



**UNIVERSIDADE FEDERAL DO MARANHÃO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM SAÚDE DO ADULTO**



JALILA ANDRÉA SAMPAIO BITTENCOURT

**PREDIÇÃO DE SÍNDROME METABÓLICA EM INDIVÍDUOS COM
DOENÇA RENAL CRÔNICA UTILIZANDO TÉCNICAS DE
APRENDIZADO DE MÁQUINA**

SÃO LUÍS

2021

JALILA ANDRÉA SAMPAIO BITTENCOURT

**PREDIÇÃO DE SÍNDROME METABÓLICA EM INDIVÍDUOS COM
DOENÇA RENAL CRÔNICA UTILIZANDO TÉCNICAS DE
APRENDIZADO DE MÁQUINA**

Dissertação apresentada ao Programa de Pós-Graduação em Saúde do Adulto (PPGSAD) da Universidade Federal do Maranhão – UFMA, como requisito parcial para obtenção do Título de Mestre em Saúde do Adulto.

Orientadora: Profa. Dra. Maria do Desterro Soares Brandão Nascimento

Coorientador: Prof. Dr. Allan Kardec Duailibe Barros Filho

SÃO LUÍS

2021

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

Bittencourt, Jalila Andréa Sampaio.

PREDIÇÃO DE SÍNDROME METABÓLICA EM INDIVÍDUOS COM
DOENÇA RENAL CRÔNICA UTILIZANDO TÉCNICAS DE APRENDIZADO DE
MÁQUINA / Jalila Andréa Sampaio Bittencourt. - 2021.

101 f.

Coorientador(a): Allan Kardec Duailibe Barros Filho.

Orientador(a): Maria do Desterro Soares Brandão

Nascimento.

Dissertação (Mestrado) - Programa de Pós-graduação em
Saúde do Adulto/ccbs, Universidade Federal do Maranhão,
São Luís, 2021.

1. Aprendizado de Máquina. 2. Doença Renal Crônica.
3. Síndrome Metabólica. I. Barros Filho, Allan Kardec
Duailibe. II. Nascimento, Maria do Desterro Soares
Brandão. III. Título.

JALILA ANDRÉA SAMPAIO BITTENCOURT

**PREDIÇÃO DE SÍNDROME METABÓLICA EM INDIVÍDUOS COM
DOENÇA RENAL CRÔNICA UTILIZANDO TÉCNICAS DE
APRENDIZADO DE MÁQUINA**

Dissertação apresentada ao Programa de Pós-Graduação em Saúde do Adulto (PPGSAD) da Universidade Federal do Maranhão – UFMA, como requisito parcial para obtenção do Título de Mestre em Saúde do Adulto.

Aprovada em / /

Profa. Dra. Maria do Desterro Soares Brandão Nascimento
Universidade Federal do Maranhão/PPGSAD (Orientador)

Prof. Dr. Allan Kardec Duailibe Barros Filho
Universidade Federal do Maranhão (Coorientador)

Profa. Dra. Deysianne Costa das Chagas
Universidade Federal do Maranhão (1º examinadora)

Prof. Dr. Marcelo Souza de Andrade
Universidade Federal do Maranhão/PPGSAD (2º examinador)

Profa. Dra. Mayara Cristina Pinto da Silva
Universidade Federal do Maranhão/PPGSAD (3º examinador)

A Deus pelas graças concedidas e por iluminar o meu caminho durante esta jornada.

Aos meus pais e irmãos pelo incentivo e apoio em todos os momentos da minha vida.

Em especial, a minha mãe por nunca ter medido esforços para me proporcionar um ensino de qualidade.

Ao meu amor Yuri Armin, por todo incentivo e paciência.

AGRADECIMENTOS

Agradeço em primeiro lugar a Deus, pela minha vida e por me permitir ultrapassar todos os obstáculos encontrados ao longo da realização deste trabalho.

Aos meus pais, Mary Dalva e Fernando Bittencourt, pelo apoio e incentivo incondicional para a realização dos meus objetivos.

Aos meus irmãos, Anderson e Fernando Jr., pelo incentivo e apoio.

A minha família e aos meus sobrinhos Lucas e Maria Helena, pelo carinho e por compreenderem a minha ausência.

Ao meu amor, Yuri Armin, pela parceria e por sempre colaborar para que este sonho se realizasse.

Aos amigos, da Nutrição, PPGSAD, IPOG, que sempre estiveram ao meu lado, pela amizade incondicional e pela colaboração e apoio demonstrado ao longo de todo o período de tempo em que me dediquei a este trabalho.

Aos amigos e colaboradores do PIB, pela parceria na pesquisa, em especial Carlos Magno, Cindy Lima e Naruna Aritana, por sempre estarem dispostos a me ajudar.

A Profa. Luciane Brito, “in memoriam”, por me acolher e me auxiliar durante boa parte do processo de desenvolvimento deste projeto.

A minha orientadora Profa. Dra. Maria do Desterro e a Profa. Dra. Nilviane Pires, por seus empenhos, orientações e pela paciência com a qual guiaram meu aprendizado.

Ao meu coorientador Prof. Dr. Allan Kardec Barros, pela oportunidade, ensinamentos e auxílio na concretização deste trabalho.

Ao CPDR e PIB, pela infraestrutura e apoio que foram fundamentais para a realização deste trabalho.

A UFMA e ao PPGSAD, pelos ensinamentos e dedicação, essenciais no meu processo de formação profissional.

A FAPEMA, pela concessão da bolsa de mestrado e pelo financiamento concedido que permitiu a realização deste estudo.

“Os sonhos não determinam o lugar onde iremos chegar, mas produzem a força necessária para tirar-nos do lugar em que estamos.”

(Augusto Cury)

RESUMO

A doença renal crônica (DRC) e a síndrome metabólica (SM) possuem íntima ligação com o excesso de peso, obesidade e fatores de risco cardiometabólicos. A detecção precoce destas patologias torna-se necessária, a fim de retardar as complicações a elas associadas. Desta forma, métodos de triagem para o rastreamento de SM são de grande importância, visto que, a SM pode impactar de forma negativa na progressão da DRC. Com base nisso, o estudo teve por objetivo desenvolver um modelo para predição do risco para SM em pessoas com DRC. Trata-se de um estudo transversal, realizado com pacientes oriundos do Centro de Prevenção de Doenças Renais (CPDR) do Hospital Universitário da Universidade Federal do Maranhão (HUUFMA). A amostra foi composta por voluntários de ambos os sexos com idade a partir de 20 anos, classificados de acordo com seu estado de saúde (DRC leve ou DRC grave). Os estágios da DRC foram classificados de acordo com a taxa de filtração glomerular (TFG), o diagnóstico sugestivo de SM foi definido de acordo com o proposto pela *International Diabetes Federation* (IDF) e foram avaliados dados antropométricos, bioquímicos, hemodinâmicos e hábitos de vida. Para o rastreamento da SM, utilizou-se o algoritmo classificador k-vizinhos mais próximos (KNN), um método do aprendizado de máquina (AM) supervisionado. Para implementação do algoritmo classificador, utilizou-se entradas de baixo custo e fácil utilização como: gênero, tabagismo, circunferência do pescoço (CP) e a relação cintura-quadril (RCQ). A construção do algoritmo classificador e implementação do *software* se deu por meio do programa Matlab®. Para o arquivo de dados e a análise estatística, utilizou-se o *software* SPSS®, sendo aplicado os seguintes testes estatísticos: Kolmogorov-Smirnov, *t* de Student, Mann-Whitney U e curva ROC. Os resultados foram considerados estatisticamente significativos para $p < 0,05$. Foram avaliados 196 indivíduos adultos com idade média de $44,73 \pm 15,96$ anos, dos quais 71,9% ($n=141$) corresponderam ao gênero feminino e 69,4% ($n=136$) possuíam excesso de peso. Da amostra investigada, 45,8% ($n=11$; $p=0,006$) dos portadores de DRC possuíam SM, com maioria apresentando até 3 componentes metabólicos alterados. Destes componentes, o grupo com DRC apresentou maiores valores de média/mediana em todos os parâmetros, com significância estatística em: circunferência da cintura (CC) ($94,85 \pm 11,7$; $p=0,02$), pressão arterial sistólica (PAS) [$134(123,25-165,5)$ mmHg; $p < 0,001$], pressão arterial diastólica (PAD) [$86,5(76,25-91)$ mmHg; $p=0,019$] e glicemia em jejum (GJ) [$81(75-88)$ mg/dL; $p=0,001$]. O algoritmo KNN mostrou-se um bom preditor para rastreamento de SM, visto que apresentou acurácia e sensibilidade de 79%, especificidade de 80%, tendo seu desempenho avaliado pela curva ROC (AUC=0,79). Desta forma, o algoritmo KNN pode ser utilizado como método de triagem com alta sensibilidade e baixo custo para avaliar a presença de SM em pessoas com DRC.

Palavras-chave: Doença Renal Crônica. Síndrome Metabólica. Aprendizado de Máquina.

ABSTRACT

The chronic kidney disease (CKD) and metabolic syndrome (MS) are closely linked to overweight, obesity and cardiovascular risk factors. In order to postpone the complications associated with them and due to the increasing incidence in all age groups, the early detection of these pathologies is necessary. Based on this, the study aimed to develop a model to predict the risk for the MS in people with the CKD. This is a cross-sectional study, carried out with patients from the Center for the Prevention of Kidney Diseases (CPDR) of the University Hospital of the Federal University of Maranhão (HUUFMA). The sample was obtained from volunteers of both genders who were 20 years old or over and were classified according to their health status (healthy or with the CKD). The stages of the CKD are classified according to the glomerular filtration rate (GFR) and the suggestive diagnosis of the MS was established according to the proposed by the International Diabetes Federation (IDF). Also, anthropometric, biochemical, hemodynamic, and lifestyle data were evaluated. For the MS tracking, the k-nearest neighbors (KNN) classifier algorithm, that is a supervised machine learning (MA) method, was used. To implement the classifier algorithm, the following entries were used: gender, smoking status, neck circumference (NC) and waist-hip ratio (WHR). The construction of the classifier algorithm and software implementation took place through the MATLAB® program. For the data file and statistical analysis, the SPSS® software was used, and the following statistical tests were applied: Kolmogorov-Smirnov, Student's t, Mann-Whitney U, in addition to the ROC curve. The results were considered statistically significant for $p < 0.05$. This study was approved by the Ethics and Research Committee of the Federal University of Maranhão, with number 67030517.5.0000.5087. A total of 196 adult individuals with a mean age of 44.73 ± 15.96 years were evaluated, of which 71.9% ($n=141$) were female and 69.4% ($n=136$) were overweight, and 12.24% ($n=24$) had CKD. Of the investigated sample, 45.8% ($n=11$; $p=0.006$) of CKD patients had MS, with the majority presenting up to 3 altered metabolic components. Of these components, the group with CKD had higher mean/median values in all parameters, with statistical significance in: waist circumference (WC) (94.85 ± 11.7 ; $p=0.02$), systolic blood pressure (SBP) [$134(123.25-165.5)$ mmHg; $p < 0.001$], diastolic blood pressure (DBP) [$86.5(76.25-91)$ mmHg; $p=0.019$] and fasting glucose (FG) [$81(75-88)$ mg/dL; $p=0.001$]. The KNN algorithm proved to be a good predictor for MS tracking, as it had 79% accuracy and sensitivity, 80% specificity, having its performance evaluated by the ROC curve ($AUC=0.79$). Thus, the KNN algorithm can be used as a screening method with high sensitivity and low cost to assess the presence of MS in people with CKD.

Keywords: Chronic Kidney Disease. Metabolic Syndrome. Machine Learning.

LISTA DE ILUSTRAÇÕES

Figura 1. Efeito iceberg da doença renal crônica Erro! Indicador não definido.

Figura 2. Etapas de desenvolvimento do classificador.....Erro! Indicador não definido.

Figura 3. Área sobre a curva demonstrando o poder discriminatório do algoritmo KNN na predição de SM no conjunto teste. Erro! Indicador não definido.

Figura 4. Layout de interface com o usuário e exemplo com parâmetros de entrada..... Erro! Indicador não definido.

LISTA DE QUADROS E TABELAS

Quadro 1. Estadiamento e classificação da DRC segundo critério do KDIGO, 2012.....	Erro! Indicador não definido.
Quadro 2. Componentes da SM segundo critérios da IDF.....	Erro! Indicador não definido.
Quadro 3. Equação estimativa do ritmo de filtração glomerular em adultos.	Erro! Indicador não definido.
Quadro 4. Métricas de avaliação.	Erro! Indicador não definido.
Quadro 5. Relação entre a área sobre a curva ROC e a precisão do diagnóstico.	Erro! Indicador não definido.
Tabela 1. Características sociodemográficas e estilo de vida da amostra (n=196)	38
Tabela 2. Características antropométricas e hemodinâmicas da amostra (n=196)	39
Tabela 3. Características antropométricas, hemodinâmicas e laboratoriais da amostra estratificada pela presença de DRC (n=196)	40
Tabela 4. Estadiamento e classificação de DRC da amostra (n=196)	41
Tabela 5. Prevalência de SM e seus componentes na amostra (n=196)	42

LISTA DE SIGLAS

ABESO – Associação Brasileira para o Estudo da Obesidade e da Síndrome Metabólica
ADA – *American Diabetes Association*
AM – Aprendizado de Máquina
CB – Circunferência do Braço
CC – Circunferência da Cintura
CKD-EPI – *Chronic Kidney Disease Epidemiology Collaboration*
CP – Circunferência do Pescoço
CPant – Circunferência da Panturrilha
CPDR – Centro de Prevenção de Doenças Renais
DCNT – Doenças Crônicas não Transmissíveis
DCV – Doenças Cardiovasculares
DM – Diabetes Mellitus
EUA – Estados Unidos da América
FRC – Fatores de Risco Cardiovascular
HAS – Hipertensão Arterial Sistêmica
HUUFMA – Hospital Universitário da Universidade Federal do Maranhão
IA – Inteligência Artificial
IDF – *International Diabetes Federation*
IMC – Índice de Massa Corporal
IR – Insuficiência Renal
KDIGO – *Kidney Disease: Improving Global Outcomes*
KNN – K-Vizinhos Mais Próximos
NCEP-ATPIII – *National Cholesterol Education Program-Third Adult Treatment Panel III*
OMS – Organização Mundial da Saúde
OPAS – Organização Pan-Americana de Saúde
PA – Pressão Arterial
PAD – Pressão Arterial Diastólica
PAS – Pressão Arterial Sistólica
POF – Pesquisa de Orçamentos Familiares
RCE – Relação Cintura-Estatura
RCQ – Relação Cintura-Quadril
SM – Síndrome Metabólica
SUS – Sistema Único de Saúde
TCLE – Termo de Consentimento Livre e Esclarecido
TFG – Taxa de Filtração Glomerular

SUMÁRIO

1 INTRODUÇÃO	Erro! Indicador não definido.
2 OBJETIVOS	Erro! Indicador não definido.
2.1 Objetivo Geral	Erro! Indicador não definido.
2.2. Objetivos Específicos	Erro! Indicador não definido.
3 FUNDAMENTAÇÃO TEÓRICA	Erro! Indicador não definido.
3.1 Doença Renal Crônica	Erro! Indicador não definido.
3.1.1 Aspectos Epidemiológicos da DRC	Erro! Indicador não definido.
3.1.2 Estadiamento da Doença Renal	Erro! Indicador não definido.
3.1.2 Fatores de Risco.....	Erro! Indicador não definido.
3.2 Síndrome Metabólica	Erro! Indicador não definido.
3.2.1 Epidemiologia da SM	Erro! Indicador não definido.
3.2.2 Fatores de Risco.....	Erro! Indicador não definido.
3.2.3 Critérios Diagnósticos	Erro! Indicador não definido.
3.3 Aprendizados de Máquina	Erro! Indicador não definido.
4 MATERIAIS E MÉTODOS	Erro! Indicador não definido.
4.1 Delineamento do Estudo	Erro! Indicador não definido.
4.2 Amostra/Amostragem	Erro! Indicador não definido.
4.3 Calculo Amostral	Erro! Indicador não definido.
4.4 Critérios de Inclusão, Não-Inclusão e Exclusão ...	Erro! Indicador não definido.
4.5 Instrumentos de Coleta e Avaliação dos Dados ...	Erro! Indicador não definido.
4.5.1 Questionário Semiestruturado	Erro! Indicador não definido.
4.5.2 Indicadores Antropométricos	Erro! Indicador não definido.
4.5.3 Perfil Hemodinâmico.....	Erro! Indicador não definido.

4.5.4 Avaliação Laboratorial	Erro! Indicador não definido.
4.5.5 Critérios para Síndrome Metabólica.....	Erro! Indicador não definido.
4.5.6 Critérios para Doença Renal Crônica	Erro! Indicador não definido.
4.6 Escolha das características de entrada do algoritmo	Erro! Indicador não definido.
4.7 Implantação do método de aprendizado de máquina.....	Erro! Indicador não definido.
4.8 Etapas do método proposto	Erro! Indicador não definido.
4.8.1 Normalização.....	Erro! Indicador não definido.
4.8.2 Aplicação dos classificadores	Erro! Indicador não definido.
4.8.3 Avaliação dos métodos de classificação.....	Erro! Indicador não definido.
4.8.4 Desenvolvimento do <i>Software</i>	Erro! Indicador não definido.
4.9 Processamento e Tratamento Estatístico.....	Erro! Indicador não definido.
4.10 Aspectos Éticos.....	Erro! Indicador não definido.
5 RESULTADOS E DISCUSSÃO	Erro! Indicador não definido.
7 CONCLUSÃO.....	Erro! Indicador não definido.
REFERÊNCIAS	Erro! Indicador não definido.
ANEXOS	Erro! Indicador não definido.
APÊNDICE.....	Erro! Indicador não definido.