Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 12, n. 1, 2026.

Trabalho apresentado no XLIV CNMAC, Fundação Getulio Vargas - Rio de Janeiro - RJ, 2025.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Algoritmos J48 e *Multilayer Perceptron* como Ferramentas para Predição do Diagnóstico de Resistência à Insulina em Adultos no Brasil

Leandro S. Teixeira[1]
IF Baiano, Catu, BA
Laércio L. Vendite[2]
IMECC/UNICAMP, Campinas, SP
Bruno Geloneze[3]
LIMED/OCRC-CEPID/UNICAMP, Campinas, SP
Ana Carolina J. Vasques[4]
FCA/UNICAMP, Campinas, SP

Resumo. A resistência à insulina (RI) é considerada um dos principais fatores de risco para o desenvolvimento do diabetes tipo 2 e ainda tem componente genético não completamente entendido. Há diferentes métodos para a avaliação da RI, com custos e complexidade distintos. Como alternativa a métodos mais complexos e invasivos, podem-se buscar padrões nas medidas antropométricas e da composição corporal, relacionando-as ao diagnóstico positivo ou negativo da RI. Para compreender a influência dos atributos envolvidos em seu diagnóstico, podem-se utilizar modelos experimentais, com base em dados obtidos junto a pacientes considerados saudáveis e com resistência à insulina. Nos últimos anos, a coleta de dados tem evoluído bastante em diversas áreas, incluindo nas ciências da saúde. Paralelamente, as técnicas de mineração de dados e aprendizagem de máquina têm sido utilizadas para o tratamento dos dados obtidos, visando a geração de conhecimento e aperfeiçoamento de técnicas. Este trabalho visou analisar o banco de dados do *Brazilian Metabolic Syndrome Study* (Brams) usando o *software* WEKA, com os algoritmos J48 (árvores de decisão) e *Multilayer Perceptron*. Ambos algoritmos tiveram resultados muito parecidos, mas as árvores de decisão trouxeram uma leitura mais simples. O J48 possibilitou obter modelos com atributos importantes e pontos de cortes para o diagnóstico de RI, sendo viável a sua utilização por profissionais da área da saúde. Concluiu-se que os atributos de insulina e glicose tiveram grande influência no resultado da RI. Valores de insulina maiores do que 13 e de glicose maiores do que 79 indicaram resistência à insulina. Por outro lado, insulina menor do que ou igual a 9 e glicose menor do que ou igual a 121 indicaram ausência dessa resistência. As medidas antropométricas (dentre elas, a medida da circunferência do pescoço) também se mostraram úteis para auxiliar no diagnóstico da resistência à insulina. No caso da medida da circunferência do pescoço, para os homens, o ponto de corte encontrado foi 42 cm (onde a resistência era mais comum em valores maiores do que esse). Para mulheres, o ponto de corte encontrado foi de 39 cm, com acurácia maior para valores menores do que ou iguais a 36 cm, com indicativo de não haver resistência à insulina.

Palavras-chave. Árvores de Decisão, Circunferência do Pescoço, *Machine Learning*, Mineração de Dados, Redes Neurais

[1]leandro.teixeira@ifbaiano.edu.br
[2]lvendite@unicamp.br
[3]geloneze@unicamp.br
[4]anavaq@unicamp.br

1 Introdução

A resistência à insulina (RI) se caracteriza como uma resposta bioquímica subnormal a concentrações normais de insulina, sendo fator de risco chave para o desenvolvimento de diabetes tipo 2 [3]. Encontra-se associada à presença de obesidade, estando mais acentuada em indivíduos com excesso de adiposidade abdominal e ectópica, além de possuir forte componente genético, e associação com dieta, estresse, privação de sono, atividade física e uso de fármacos [2, 4, 6, 11]. No atual ambiente obesogênico em que vivemos, a RI tem se tornado cada vez mais frequente e uma proporção substancial de indivíduos aparentemente saudáveis apresenta resistência à insulina [3].

Diversos são os métodos disponíveis para a avaliação da resistência à insulina, variando em custo e complexidade. Em 1985, Mathews et al. (1985) validaram o índice HOMA-IR (*homeostatic model assessment – insulin resistance*) frente ao *clamp* [9]. Trata-se de uma equação matemática que avalia a RI a partir das dosagens de glicemia e insulina de jejum. O HOMA-IR tem sido amplamente utilizado e seu espectro nas populações possui forte influência étnica, demandando pontos de corte específicos para uma adequada diferenciação do estado de RI [8].

Para compreender a influência das diferentes variáveis (atributos) clínicas e metabólicas associadas ao quadro de RI, podem-se utilizar modelos experimentais, com base em dados obtidos de pacientes considerados saudáveis e com resistência à insulina. Porém, analisar um atributo individualmente ou sem a devida profundidade pode gerar conclusões equivocadas.

A mineração de dados usa técnicas para encontrar e descrever padrões estruturais em dados como ferramentas para ajudar a explicar esses dados e fazer previsões a partir desses [15]. Por isso, a utilização de algumas técnicas de mineração de dados e aprendizagem de máquina aparece como opção interessante para subsidiar os profissionais da área de saúde na devida análise dos dados obtidos junto aos pacientes. De modo simples, mineração de dados se refere ao processo de extração ou mineração de conhecimento de grandes quantidades de dados. Como a busca é pelo conhecimento que pode ser extraído dos bancos de dados, mais apropriadamente, poderia ser entendida como mineração de conhecimentos a partir dos dados [1, 7].

Este trabalho tem a finalidade de usar algoritmos de árvores de decisão (J48) e redes neurais (*Multilayer Perceptron*) para predição de resistência à insulina em adultos no Brasil. Apesar de existirem trabalhos que usam mineração de dados e aprendizagem de máquina na área de saúde, durante a pesquisa, não foi encontrado neles a aplicação de tais algoritmos com a finalidade aqui proposta, representando, assim, algo novo. Outra inovação consiste na utilização do algoritmo J48 para a definição de pontos de cortes nos atributos analisados para a predição do diagnóstico da resistência à insulina.

2 Resistência à Insulina

A resistência à insulina (RI) consiste em uma falha dos órgãos-alvos em responder normalmente à ação da insulina. Tal resistência causa supressão incompleta da produção de glicose hepática e captação da glicose mediada por insulina prejudicada nos tecidos periféricos alvos, como o músculo esquelético, o fígado e o tecido adiposo, causando menor efeito fisiológico desse hormônio, mais especificamente, menor captação da glicose [10].

A RI pode estar associada a outras condições, como obesidade central, hipertensão e dislipidemia, que são fatores de risco para desenvolvimento da síndrome metabólica. Adicionalmente, essa condição é apontada como responsável pelo desenvolvimento de outras alterações que a compõem, a exemplo da pressão arterial, da elevação da glicemia de jejum, do aumento dos triglicerídeos, da diminuição do colesterol HDL e do quadro de obesidade abdominal. Além do risco para desenvolvimento da síndrome metabólica, os pacientes com RI apresentam maior predisposição para diagnósticos de diabetes *mellitus* tipo II (DM2) e doenças cardiovasculares [12].

DOI: 10.5540/03.2026.012.01.0242

3 Metodologia

Este trabalho consiste em estudo com base em dados fornecidos pelo *Brazilian Metabolic Syndrome Study* (Brams). A pesquisa do Brams tem início em 2011 e avalia características antropométricas, clínicas, hormonais e metabólicas da síndrome de resistência à insulina na população adulta.

3.1 Descrição da Base de Dados

Este trabalho utilizou base de dados do Brams, obtida no Hospital das Clínicas da Universidade Estadual de Campinas (Unicamp) e em Unidades Básicas de Saúde (UBS). A pesquisa do Brams envolveu indivíduos de três estados do Brasil: São Paulo, Ceará e Minas Gerais. As amostras foram selecionadas utilizando uma amostragem não-probabilística intencional. No total, 5668 pacientes foram avaliados e, desses, 2607 foram compatíveis com os critérios desejados: idade entre 18 e 65 anos, índice de massa corporal (IMC) de 16 a 69, glicemia em jejum de 59 a 458 e insulina em jejum de 1 a 97.

Os critérios de exclusão foram evidência clínica ou laboratorial de doença cardíaca, renal, do fígado ou endócrina, doença sistêmica severa (por exemplo, câncer ou AIDS), consumo exagerado de álcool (cinco ou mais unidades a cada cinco ou mais dias nos últimos 30 dias). Também foram excluídos fisiculturistas, atletas, mulheres grávidas ou lactantes. Nenhum dos selecionados fazia uso de medicamentos que afetassem níveis de glicose plasmática ou sensibilidade à insulina, com exceção dos pacientes diabéticos. O grupo de pessoas diabéticas era composto por pacientes com diabetes leve bem controlada.

Todos os pacientes foram submetidos a um exame antropométrico detalhado, no qual foram coletadas informações sobre massa (apontada como peso na tabela), altura, circunferência da cintura, circunferência do quadril, circunferência do pescoço e diâmetro abdominal sagital, conforme explicado na pesquisa *Neck circumference as a simple tool for identifying the metabolic syndrome and insulin resistance* [13].

3.2 Etapas da Pesquisa

No pré-processamento, foi necessária uma redução na dimensionalidade, visando melhorar a eficácia e a eficiência do processo de mineração de dados. Essa etapa auxiliou para uma melhor compreensão dos dados disponíveis, reduziu o custo computacional, pois eliminou atributos que contribuíam menos, e, consequentemente, aumentou o desempenho do modelo.

Para o processo de modelagem, foram utilizados dois algoritmos disponíveis no WEKA: *Multilayer Perceptron* (Redes Neurais) e J48 (Árvores de Decisão) [14]. Os desempenhos com os dois algoritmos foram comparados, bem como sua utilização segundo os objetivos do trabalho e a clareza dos resultados para os profissionais da área de saúde.

Os dois algoritmos são utilizados como métodos de classificação. Essa escolha se deve à possibilidade de utilização de tais algoritmos em diversas áreas [5, 15, 16]. Além disso, o algoritmo J48 indica pontos de cortes nos atributos, o que é interessante para o estudo em questão.

O J48 consiste em um algoritmo de classificação e é uma implementação do antigo C4.5, versão 8, que foi proposto para o WEKA por Ross Quinlan, em 1993. Esse algoritmo gera uma árvore de decisão a partir do conjunto de dados de treinamento disponível. O modelo resultante é usado para classificar as instâncias no conjunto de teste.

Esse algoritmo cria uma árvore de decisão "de cima para baixo". O objetivo consiste em selecionar o melhor atributo para cada nó da árvore criada. Desse modo, cria-se um processo recursivo que escolhe o atributo para um nó (iniciando pela raiz) e, por meio de um ponto de corte, inicia

a classificação das instâncias, aplicando o mesmo processo (com os atributos escolhidos para cada nó) nos nós subsequentes, até que os critérios de parada sejam alcançados.

O *Multilayer Perceptron* consiste em um algoritmo de classificação com base em redes neurais artificias. Nele, utiliza-se técnica de retropropagação para a aprendizagem de um *perceptron* multicamadas para classificar as instâncias. Sua composição se dá por uma camada de entrada, uma ou mais camadas intermediárias (ocultas) e uma camada de saída.

Em uma camada específica, as saídas de seus neurônios se ligam apenas às entradas dos neurônios das camadas seguintes. Cada neurônio na rede apresenta uma função de ativação.

A validação do modelo se deu por validação cruzada (*cross-validation*). A base de dados foi dividida de forma aleatória em alguns *folds* (subconjuntos) mutuamente exclusivos e com aproximadamente a mesma quantidade de amostras. O número de *folds* escolhido foi 10. A cada iteração, 9 desses subconjuntos são usados para treinamento e o subconjunto restante é utilizado para teste, gerando uma métrica como resultado para avaliação. Nesse processo, cada subconjunto é usado para teste em algum momento da avaliação.

Essa técnica é bastante utilizada para avaliação de desempenho de modelos de aprendizagem de máquina. Seu uso ajuda a detectar se o modelo se encontra sobreajustado aos dados de treinamento, isto é, se aconteceu o *overfitting*. A validação cruzada pode ser computacionalmente intensiva. Contudo, dado o tamanho da base de dados da presente pesquisa, esse método se mostrou adequado.

4 Resultados e Discussão

Como um dos objetivos era analisar a influência das medidas antropométricas, dentre elas, a medida da circunferência do pescoço, aqueles que tinham dado faltante para esse atributo foram retirados. Por serem 375 (14% do total), sobraram 2232, o que ainda representou um número razoável para os testes.

Para o caso dos homens, os testes foram feitos com 734 pacientes. A árvore de decisão resultante do teste que manteve os atributos insulina e glicose trouxe os valores de corte obtidos. Para valores de insulina menores do que ou iguais a 11, há uma forte tendência ao paciente não apresentar resistência à insulina. Quando a glicose é menor do que ou igual a 114, o índice de acerto foi aproximadamente 99,53%. Nessa faixa, os pacientes só apresentavam tendência a ter a resistência à insulina quando a glicose era superior 114 e a insulina maior do que 8 ou quando a glicose era maior do que 186 e a insulina maior do que 4.

Por outro lado, quando a insulina era maior do que 11 entre os homens, havia tendência a existir a resistência à insulina. Os pacientes nessa faixa, combinado à glicose acima de 89, apresentaram todos resistência à insulina. A exceção estava na faixa de insulina menor do que ou igual a 14 e glicose menor do que ou igual a 86.

Alguns testes foram realizados com esses mesmos dados, dessa vez, utilizando redes neurais. O melhor resultado veio com o número de camadas ocultas (*hidden layers*) igual a 4. A Tabela 1 traz a comparação entre os resultados obtidos com os dois algoritmos.

Tabela 1: Resultados com os algoritmos J48 e *Multilayer Perceptron* (homens).

Algoritmo	Acurácia	Classificação correta SIM	Classificação correta NÃO	*Kappa*	Área sob a curva ROC
J48	97,56%	96,9%	97,9%	0,9467	0,968
Multilayer Perceptron	97,82%	97,3%	98,1%	0,9526	0,995

Ao considerar apenas a medida da circunferência do pescoço entre os homens, o ponto de corte

obtido foi 42 centímetros, no qual valores menores ou iguais a este indicavam a não existência da resistência à insulina. Nos testes com redes neurais, os resultados com 1 e 2 *hidden layers* foram muito parecidos. Ambos alcançaram 70,71% de acurácia aproximadamente. A Tabela 2 faz a comparação dos resultados com os dois algoritmos.

Tabela 2: Resultados com os algoritmos J48 e *Multilayer Perceptron* para testes com a circunferência do pescoço (homens).

Algoritmo	Acurácia	Classificação correta SIM	Classificação correta NÃO	*Kappa*	Área sob a curva ROC
J48	71,25%	41,2%	87,9%	0,3176	0,609
Multilayer Perceptron	70,71%	43,9%	85,6%	0,3161	0,691

Os resultados indicam que a medida da circunferência do pescoço pode ser usada muito mais para afastar a hipótese do diagnóstico de presença de resistência à insulina em homens adultos do que para confirmar a classe positiva. Além disso, o uso de árvores de decisão possibilitou encontrar um ponto de corte para a predição do diagnóstico: 42 centímetros.

Ao analisar as mulheres, após a retirada daquelas para as quais faltavam valores da medida da circunferência do pescoço, havia 1498 instâncias. O teste que incluiu insulina e glicose apresentou acurácia de 97,33% (aproximadamente) e estatística *kappa* igual a 0,9429. A árvore de decisão gerada por esse modelo trouxe um ponto de corte para a insulina igual a 13.

Nesse teste, valores de insulina maiores do que 13 associados a valores de glicose maiores do que 79 indicaram resistência à insulina em todas as pacientes. Por outro lado, valores de glicose menores do que ou iguais a 122 e de insulina menores do que ou iguais a 9 tiveram como resposta não resistência à insulina em todas as mulheres do banco de dados. Além disso, podem-se observar outros intervalos com alto índice de acerto (tanto para a existência da resistência à insulina quanto para a ausência desta).

O melhor resultado dos testes com o algoritmo *Multilayer Perceptron* ocorreu com 3 camadas ocultas. Na Tabela 3, pode-se observar a comparação entre os resultados obtidos com os dois algoritmos.

Tabela 3: Resultados com os algoritmos J48 e *Multilayer Perceptron* (mulheres).

Algoritmo	Acurácia	Classificação correta SIM	Classificação correta NÃO	*Kappa*	Área sob a curva ROC
J48	97,33%	95,9%	98,2%	0,9429	0,982
Multilayer Perceptron	97,13%	94,7%	98,6%	0,9394	0,997

Testando apenas a medida da circunferência do pescoço para o diagnóstico de resistência à insulina no banco de dados considerado, a árvore de decisão gerada trouxe como ponto de corte 39 cm. Acima desse valor, a tendência era para a indicação de haver resistência à insulina (com acerto de aproximadamente 68,92%). Abaixo desse valor, o modelo classificava como não havendo resistência à insulina, mas com maior índice de acerto para valores menores do que ou iguais a 36 cm, apresentando acurácia aproximada de 77,09%.

Os testes realizados com o algoritmo *Multilayer Perceptron* obtiveram 68,29% de acurácia, estatística *kappa* igual a 0,2952 e área sob a curva ROC igual a 0,714. O valor de *kappa* indicou que o modelo tem concordância fraca. Contudo, ao analisar as instâncias da classe negativa que

foram corretamente classificadas, tem-se 80,6% de acerto. Foram usadas 2 camadas intermediárias. A Tabela 4 traz os resultados dos testes com os dois algoritmos.

Tabela 4: Resultados com os algoritmos J48 e *Multilayer Perceptron para testes com a circunferência do pescoço (mulheres).*

Algoritmo	Acurácia	Classificação correta SIM	Classificação correta NÃO	*Kappa*	Área sob a curva ROC
J48	68,89%	30,8%	91,7%	0,2532	0,667
Multilayer Perceptron	68,291%	47,8%	80,6%	0,2952	0,714

Nas análises das medidas da circunferência do pescoço para o diagnóstico de resistência à insulina, os valores encontrados para pontos de corte para homens e mulheres foram os mesmos de estudo já realizado pelo BRAMS, quando o banco de dados estava menor (cerca de 1000 participantes). Para homens, o ponto de corte foi 42 cm e, para mulheres, 39 cm [13].

5 Considerações Finais

Este trabalho visou propor modelos preditivos para auxiliar profissionais de saúde quanto ao diagnóstico de resistência à insulina em adultos no Brasil. Os atributos de insulina e glicose apresentaram grande influência no resultado da resistência à insulina. Isso era esperado, visto que esses dois fatores são usados no cálculo do HOMA-IR. Ambos foram bem ranqueados pela hierarquia das árvores de decisão geradas pelos modelos. Valores de insulina maiores do que 13 e de glicose maiores do que 79 indicaram existência de resistência à insulina. Por outro lado, insulina menor do que ou igual a 9 e glicose menor do que ou igual a 121 indicaram ausência dessa resistência.

As medidas antropométricas também podem ser usadas para auxiliar no diagnóstico da resistência à insulina. Entre elas, a medida da circunferência do pescoço. Para os homens, o ponto de corte encontrado foi 42 cm (onde a resistência era mais comum em valores maiores do que este). Para mulheres, o ponto de corte encontrado foi 39 cm, porém, valores menores do que ou iguais a 36 cm aumentavam a precisão de não haver resistência à insulina.

Os resultados obtidos trazem informações importantes que podem auxiliar no diagnóstico de resistência à insulina em adultos no Brasil e trazer economicidade nesse processo. Mesmo nos modelos com acurácia menor, os pontos de corte indicaram bons resultados na predição da ausência de resistência à insulina. Isso pode afastar a hipótese de diagnóstico positivo e poupar os pacientes da realização de mais exames. Além disso, as medidas antropométricas podem ser obtidas de forma simples e com baixo custo.

Pesquisas futuras podem aplicar esses algoritmos na predição de diagnósticos de resistência à insulina em povos de diferentes países. Adicionalmente, esses métodos podem ser aplicados para auxiliar no diagnóstico de outras doenças. Além disso, os resultados são de entendimento acessível à população em geral e podem auxiliar na prevenção do desenvolvimento de fatores de risco (principalmente, no que diz respeito à glicose e insulina).

Agradecimentos

Agradecemos ao Instituto Federal Baiano (IF Baiano) e à Universidade Estadual de Campinas (UNICAMP) por possibilitar o desenvolvimento da pesquisa, em especial, ao IMECC, ao LIMED, ao OCRC-CEPID e ao BRAMS.

Referências

[1] E. Alpaydin. **Introduction to machine learning**. 3a. ed. Cambridge: MIT Press (MA), 2014. ISBN: 9780262028189.

[2] M. H. C. Carvalho, A. L. Colaço e Z. B. Fortes. "Citocinas, disfunção endotelial e resistência à insulina". Em: **Arquivos Brasileiros de Endocrinologia Metabologia** 50(2) (2006), pp. 304–312. DOI: 10.1590/S0004-27302006000200016.

[3] S. Chatterjee, K. Khunti e M. J. Davies. "Type 2 diabetes". Em: **Lancet** 389(10085) (2017), pp. 2239–2251. DOI: 10.1016/S0140-6736(17)30058-2.

[4] R. A. Defronzo e D. Tripathy. "Skeletal muscle insulin resistance is the primary defect in type 2 diabetes". Em: **Diabetes Cares** 32(2) (2009), pp. 157–163. DOI: 10.2337/dc09-S302.

[5] G. W. Dekker, M. Pechenizkiy e J. M. Vleeshouwers. "Predicting students drop out : a case study." Em: jan. de 2009, pp. 41–50. ISBN: 978-84-613-2308-1.

[6] J. P. Despres e I. Lemieux. "Abdominal obesity and metabolic syndrome". Em: **Nature** 444(7121) (2006), pp. 881–887. DOI: 10.1038/nature05488.

[7] J. Han, M. Kamber e J. Pei. **Data mining: concepts and techniques**. 3a. ed. Waltham: Morgan Kaufmann, 2012. ISBN: 9780123814791.

[8] K. Kodama, D. Tojjar, S. Yamada, K. Toda, C. J. Patel e A. J. Butte. "Ethnic differences in the relationship between insulin sensitivity and insulin response: a systematic review and meta-analysis". Em: **Diabetes Care** 36(6) (2013), pp. 1789–1796. DOI: 10.2337/dc12-1235.

[9] D. R. Mattews, J. P. Hoskers, A. S. Rudenski, B. A. Naylor, D. f. Treacher e R. C. Turner. "Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man". Em: **Diabetologia** 28(7) (1985), pp. 412–419. DOI: 10.1007/BF00280883.

[10] A. G. Pittas, N. A Joseph e A. S. Greenberg. "Adipocytokines and insulin resistance". Em: **J Clin Endocrinol Metab** 89(2) (2004), pp. 447–452. DOI: 10.1210/jc.2003-031005.

[11] G. Reaven. "All obese individuals are not created equal: insulin resistance is the major determinant of cardiovascular disease in overweight/ obese individuals". Em: **Diab Vasc Dis Res** 2(3) (2005), pp. 105–112. DOI: 10.3132/dvdr.2005.017.

[12] M. C. Romualdo, F. J. Nóbrega e M. A. Escrivão. "Insulin resistance in obese children and adolescents". Em: **J Pediatr (Rio J)** 90(6) (2014), pp. 600–607. DOI: 10.1016/j.jped.2014.03.005.

[13] C. Stabe, A. C. Vasques, M. M. Lima, M. A. Tambascia, J. C. Pareja, A. Yamanaka e B. Geloneze. "Neck circumference as a simple tool for identifying the metabolic syndrome and insulin resistance: results from the Brazilian Metabolic Syndrome Study". Em: **Clin Endocrinol (Oxf)** 78(6) (2013), pp. 874–81. DOI: 10.1111/j.1365-2265.2012.04487.x.

[14] WEKA. **The Data Platform for Cloud AI | WEKA**. Online. Acessado em 28/05/2025, https://www.weka.io.

[15] I. H. Witten, E. Frank e M. A. Hall. **Data mining: practical machine learning tools and techniques**. 3a. ed. Burlington: Morgan Kaufmann, 2011. ISBN: 9780123748560.

[16] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z. H. Zhou, M. Steinbach, D. J. Hand e D. Steinberg. "Top 10 algorithms in data mining". Em: **Knowledge and information systems** 14 (2008), pp. 1–37. DOI: 10.1007/s10115-007-0114-2.

