatura			
IMC			
CB			
CC			
CQ			
PCT			
PCB			
PCSI			
PCSE			
BIA			
C Pescoço			
C panturrilha			
PAS			
PAD			
Pulso			

Além do paciente, algum familiar foi orientado? _____

Informações Adicionais: _____

Responsável pela coleta: _____