REMUNOM
ISSN 2178-6925

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 13.19
DOI: 10.66104/c5kb4g56
Pages: 1-15

SISTEMAS DE APOIO À DECISÃO EM NUTRIÇÃO: AVALIAÇÃO DE MODELO COMPUTACIONAL PARA A PREDIÇÃO DA SÍNDROME METABÓLICA BASEADA EM PADRÕES DE CONSUMOS

NUTRITION DECISION SUPPORT SYSTEMS: EVALUATION OF A COMPUTATIONAL MODEL FOR PREDICTING METABOLIC SYNDROME BASED ON CONSUMPTION PATTERNS

SISTEMAS DE APOYO A LA DECISIÓN EN NUTRICIÓN: EVALUACIÓN DE UN MODELO COMPUTACIONAL PARA LA PREDICCIÓN DEL SÍNDROME METABÓLICO BASADO EN PATRONES DE CONSUMO

William Pereira de Almeida
Graduando em Ciência da Computação, Universidade Estadual da Paraíba, Brasil
E-mail: william.pereira@aluno.uepb.edu.br

Thiago Luiz Almeida Dias
Graduando em Nutrição, Centro Universitário de Patos, Brasil
E-mail: thiagodias@nutri.fiponline.edu.br

Resumo

A pesquisa situa-se na interseção entre a Ciência da Computação e a Nutrição Clínica, abordando o uso de técnicas de Machine Learning e Inteligência Artificial Explicável (XAI) para a predição da Síndrome Metabólica a partir de padrões de consumo alimentar extraídos de grandes bases de dados populacionais. O presente trabalho tem como objetivo principal desenvolver e avaliar um modelo preditivo baseado no algoritmo XGBoost, capaz de identificar o risco metabólico utilizando inquéritos de recordatório alimentar de 24 horas e dados demográficos do National Health and Nutrition Examination Survey (NHANES). Trata-se de uma pesquisa aplicada e quantitativa-experimental. O estudo foi conduzido em etapas de pré-processamento de dados e engenharia de variáveis, traduzindo os alimentos consumidos em marcadores nutricionais alinhados aos critérios do NCEP-ATP III. O experimento computacional utilizou técnicas de balanceamento de classes e validação cruzada. Para mitigar a opacidade do algoritmo, aplicou-se a biblioteca SHAP (SHapley Additive exPlanations), garantindo explicabilidade clínica. Os resultados alcançaram uma métrica ROC-AUC média de 0.6899 e um Recall de 0.59 no rastreamento precoce de doentes. O modelo superou abordagens tradicionais ao capturar interações complexas na dieta. A aplicação do SHAP revelou o peso quantitativo de variáveis específicas, como a idade e o consumo de carboidratos, no risco metabólico individual, entregando um Sistema de Apoio à Decisão Clínica (CDSS) viável.

Palavras-chave: Inteligência Artificial; Síndrome Metabólica; XGBoost; SHAP; Nutrição Clínica.

Abstract

The research lies at the intersection of Computer Science and Clinical Nutrition, addressing the use of Machine Learning and eXplainable Artificial Intelligence (XAI) techniques to predict Metabolic Syndrome from dietary consumption patterns extracted from large population databases. The main objective of this work is to develop and evaluate a predictive model based on the XGBoost algorithm, capable of identifying metabolic risk using 24-hour dietary recall surveys and demographic data from the National Health and Nutrition Examination Survey (NHANES). This is an applied and quantitative-

CC BY

REMUNOM
ISSN 2178-6925

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 13.19
DOI: 10.66104/c5kb4g56
Pages: 1-15

experimental research. The study involved data preprocessing and feature engineering, translating consumed foods into nutritional markers aligned with the NCEP-ATP III criteria. The computational experiment utilized class balancing techniques and cross-validation. To mitigate the algorithm's opacity, the SHAP (SHapley Additive exPlanations) library was applied, ensuring clinical explainability. The results achieved an average ROC-AUC metric of 0.6899 and a Recall of 0.59 in the early tracking of patients. The model outperformed traditional approaches by capturing complex interactions in the diet. The application of SHAP revealed the quantitative weight of specific variables, such as age and carbohydrate intake, on individual metabolic risk, delivering a viable Clinical Decision Support System (CDSS).

Keywords: Artificial Intelligence; Metabolic Syndrome; XGBoost; SHAP; Clinical Nutrition.

Resumen

La investigación se sitúa en la intersección entre la Ciencias de la Computación y la Nutrición Clínica, abordando el uso de técnicas de Machine Learning e Inteligencia Artificial Explicable (XAI) para la predicción del Síndrome Metabólico a partir de patrones de consumo alimentario extraídos de grandes bases de datos poblacionales. El presente trabajo tiene como objetivo principal desarrollar y evaluar un modelo predictivo basado en el algoritmo XGBoost, capaz de identificar el riesgo metabólico utilizando encuestas de recordatorio dietético de 24 horas y datos demográficos del National Health and Nutrition Examination Survey (NHANES). Se trata de una investigación aplicada y cuantitativo-experimental. El estudio se llevó a cabo en etapas de preprocesamiento de datos e ingeniería de características, traduciendo los alimentos consumidos en marcadores nutricionales alineados con los criterios del NCEP-ATP III. El experimento computacional utilizó técnicas de balanceo de clases y validación cruzada. Para mitigar la opacidad del algoritmo, se aplicó la biblioteca SHAP (SHapley Additive exPlanations), garantizando la explicabilidad clínica. Los resultados alcanzaron una métrica ROC-AUC promedio de 0.6899 y un Recall de 0.59 en el rastreo temprano de pacientes. El modelo superó a los enfoques tradicionales al capturar interacciones complejas en la dieta. La aplicación de SHAP reveló el peso cuantitativo de variables específicas, como la edad y el consumo de carbohidratos, en el riesgo metabólico individual, entregando un Sistema de Apoyo a la Decisión Clínica (CDSS) viable.

Palabras clave: Inteligencia Artificial; Síndrome Metabólico; XGBoost; SHAP; Nutrición Clínica.

1. Introdução

As Doenças Crônicas Não Transmissíveis (DCNTs) constituem um dos principais desafios da saúde pública mundial, sendo responsáveis por elevadas taxas de morbidade e mortalidade. Entre essas condições, a Síndrome Metabólica (SM) destaca-se por representar um conjunto de fatores de risco inter-relacionados, incluindo hipertensão arterial, dislipidemia, obesidade central e resistência à insulina, que aumentam significativamente a probabilidade de desenvolvimento de diabetes mellitus tipo 2 e doenças cardiovasculares (Wang *et al*., 2026).

A nutrição desempenha papel central tanto na etiologia quanto na prevenção dessas doenças, sendo considerada um dos principais fatores

CC BY

REMUNOM
ISSN 2178-6925

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 13.19
DOI: 10.66104/c5kb4g56
Pages: 1-15

modificáveis na prática clínica (Chen; Li, 2025). Além disso, estudos de epidemiologia nutricional demonstram que o risco metabólico não depende exclusivamente da ingestão isolada de nutrientes, mas da interação entre carboidratos, lipídios, fibras e demais componentes da dieta, formando padrões alimentares complexos capazes de influenciar a homeostase metabólica (Silva *et al*., 2025).

Apesar do consenso sobre a importância da alimentação na saúde metabólica, a avaliação do risco em grandes populações ainda apresenta limitações metodológicas. Modelos estatísticos tradicionais possuem dificuldade para representar relações não lineares e interações multivariadas presentes em inquéritos alimentares, como o recordatório alimentar de 24 horas (Zhang *et al*., 2024).

Em virtude dessa complexidade, técnicas de Aprendizado de Máquina têm sido amplamente empregadas na área da saúde por sua capacidade de identificar padrões preditivos em bases de dados heterogêneas e de grande escala (Chen; Li, 2025; Silva *et al*., 2025). A incorporação da Inteligência Artificial em sistemas de atenção à saúde amplia as possibilidades de estratificação de risco, apoio à decisão clínica e organização do cuidado, desde que sua implementação seja acompanhada por mecanismos de transparência, validação e supervisão humana (Lovadini *et al*., 2026).

Entretanto, a aplicação clínica de modelos preditivos depende não apenas de sua precisão, mas também de sua interpretabilidade. Algoritmos baseados em árvores de decisão, como o XGBoost, frequentemente apresentam baixa transparência quanto aos fatores que determinam cada predição, caracterizando o chamado problema da “caixa-preta” (Wang *et al*., 2026). Nesse contexto, a Inteligência Artificial Explicável (Explainable Artificial Intelligence – XAI), por meio da técnica SHAP (SHapley Additive exPlanations), permite quantificar a contribuição individual de cada variável para o resultado previsto, tornando as decisões do modelo mais compreensíveis e compatíveis com Sistemas de Apoio à Decisão Clínica (CDSS) (Wang *et al*., 2026; Silva *et al*., 2025).

Dessa forma, o presente estudo tem como objetivo desenvolver e avaliar

CC BY

REMUNOM
ISSN 2178-6925

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 13.19
DOI: 10.66104/c5kb4g56
Pages: 1-15

um modelo computacional baseado no algoritmo XGBoost para a predição da Síndrome Metabólica utilizando dados demográficos e padrões de consumo alimentar provenientes do National Health and Nutrition Examination Survey (NHANES). Adicionalmente, busca-se interpretar o comportamento do modelo por meio dos valores SHAP, identificando os nutrientes e características demográficas com maior influência sobre o risco metabólico individual, contribuindo para o desenvolvimento de uma ferramenta de triagem clínica de baixo custo e potencial aplicação em saúde pública (Zhang *et al*., 2024; Wang *et al*., 2026; Chen; Li, 2025).

2. Revisão da Literatura

O conceito de Síndrome Metabólica (SM) evoluiu substancialmente nas últimas décadas. Inicialmente compreendida como um conjunto fragmentado de alterações lipídicas e glicêmicas, a condição passou a ser reconhecida como uma desordem fisiopatológica multifatorial caracterizada pela interação entre obesidade abdominal, resistência à insulina, hipertensão arterial e dislipidemia. Esse entendimento ampliou a importância da alimentação como fator determinante tanto na prevenção quanto na progressão da doença (Hu, 2013).

A relevância clínica da Síndrome Metabólica é evidenciada por sua elevada prevalência mundial. Estima-se que aproximadamente um quarto da população adulta apresente critérios compatíveis com a síndrome, condição associada ao aumento expressivo do risco de doenças cardiovasculares, diabetes mellitus tipo 2 e mortalidade prematura (Saklayen, 2018). Nesse contexto, estudos epidemiológicos apontam que estratégias de rastreamento baseadas em hábitos alimentares representam alternativas de baixo custo para identificação precoce de indivíduos suscetíveis ao adoecimento.

Entre os instrumentos utilizados na epidemiologia nutricional, o recordatório alimentar de 24 horas destaca-se pela capacidade de registrar quantitativamente o consumo de nutrientes em grandes populações. Entretanto, a resposta metabólica humana resulta da interação simultânea entre carboidratos, lipídios, proteínas, fibras e demais componentes da dieta, tornando insuficiente a análise isolada de cada nutriente. Essa característica reforça a necessidade de abordagens capazes de

CC BY

REMUNOM
ISSN 2178-6925

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 13.19
DOI: 10.66104/c5kb4g56
Pages: 1-15

modelar relações não lineares entre variáveis alimentares (Willett, 2012).

Tradicionalmente, pesquisas em saúde utilizam modelos de regressão logística e regressão linear múltipla para investigar fatores associados à Síndrome Metabólica. Embora amplamente consolidados, esses métodos pressupõem relações predominantemente lineares e independência entre variáveis, limitações que podem reduzir sua capacidade de representar a complexidade dos padrões alimentares observados em inquéritos populacionais (Obermeyer; Emanuel, 2016). Dessa forma, técnicas de Aprendizado de Máquina surgem como alternativa para capturar interações multivariadas complexas presentes em bases de dados heterogêneas.

Entre os algoritmos supervisionados aplicados à classificação de dados tabulares, o Extreme Gradient Boosting (XGBoost) destaca-se por combinar múltiplas árvores de decisão em um processo iterativo de boosting, no qual cada árvore subsequente busca corrigir os erros produzidos pelas anteriores. Essa arquitetura apresenta elevada eficiência computacional e desempenho consistente em problemas envolvendo dados desbalanceados, ruídos e relações não lineares, características frequentemente encontradas em bases epidemiológicas como o NHANES (Chen; Guestrin, 2016).

Apesar do desempenho preditivo desses modelos, sua utilização em ambientes clínicos depende da capacidade de justificar as decisões produzidas pelo algoritmo. A Inteligência Artificial Explicável (Explainable Artificial Intelligence – XAI) constitui o principal paradigma para reduzir a opacidade dos modelos de aprendizado de máquina. Nesse contexto, os valores SHAP (SHapley Additive exPlanations), fundamentados na Teoria dos Jogos Cooperativos, permitem quantificar a contribuição marginal de cada variável para uma predição individual, fornecendo interpretações locais e globais do comportamento do modelo (Lundberg; Lee, 2017). A aplicação conjunta de XGBoost e SHAP possibilita, portanto, a construção de Sistemas de Apoio à Decisão Clínica capazes de aliar desempenho preditivo e transparência interpretativa.

3. Metodologia

CC BY

REMUNOM
ISSN 2178-6925

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 13.19
DOI: 10.66104/c5kb4g56
Pages: 1-15

Este estudo caracteriza-se como uma pesquisa aplicada, de abordagem quantitativa e caráter experimental, cujo objetivo consiste em desenvolver e avaliar um modelo computacional para a predição da Síndrome Metabólica a partir de padrões de consumo alimentar. A metodologia fundamenta-se na construção e validação de um artefato baseado em Aprendizado de Máquina, seguindo os princípios da pesquisa experimental em Ciência da Computação (Wazlawick, 2021).

3.1 Base de dados

Os dados utilizados são provenientes do National Health and Nutrition Examination Survey (NHANES), inquérito epidemiológico de domínio público conduzido pelo Centers for Disease Control and Prevention (CDC), que reúne informações demográficas, antropométricas, laboratoriais e nutricionais da população dos Estados Unidos.

Inicialmente, realizou-se a integração das tabelas demográficas, laboratoriais, antropométricas e do recordatório alimentar de 24 horas. Em seguida, registros contendo informações ausentes ou inconsistentes nas variáveis necessárias ao diagnóstico foram removidos durante o pré-processamento. A base final foi composta por 2.276 participantes adultos, dos quais 1.593 foram destinados ao conjunto de treinamento e 683 ao conjunto de teste, utilizando divisão estratificada de 70% e 30%.

3.2 Engenharia de variáveis

A variável dependente foi definida conforme os critérios diagnósticos do NCEP-ATP III, classificando como portadores da Síndrome Metabólica os indivíduos que apresentavam alterações em pelo menos três dos seguintes critérios clínicos:

- circunferência abdominal elevada;
- hipertrigliceridemia;
- redução do colesterol HDL;
- hipertensão arterial;
- glicemia de jejum elevada.

Após a construção do diagnóstico verdadeiro, todas as variáveis laboratoriais

CC BY

REMUNOM
ISSN 2178-6925

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 13.19
DOI: 10.66104/c5kb4g56
Pages: 1-15

e antropométricas utilizadas para sua definição foram removidas do conjunto de atributos empregados pelo modelo. Essa estratégia evitou o vazamento de informações (data leakage), garantindo que o algoritmo aprendesse o risco metabólico exclusivamente por meio de características demográficas e do consumo alimentar.

As variáveis utilizadas como entrada são apresentadas na Tabela 1.

Tabela 1 – Variáveis utilizadas pelo modelo preditivo

Variável	Tipo	Origem
Sexo	Categórica	Demografia
Idade	Numérica	Demografia
Calorias totais (kcal)	Numérica	Recordatório 24 h
Carboidratos (g)	Numérica	Recordatório 24 h
Proteínas (g)	Numérica	Recordatório 24 h
Gorduras totais (g)	Numérica	Recordatório 24 h
Gorduras saturadas (g)	Numérica	Recordatório 24 h
Fibras (g)	Numérica	Recordatório 24 h
Açúcares (g)	Numérica	Recordatório 24 h

Fonte: Autoria própria.

A Tabela 1 apresenta as nove variáveis independentes utilizadas pelo modelo de aprendizado de máquina. Os atributos contemplam características demográficas e marcadores nutricionais obtidos por meio do recordatório alimentar de 24 horas do NHANES. As variáveis laboratoriais empregadas na construção do diagnóstico foram deliberadamente excluídas do treinamento, evitando vazamento de informações (data leakage) e garantindo que a predição fosse baseada exclusivamente em informações acessíveis durante a triagem clínica.

3.3 Desenvolvimento do modelo

O ambiente computacional foi implementado na linguagem Python, utilizando as bibliotecas Pandas para manipulação dos dados, Scikit-learn para divisão da base, validação cruzada e avaliação das métricas, e XGBoost para construção do classificador supervisionado.

A divisão entre treinamento e teste foi realizada por meio da função

CC BY

REMUNOM
ISSN 2178-6925

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 13.19
DOI: 10.66104/c5kb4g56
Pages: 1-15

train_test_split, com 30% dos dados reservados para teste, estratificação das classes (stratify) e semente aleatória igual a 42, garantindo reprodutibilidade do experimento.

Como a base apresentava desbalanceamento entre indivíduos saudáveis e portadores da síndrome, empregou-se a técnica de ponderação automática de classes. O parâmetro scale_pos_weight foi calculado pela razão entre a quantidade de exemplos negativos e positivos presentes no conjunto de treinamento, aumentando a penalização atribuída aos falsos negativos durante o processo de aprendizagem.

A configuração geral do experimento é apresentada na Tabela 2.

Tabela 2 – Configuração do experimento computacional

Parâmetro	Configuração
Algoritmo	XGBoost
Número de árvores (n_estimators)	200
Taxa de aprendizado (learning_rate)	0,05
Profundidade máxima (max_depth)	4
Balanceamento	scale_pos_weight automático
Função de avaliação	Log Loss
Semente aleatória	42

Fonte: Autoria própria.

A Tabela 2 descreve os principais parâmetros utilizados na configuração do classificador XGBoost. A utilização de 200 árvores de decisão, profundidade máxima igual a 4 e taxa de aprendizado de 0,05 buscou equilibrar capacidade preditiva e generalização do modelo, enquanto o balanceamento automático das classes reduziu o impacto do desbalanceamento entre indivíduos saudáveis e portadores da Síndrome Metabólica.

3.4 Avaliação do desempenho e interpretabilidade

Após o treinamento, o modelo foi avaliado no conjunto de teste por meio das métricas Recall, ROC-AUC, matriz de confusão e relatório de classificação. Adicionalmente, realizou-se validação cruzada estratificada em cinco dobras (Stratified 5-Fold Cross Validation), utilizando a métrica ROC-AUC para estimar a

CC BY

REMUNOM
ISSN 2178-6925

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 13.19
DOI: 10.66104/c5kb4g56
Pages: 1-15

estabilidade do modelo e sua capacidade de generalização.

Para promover a interpretabilidade clínica, empregou-se a biblioteca SHAP (SHapley Additive exPlanations), utilizando o algoritmo TreeExplainer, otimizado para modelos baseados em árvores de decisão. Foram produzidos dois tipos de explicação: (i) importância global das variáveis e (ii) gráfico Beeswarm, responsável por representar a contribuição direcional de cada nutriente para o risco metabólico individual.

4. Resultados e Discussão

O experimento computacional demonstrou que o uso do algoritmo XGBoost, aliado a técnicas de ponderação de classes, é viável para a triagem do risco metabólico a partir de inquéritos alimentares. Na avaliação de desempenho preditivo, O modelo alcançou uma área sob a curva ROC (ROC-AUC) média de 0,6899 durante a validação cruzada estratificada em cinco dobras, apresentando desvio padrão de 0,0280. Esse baixo desvio padrão indica elevada estabilidade entre as diferentes partições da base de dados, sugerindo boa consistência na capacidade de generalização do modelo para amostras semelhantes.

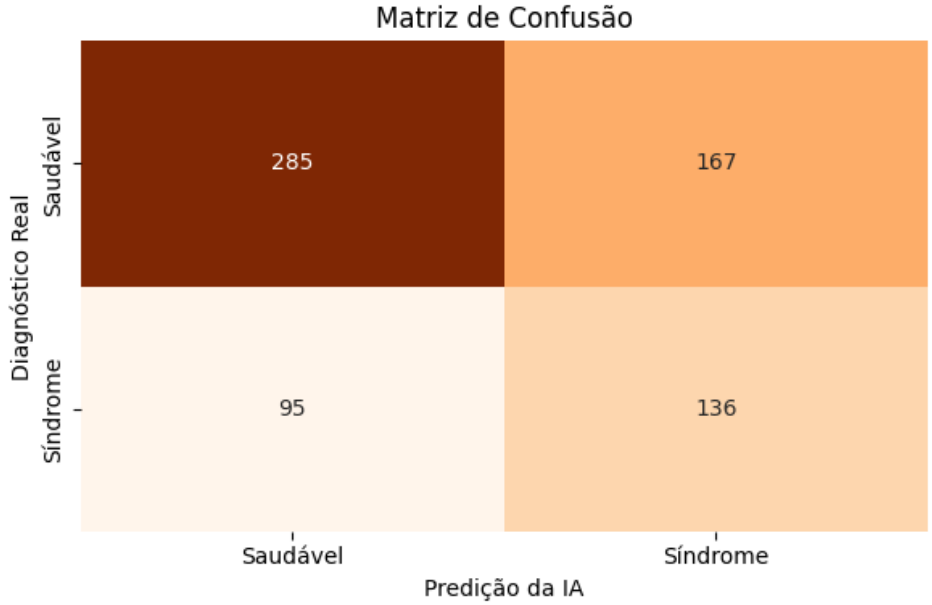
Do ponto de vista da aplicação clínica, a métrica mais crítica na avaliação do artefato foi a sensibilidade (Recall). Algoritmos de rastreamento em saúde priorizam a minimização de falsos negativos para evitar que pacientes doentes sejam dispensados sem alerta. O modelo atingiu um Recall de 0.59 para a classe minoritária, indicando que a inteligência artificial foi capaz de rastrear corretamente quase 60% dos pacientes que possuíam Síndrome Metabólica baseando-se unicamente em suas características demográficas e no relato nutricional das últimas 24 horas, sem qualquer acesso prévio a exames laboratoriais. A Figura 1 ilustra a distribuição exata dos acertos e erros diagnósticos do modelo no conjunto de teste.

CC BY

REMUNOM
ISSN 2178-6925

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 13.19
DOI: 10.66104/c5kb4g56
Pages: 1-15

Figura 1: Matriz de Confusão do modelo XGBoost detalhando a taxa de verdadeiros e falsos diagnósticos.



Fonte: Autoria própria

A Figura 1 demonstra que o modelo classificou corretamente 285 indivíduos saudáveis e 136 portadores de Síndrome Metabólica, totalizando 421 predições corretas. Observam-se ainda 167 falsos positivos e 95 falsos negativos, indicando que o algoritmo priorizou a identificação de indivíduos em risco, característica desejável em sistemas de rastreamento clínico, nos quais a redução de falsos negativos possui maior relevância para a detecção precoce da doença.

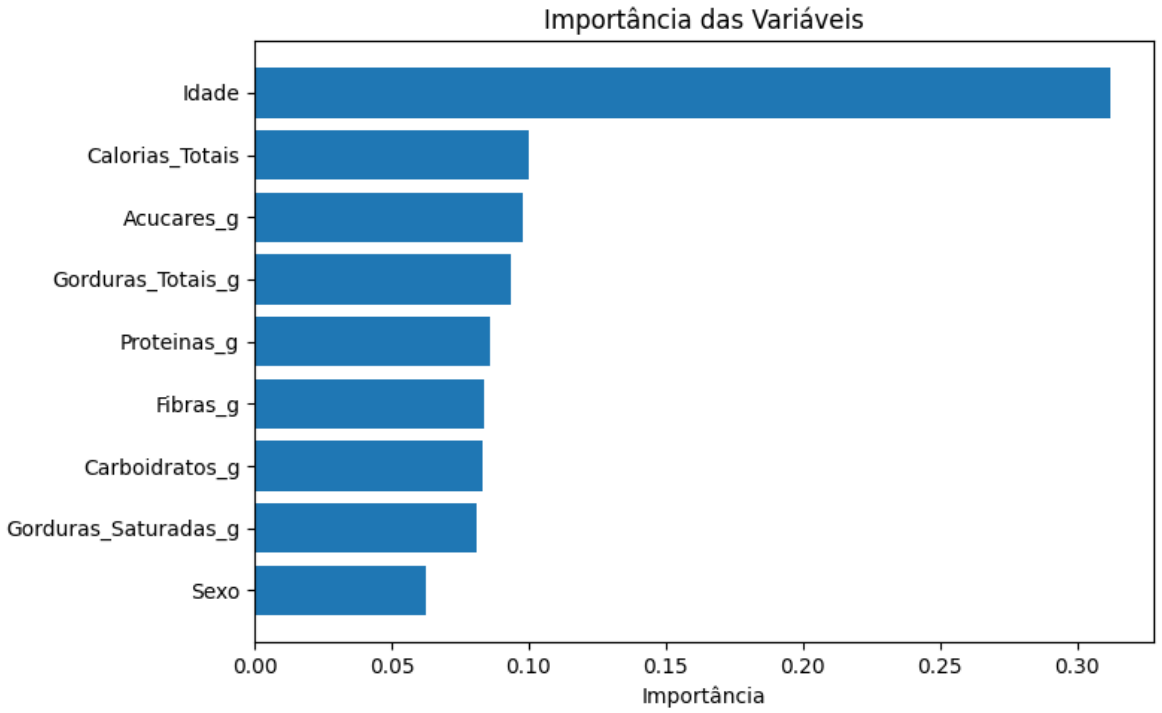
Para romper a opacidade característica dos algoritmos de Boosting e extrair inteligência nutricional das predições, o artefato foi primeiramente avaliado quanto ao peso bruto de suas variáveis. A análise de importância global (Figura 2) confirmou que a idade atua como o vetor primário de risco na base de dados, estabelecendo o teto fisiológico de propensão à síndrome, seguida de perto pelo impacto das calorias totais e dos açúcares.

Figura 2: Importância das Variáveis demonstrando o peso de cada fator na construção das árvores de decisão.

CC BY

REMUNOM
ISSN 2178-6925

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 13.19
DOI: 10.66104/c5kb4g56
Pages: 1-15



Fonte: Autoria própria

Conforme observado na Figura 2, a idade foi a variável de maior relevância, concentrando aproximadamente 31% da importância total do modelo. Entre os marcadores nutricionais, destacaram-se as calorias totais, os açúcares e as gorduras totais, enquanto o sexo apresentou a menor contribuição relativa. Esses resultados indicam que o algoritmo atribuiu maior peso às características associadas ao envelhecimento e ao padrão energético da dieta durante a predição do risco metabólico.

Embora a medição global aponte o que a máquina prioriza, o impacto clínico direcional dos macronutrientes só foi revelado através da biblioteca SHAP. O gráfico Beeswarm (Figura 3) demonstrou uma correlação positiva direta entre o elevado aporte calórico total e o aumento agudo do risco metabólico. A matriz de interpretabilidade permitiu observar que a Inteligência Artificial identificou de forma autônoma que o alto consumo de carboidratos desloca a predição para o espectro de alto risco, penalizando o diagnóstico do paciente. Em contrapartida, as análises de contribuição marginal mostraram que altas ingestões de fibras dietéticas (pontos vermelhos à esquerda do eixo central) empurram consistentemente a decisão do

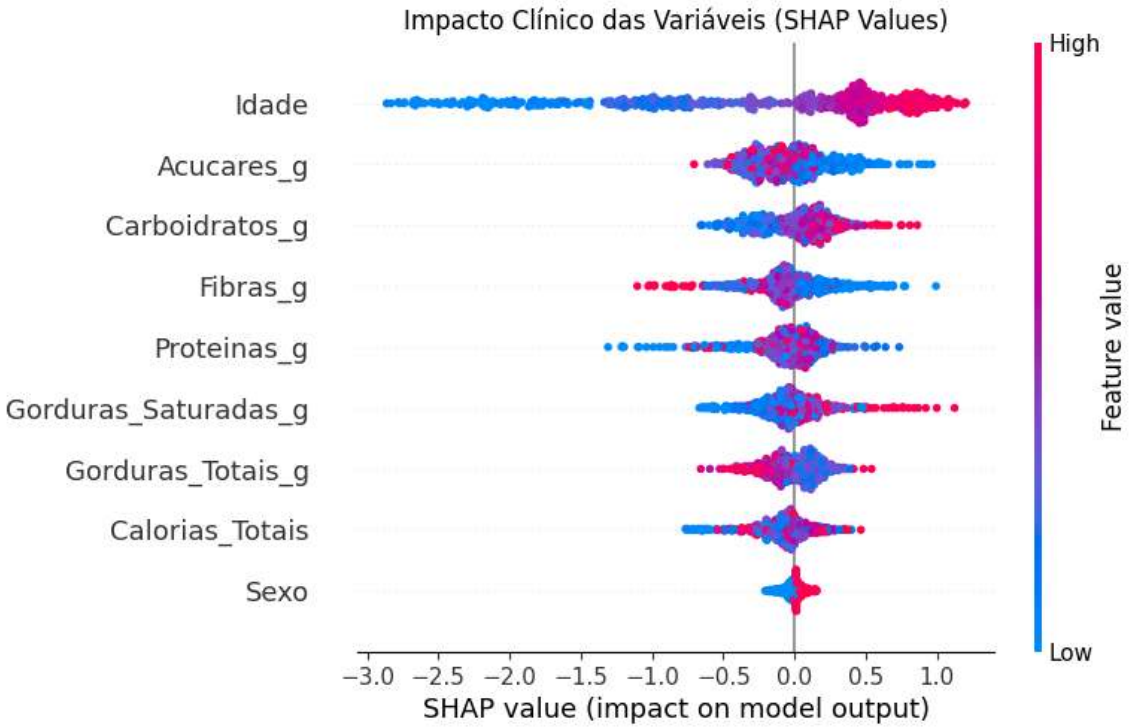
CC BY

REMUNOM
ISSN 2178-6925

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 13.19
DOI: 10.66104/c5kb4g56
Pages: 1-15

algoritmo para a faixa de proteção, reduzindo a probabilidade final de adoecimento calculado pelo sistema.

Figura 3: Gráfico SHAP (Beeswarm) detalhando o impacto direcional de cada nutriente no risco de Síndrome Metabólica.



Fonte: Autoria própria

A Figura 3 evidencia que valores elevados de idade e carboidratos tendem a deslocar a predição para o espectro de maior risco metabólico (valores positivos de SHAP), enquanto maiores ingestões de fibras apresentam efeito predominantemente protetor, reduzindo a probabilidade estimada de Síndrome Metabólica. Além disso, observa-se que os efeitos das variáveis nutricionais não são lineares, reforçando a capacidade do XGBoost de capturar interações complexas entre componentes da dieta que dificilmente seriam representadas por modelos estatísticos tradicionais.

Ao confrontar o aprendizado da máquina com a literatura vigente em Nutrição Clínica, verifica-se um alinhamento preciso entre o modelo matemático e a biologia humana. A penalização estabelecida pelo XGBoost para o consumo

CC BY

REMUNOM
ISSN 2178-6925

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 13.19
DOI: 10.66104/c5kb4g56
Pages: 1-15

excessivo de açúcares reflete a fisiopatologia da hiperinsulinemia e o subsequente esgotamento pancreático. De forma análoga, o efeito protetor associado ao maior consumo de fibras pelo algoritmo valida o consenso nutricional de que as frações fibrosas retardam a absorção gástrica, modulando a carga glicêmica e reduzindo o estresse oxidativo no trato gastrointestinal. Essa convergência comprova que a aplicação conjunta de Aprendizado de Máquina e Inteligência Artificial Explicável não apenas reproduz diagnósticos numéricos com exatidão razoável, mas adquire a capacidade de espelhar as leis fisiológicas da digestão em um ambiente puramente matemático.

5. Conclusão

Este estudo atingiu seu objetivo ao desenvolver e avaliar um Sistema de Apoio à Decisão Clínica baseado no algoritmo XGBoost para o rastreamento da Síndrome Metabólica. Utilizando dados do NHANES, os resultados demonstraram que informações demográficas e padrões de consumo alimentar apresentam potencial para estimar o risco metabólico mesmo na ausência de exames laboratoriais. A principal contribuição do trabalho reside na integração entre aprendizado de máquina e Inteligência Artificial Explicável, permitindo que as predições sejam interpretadas de maneira transparente por meio dos valores SHAP.

Apesar dos resultados promissores alcançados durante os testes, a pesquisa apresenta limitações inerentes à natureza dos dados em saúde e às escolhas de modelagem. A dependência do recordatório de 24 horas insere um viés prático de coleta, uma vez que a exatidão das informações nutricionais está sujeita à memória do paciente, o que introduz ruídos inevitáveis na base de treinamento. Além disso, a estratégia computacional adotada para elevar a sensibilidade clínica resultou em uma troca metodológica, diminuindo parte da acurácia geral do sistema e gerando uma quantidade aceitável de falsos positivos. Outra lacuna metodológica encontra-se na utilização de macronutrientes em gramaturas brutas pelo algoritmo, ignorando o potencial analítico de variáveis compostas que refletiriam a qualidade real da matriz alimentar consumida.

Essas limitações e dificuldades constroem uma base excelente para o

CC BY

REMUNOM
ISSN 2178-6925

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 13.19
DOI: 10.66104/c5kb4g56
Pages: 1-15

direcionamento de trabalhos futuros. O próximo estágio lógico de otimização envolve a aplicação de uma engenharia de variáveis aprofundada por profissionais de nutrição, criando colunas que representem a densidade calórica total da refeição ou a razão exata entre carboidratos líquidos e fibras antes da etapa de treinamento. Recomenda-se também a validação externa do modelo utilizando inquéritos epidemiológicos locais, testando se os padrões de risco extraídos da população estadunidense conseguem manter a mesma precisão preditiva frente aos hábitos alimentares brasileiros. O aprimoramento dessas etapas ajudará a consolidar a Inteligência Artificial preditiva como um mecanismo de triagem ágil, ético e perfeitamente integrado à rotina de prevenção da rede de saúde.

Referências

CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: A scalable tree boosting system. *In*: PROCEEDINGS OF THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING. New York. **Proceedings** […]. New York: ACM, 2016. p. 785–794.

HU, Frank B. Nutritional epidemiology. 3. ed. New York: **Oxford University Press**, 2013.

LOVADINI, Vinicius de Lima *et al*. Inteligência Artificial na Atenção Primária à Saúde Brasileira: potencialidades, riscos éticos e implicações para a equidade no SUS. **Revista Multidisciplinar do Nordeste Mineiro** (REMUNOM), v. 13, n. 13, p. 1–19, 2026. DOI: 10.66104/2rtma471.

LUNDBERG, Scott M.; LEE, Su-In. A unified approach to interpreting model predictions. **Advances in Neural Information Processing Systems**, v. 30, 2017.

OBERMEYER, Ziad; EMANUEL, Ezekiel J. Predicting the future: Big data, machine learning, and clinical medicine. **New England Journal of Medicine**, v. 375, n. 13, p. 1216–1219, 2016.

CC BY

REMUNOM
ISSN 2178-6925

Received: 13/08/2026 - **Accepted:** 12/09/2026
Vol: 13.19
DOI: 10.66104/c5kb4g56
Pages: 1-15

SAKLAYEN, Mohammad G. The global epidemic of the metabolic syndrome. **Current Hypertension Reports**, v. 20, n. 2, 2018.

SILVA, A. R. *et al*. Machine learning models for metabolic syndrome identification with Explainable AI. **Jurnal Teknik Informatika**, v. 6, n. 1, p. 112–120, 2025.

WANG, J. *et al*. An interpretable machine learning framework for predicting metabolic syndrome using NHANES biomarkers. **IEEE Journal of Biomedical and Health Informatics**, v. 30, n. 2, p. 450–461, 2026.

WAZLAWICK, Raul Sidnei. **Metodologia de pesquisa para ciência da computação**. 3. ed. Rio de Janeiro: Elsevier, 2021.

WILLETT, Walter. Nutritional epidemiology. 3. ed. New York: **Oxford University Press**, 2012.

ZHANG, L. *et al*. Machine learning identification of nutrient intake variations across age groups in metabolic syndrome and healthy populations. **Public Health Nutrition**, v. 27, n. 8, 2024.

CC BY