Diabetes detection in adult patients using explainable artificial intelligence

xxx
xxx
xxx, xxx
xxx

xxx
xxx
xxx, xxx

xxx
xxx
xxx, xxx

Abstract

Diabetes, a rapidly expanding chronic condition worldwide, requires diagnostic methods that are both accurate and efficient. This study presents an artificial intelligence-based approach, utilizing a multilayer perceptron neural network designed to identify diabetes based on multiple influential factors. The network was trained using a database, achieving an 82% precision and 83% accuracy. Additionally, the explainable artificial intelligence method SHAP was applied, revealing the significant contribution of parameters such as age and high blood pressure in the diagnostic process. These results highlight the importance of early detection and proper treatment of diabetes, reinforcing the potential of AI as a clinical support tool.

Keywords

Diabetes, Artificial Intelligence, Neural Networks, Machine Learning.

ACM Reference Format:
xxx, xxx, and xxx. 2018. Diabetes detection in adult patients using explainable artificial intelligence. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introdução

A diabetes é um dos principais fatores de morbidade e mortalidade, não apenas pelos efeitos imediatos da doença, mas também pelas complicações crônicas que pode provocar [6]. Entre essas complicações estão as doenças dos grandes vasos sanguíneos, como a doença coronariana e a doença arterial periférica, além de condições que afetam os pequenos vasos sanguíneos, como a nefropatia e a retinopatia [25]. Adicionalmente, a diabetes pode levar ao desenvolvimento de neuropatias, impactando negativamente a saúde geral do paciente [15]. Além disso, a diabetes é uma condição de saúde crônica que pode vim afetar a maneira em que seu corpo faz o processo de absorvição dos nutrientes alimentares em energia física no corpo, onde se tem definido três tipos correspondentes a diabetes, que são mostrados a seguir [10]:

(1) O diabetes do tipo 1: que é uma doença autoimune que faz com que seu próprio corpo reaga atacando as células do pâncreas que produzem o hormônio da insulina. Esse hormônio é quem ajuda o corpo a usar glicose para carregar a energia [4].
(2) O diabetes do tipo 2: é o tipo em que se é mais encontrado entre as pessoas que possuem essa condição, pois ela ocorre quando seu corpo não responde mais tão normalmente a insulina, ou quando o corpo não consegue mais produzir uma quantidade que seja suficiente de insulina [2, 4]. Para a falta de insulina é fundamental que exista aplicações diárias para que é portador da diabetes, num processo conhecido como insulinoterapia Isto é, passam a ter de administrar insulina todos os dias. Esta estratégia visa, sobretudo, controlar a glicemia, mas é também fundamental para evitar ao máximo as complicações provocadas pela doença [14, 21].
(3) Diabetes gestional: esse tipo de diabetes pode ocorrer durante a gravidez, porém é costumamento que deixe de existir após o nascimento do bebe[4].

A IA pode analisar grandes volumes de dados de saúde, incluindo informações demográficas, históricos médicos, exames laboratoriais e dados de monitoramento contínuo de glicose [9, 22]. Com algoritmos de aprendizado de máquina, é possível identificar padrões e correlações que poderiam passar despercebidos em análises tradicionais [7, 19]. Técnicas como redes neurais e máquinas de vetores de suporte têm sido aplicadas para classificar pacientes em diferentes categorias de risco de diabetes, como não-diabéticos, pré-diabéticos e diabéticos [13, 23, 31]. Esses modelos podem ser treinados com conjuntos de dados robustos para prever a probabilidade de desenvolvimento da doença, auxiliando os médicos na tomada de decisões informadas e na adoção de intervenções precoces [28].

A aplicação da rede neural com inteligência artificial explicável (XAI) tem como objetivo prever o desenvolvimento de pré-diabetes para ambos os tipos da doença (Tipo 1 e Tipo 2) [16]. Utilizando variáveis extraídas do banco de dados, o modelo é capaz de estimar a probabilidade de manifestação da condição, contribuindo para a tomada de decisões clínicas [1]. Este estudo propõe a utilização de uma rede neural combinada com o método SHAP (SHapley Additive exPlanations)

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

para identificar os parâmetros que exercem maior influência na predição de diabetes, conforme o modelo de rede neural implementado.

A explicabilidade da IA, por meio de métodos como SHAP, permite que os médicos compreendam melhor como os modelos chegaram a determinadas conclusões, aumentando a confiança nas recomendações feitas pela IA [18, 32]. Isso é fundamental em áreas sensíveis, como a saúde, onde decisões clínicas devem ser transparentes e baseadas em evidências [26].

2 Trabalhos relacionados

A busca por métodos eficazes para a classificação de condições de saúde, como diabetes, por meio de modelos de redes neurais, tem sido amplamente explorada em diversos estudos. Portanto, é relevante analisar os resultados alcançados em outras pesquisas e compará-los com os deste projeto[5, 11].

A Tabela 1 apresenta os principais modelos de aprendizado de máquina utilizados em três diferentes estudos, focados na classificação de condições de saúde relacionadas a detecção de diabetes. Em todos os trabalhos mencionados, a Floresta Aleatória foi o modelo que apresentou os melhores resultados em termos de acurácia.

Table 1: Modelos utilizados nos trabalhos e suas respectivas acurácias

Artigo	Modelos	Melhor Modelo	Acurácia
[30]	Floresta Aleatória	Floresta Aleatória	97%
[20]	Árvore de Decisão, ANN, Perceptron Multicamadas, Naive Bayes, Floresta Aleatória	Floresta Aleatória	96%
[24]	Naive Bayes, Máquina de Vetores de Suporte (SVM), Perceptron Multicamadas, Floresta Aleatória	Floresta Aleatória	NA

O artigo [30] apresentou uma aplicação de aprendizado de máquina utilizando um dos algoritmos mais comuns, a floresta aleatória. Esse algoritmo, que combina as saídas de várias árvores de decisão para gerar um único resultado, foi aplicado para classificar dados de saúde em categorias como não-diabéticos, pré-diabéticos e diabéticos. Após a realização de múltiplos testes cegos, a eficácia do algoritmo foi validada, alcançando uma precisão de classificação superior a 97% para a distinção entre diabéticos e não-diabéticos. No entanto, ao classificar pré-diabéticos, a precisão foi ligeiramente inferior, ficando em torno de 92%.

Com o mesmo objetivo de classificar os dados de saúde em diferentes categorias relacionadas ao diabete, o artigo [20] desenvolveu diversos algoritmos de aprendizado de máquina, utilizando o conjunto de dados mais recente disponibilizado pelo HANES. O diferencial desse trabalho foi a construção de múltiplos modelos de aprendizado de máquina, incluindo a árvore de decisão, que, devido à sua estrutura semelhante a um fluxograma, é fácil de visualizar e interpretar; o ANN, que classifica as amostras com base na proximidade com seus vizinhos mais próximos; o perceptron multicamadas, modelo também utilizado no presente artigo; o Naive Bayes, que se baseia nas descobertas de Thomas Bayes para fazer predições; e a floresta aleatória, o mesmo modelo utilizado no artigo anterior. A floresta aleatória apresentou o melhor resultado, com uma acurácia de aproximadamente 96%, enquanto o modelo Naive Bayes apresentou o menor desempenho, com 89% de acurácia.

Por fim, o artigo [24] também utilizou diferentes modelos de aprendizado de máquina, incluindo o Naive Bayes, a máquina de vetores de suporte (SVM), o perceptron multicamadas e a floresta aleatória. Todos os modelos foram avaliados com base em parâmetros de desempenho como sensibilidade, especificidade e acurácia balanceada. Utilizando o conjunto de dados NHANES.

Diferentemente dos trabalhos citados [30], [20] e [24], que se concentram principalmente na aplicação e comparação de diferentes algoritmos de aprendizado de máquina, como floresta aleatória, árvore de decisão, perceptron multicamadas, Naive Bayes e máquina de vetores de suporte, o presente trabalho foca na interpretação dos resultados gerados pelos modelos de IA. Enquanto os artigos destacam o desempenho em termos de métricas como acurácia e precisão, o diferencial do trabalho proposto está na explicabilidade dos modelos, utilizando uma rede neural perceptron multicamadas (MLP) para a classificação de dados de saúde relacionados ao diabetes, em conjunto com a técnica SHAP. Através do SHAP, é possível interpretar a contribuição individual de cada característica para a previsão final do modelo, promovendo maior transparência e confiabilidade nas decisões geradas pela IA [3, 27].

3 Metodologia

O BRFSS (*Behavioral Risk Factor Surveillance System*) [17] é uma pesquisa telefônica continua orientada a coleta de informações sobre possíveis comportamentos que podem causar riscos relacionados a saúde, condições crônicas e uso de serviços preventivos, onde esses dados sobre o estado de saúde são dirigidas ao público de adultos acima dos 18 anos. Os resultados das pesquisas são fielmente disponibilizadas todos os anos ao público para contribuição cientifica desde sua criação em 1984, nos Estados Unidos, realizado pelo CDC (*Centers for Disease Control and Prevention*). O BRFSS chega a realizar uma quantidade superior a 400.000 entrevistas com adultos a cada ano, atuando em todos os 50 estados e sendo a principal e às vezes a única fonte na maioria deles, o que torna o BRFSS o maior sistema de pesquisa de saúde conduzido continuamente em múltiplos modos por meio de correios telefones fixos e celulares em todo mundo[4].

O banco de dados empregado neste projeto é composto por 236.378 respostas à pesquisa BRFSS do CDC para o ano de 2021. A pesquisa utiliza valores ou respostas binárias (sim ou não), distribuídos em 22 parâmetros. Abaixo estão listados os nomes dos parâmetros e suas respectivas descrições:

(1) Diabetes: Utilizando 0 como valor falso ou neste caso definir como sem diabetes, e o valor 1 para risco de pré-diabetes e diabetes.
(2) HighBP (Pressão Alta): Que é referente à análise para pressão alta, para entrevistados que têm essa particularidade que pode interferir, sendo 0 para ausência de pressão alta e 1 para pressão alta confirmada.
(3) HighChol (Colesterol Alto): O HighChol é determinado para a taxa de colesterol, onde a definição binária que contém 0 é para sem colesterol alto, e o 1 para quem tem colesterol alto.
(4) CholCheck (Checagem de Colesterol): Neste caso seria uma verificação da taxa de colesterol nos últimos 5 anos, onde o binário 0 tem como resultado de checagem negativa para sem checagem de colesterol nos últimos 5 anos e 1 para quem realizou a checagem nos últimos 5 anos.
(5) BMI (Índice de Massa Corporal): É o índice de massa corporal que é para análise se a pessoa está com o peso ideal.
(6) Smoker (Fumante): Campo destinado a fumantes, ou que pelo menos tenham fumado um valor mínimo de 10 cigarros em toda a sua vida, que é equivalente a 5 maços. Os valores binários são 0 para não e 1 para sim.
(7) Stroke (Acidente Vascular Cerebral - AVC): Relacionado a quem teve AVC. 0 = não e 1 = sim.
(8) HeartDiseaseorAttack (Doença Cardíaca ou Infarto): Doença arterial coronariana (CHD) ou infarto do miocárdio (IM), 0 = não e 1 = sim.
(9) PhysActivity (Atividade Física): Atividade física praticada no intervalo dos últimos 30 dias, sendo que não é relevante incluir trabalho físico. 0 = não e 1 = sim.
(10) Fruits (Frutas): Este campo é para a verificação se houve ou se há o consumo diário de uma ou mais frutas durante o dia, 0 = não e 1 = sim.
(11) Veggies (Vegetais): Verifica se foi consumido um ou mais vegetais por dia, 0 = não e 1 = sim.
(12) HvyAlcoholConsump (Consumo Exagerado de Álcool): Corresponde a quem tem um alto consumo de álcool, homens adultos que tomam mais de 14 doses por semana e mulheres adultas que tomam mais de 7 doses por semana. 0 = não e 1 = sim.
(13) AnyHealthcare (Qualquer Tipo de Assistência Médica): Relacionado a pessoas que têm qualquer tipo de plano ou cobertura médica, incluindo seguro saúde, planos pré-pagos como o HMO, ou até mesmo planos de afinidade gratuitos. 0 = não e 1 = sim.
(14) NoDocbcCost (Não Consultou o Médico por Custo): Corresponde se houve em algum dos últimos 12 meses em que foi necessário realizar uma consulta médica, mas não a realizou devido ao custo. 0 = não e 1 = sim.
(15) GenHlth (Saúde Geral): Informa o nível de saúde geral segundo a pessoa, em uma escala de 1-5: 1 = excelente, 2 = muito bom, 3 = bom, 4 = razoável, 5 = ruim.
(16) MentHlth (Saúde Mental): Relacionado à saúde mental, incluindo fatores como estresse, depressão e problemas emocionais, indicando por quantos dias nos últimos 30 dias sua saúde mental não foi boa. A escala varia de 0 a 30.
(17) PhysHlth (Saúde Física): Relacionado à saúde física, incluindo doenças e lesões físicas, e indica por quantos dias nos últimos 30 dias a saúde física não foi boa. A escala varia de 0 a 30.
(18) DiffWalk (Dificuldade para Caminhar): Consiste na dificuldade da pessoa para caminhar ou subir escadas. 0 = não e 1 = sim.
(19) Sex (Sexo): Classifica o gênero entre homem e mulher, sendo 0 = feminino e 1 = masculino.
(20) Age (Idade): Categoria relacionada à idade e dividida em níveis, com a escala 1-13: 1 = 18-24 anos, 8 = 55-59 anos, 13 = 80 anos ou mais.
(21) Education (Educação): Categoria para o nível de educação, utilizando a escala 1-6: 1 = Nunca frequentou a escola ou apenas jardim de infância, 2 = 1ª a 8ª série.
(22) Income (Renda): Escala de renda, utilizando a escala 1-8: 1 = menos de 10.000, 5 = menos de 35.000, 11 = 200.000 ou mais.

Os dados se se dividem em sua maior parte na classe 0, em quase 7 vezes mais do que na classe 1, com isso é de necessidade então realizar um balanceamento mediante uma implementação de dados por intermédio da *Synthetic Minority Oversampling Technique* (SMOTE), se caso seja considerado a não utilização, o modelo pode ser consideravelmente afetado pelo paradoxo da acurácia, onde a classe menor, que neste caso seria a 1, não seria corretamente diferenciada das outras categorias, o que levaria a uma falsa impressão que o modelo está atingindo bons resultados, pois sua acurácia estaria apropriadamente alta [8].

4 Modelo de Rede Neural

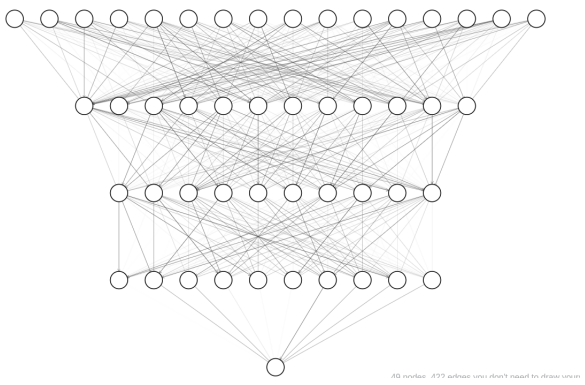

Figure 1: Modelo de rede neural desenvolvido.

O modelo computacional da rede neural perceptron multicamadas, mostrado na Figura 1, pode ser descrito matematicamente pela representação dos sinais de entrada como um vetor $\mathbf{x}$, composto por N elementos:

$$\mathbf{x} = [x_1, x_2, \ldots, x_N] \tag{1}$$

Cada sinal de entrada x_i é associado a um peso sináptico w_i, descrito pelo vetor de pesos:

$$\mathbf{w} = [w_1, w_2, \ldots, w_N] \tag{2}$$

No processo de processamento de sinais, os valores de entrada são multiplicados pelos respectivos pesos sinápticos, e, em seguida, é adicionado um termo conhecido como "bias" b, que confere maior flexibilidade ao modelo, permitindo que a função de ativação seja deslocada e não seja exclusivamente influenciada pelas entradas. Essa operação de combinação linear entre as entradas e os pesos sinápticos, acrescida do bias, pode ser expressa pela seguinte equação:

$$z = \sum_{i=1}^{N} w_i x_i + b \tag{3}$$

A variável z representa a soma ponderada das entradas, que posteriormente serve como entrada para a função de ativação, a qual determina a resposta final do neurônio artificial.

Para o desenvolvimento deste trabalho, o modelo mais adequado para classificar se a pessoa contém sinais de que tenha falta de insulina no sangue ou neste caso diabete é por meio de classificação binária. Para encontrar o melhor modelo a ser implementado, foi realizado um cálculo com o valor de amostras, sobre fator de escala multiplicado pelo número de neurônios de entrada mais a quantidade de neurônios de saída, o que resulta a quantidade de magnitude mais adequada para o modelo. A arquitetura do modelo desenvolvido para a classificação binária de diabetes utiliza uma rede neural densa com várias camadas totalmente conectadas (fully connected layers) implementadas com a biblioteca TensorFlow e Keras. O objetivo é classificar os dados em duas classes, representadas por uma saída binária (0 ou 1), utilizando a função de ativação sigmoide na camada final.

A arquitetura do modelo foi desenvolvida com 4 camadas, sendo 3 delas camadas ocultas e uma camada de saída, com o número de neurônios, 256 na primeira camada, 128 na segunda, 64 na terceira e 1 na quarta camada, em todas camadas ocultas utilizando a função de ativação ReLU (Rectified Linear Unit) dada por,

$$\mathrm{ReLU}(x) = \max(0, x). \tag{4}$$

Na camada de saída a função sigmoide foi utilizada. A função sigmoide é dada por,

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

A função de perda `binary crossentropy` dada por

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \tag{6}$$

foi utilizada no treinamento do modelo, em que N é o número total de amostras no conjunto de dados, y_i é o valor verdadeiro da classe para a amostra i, onde $y_i \in \{0, 1\}$ e $\hat{y}_i$ a probabilidade predita da classe 1 para a amostra i, onde $\hat{y}_i \in [0, 1]$. Portanto, a saída da rede pode ser dada por,

$$\hat{y} = \sigma\Big(W_4 \cdot \mathrm{ReLU}\Big(W_3 \cdot \mathrm{ReLU}\Big(W_2 \cdot \mathrm{ReLU}\big(W_1 \cdot x + b_1\big) + b_2\Big) + b_3\Big) + b_4\Big) \tag{7}$$

A equação de atualização dos pesos pode ser dada por,

$$W_{\text{novo}} = W_{\text{antigo}} - \eta \cdot \frac{\partial L}{\partial W} \tag{8}$$

em que W_{novo} representa o novo valor do peso após a atualização, W_{antigo} é o valor anterior do peso antes da atualização, η denota a taxa de aprendizado, que controla o tamanho do passo dado em direção à minimização da função de perda, e $\frac{\partial L}{\partial W}$ é o gradiente da função de perda L em relação ao peso W, que indica a direção e a magnitude em que o peso deve ser ajustado para reduzir a função de perda.

Para a definição da taxa de aprendizado, foi identificado que o otimizador Adam (*Adaptive Moment Estimation*) com um valor de 0.001 proporcionou resultados satisfatórios.

A fim de aprimorar o desempenho do modelo, foi implementada a técnica de *dropout*, que realiza atualizações nos pesos de forma a mitigar um dos principais problemas associados à implementação de redes neurais: o *overfitting*. A equação que define a operação do dropout pode ser expressa como

$$\mathbf{a}_{\text{dropout}} = \begin{cases} \frac{a}{p} & \text{com probabilidade } p \\ 0 & \text{com probabilidade } (1 - p) \end{cases} \tag{9}$$

em que $\mathbf{a}_{\text{dropout}}$ representa a saída após a aplicação do *dropout*, onde a é a ativação do neurônio antes da aplicação do *dropout*. A variável p denota a probabilidade de manter um neurônio ativo durante o treinamento, ou seja, a fração de neurônios que permanecem ativados. Com uma probabilidade p, a ativação é escalonada (normalizada) pela probabilidade p para manter a saída esperada, enquanto, com uma probabilidade $1 - p$, a ativação é definida como zero, efetivamente desativando o neurônio. Para converter qualquer número real resultante da soma das ativações dos neurônios em um valor dentro do intervalo [0, 1], foi aplicada a função de ativação sigmoide na camada oculta. Esta função retorna valores próximos a 0 para entradas pequenas e valores próximos a 1 para entradas grandes. A camada oculta está diretamente conectada à camada de saída, a qual é composta por dois neurônios [12].

Adicionalmente, diversas métricas foram empregadas para avaliar a evolução do modelo, incluindo acurácia, precisão, sensibilidade e F1-score. Os resultados dessas métricas serão discutidos em detalhes na seção de resultados.

4.1 IA explicável

Modelos de aprendizado de máquina podem ser complexos e difíceis de entender. Usar Inteligência Artificial Explicável (XAI) permite que médicos e pacientes compreendam como as previsões sobre o risco de diabetes são feitas. Além disso,

a explicabilidade permite identificar quais variáveis, como nível de glicose, IMC e histórico familiar, mais influenciam a classificação do risco de diabetes. Isso não apenas fornece informações valiosas para o tratamento, mas também pode orientar mudanças no estilo de vida.

O método de XAI SHAP é uma abordagem que busca oferecer interpretações compreensíveis sobre como modelos de aprendizado de máquina tomam decisões. Baseado na teoria dos jogos, o SHAP atribui a cada recurso (ou variável) uma contribuição para a previsão do modelo, permitindo que os usuários entendam como e por que uma determinada decisão foi tomada [29].

Matematicamente, a explicação de SHAP para uma previsão pode ser expressa da seguinte forma:

$$\phi_j(x) = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} \left(f(S \cup \{j\}) - f(S)\right) \tag{10}$$

em que $\phi_j(x)$ representa o valor de SHAP para a característica j. O conjunto N refere-se ao total de características disponíveis. A variável S denota um subconjunto de características que não inclui j. A função $f(S)$ representa a previsão do modelo utilizando apenas as características presentes no subconjunto S. O termo $f(S \cup \{j\}) - f(S)$ quantifica a contribuição marginal da característica j quando ela é adicionada ao subconjunto S.

Ao fornecer explicações para as previsões, SHAP ajuda a validar se o modelo está tomando decisões lógicas e coerentes com o conhecimento prévio sobre o problema.

5 Resultados

Para treinar o modelo, o dataset foi dividido em subconjuntos, com 80% dos dados alocados para o treino e os outros e 20% para teste.

O modelo de rede neural sequencial foi definido com várias camadas densas e *dropout* para regularização. A primeira camada tinha 256 neurônios com ativação ReLU, seguida por uma camada *dropout* com taxa de 25% para prevenir *overfitting*. A segunda camada possuía 128 neurônios também com ativação ReLU, seguida por outra camada *dropout* com taxa de 1%, e a terceira camada contava com 64 neurônios com ReLU. Na camada final de saída a função ativação sigmoide foi utilizada para gerar uma única previsão, adequada para um problema de classificação binária. O modelo foi compilado com a função de perda de entropia cruzada, o otimizador Adam com taxa de aprendizado de 0,0001. Durante o treinamento, utilizou-se a técnica de parada antecipada, que monitorou a perda de validação e restaurou os melhores pesos, interrompendo o treinamento quando o desempenho sobre os dados de validação deixou de melhorar após 80 épocas, garantindo a melhor performance do modelo sem desperdício de recursos computacionais. A Figura 2 mostra o gráfico da função perda.

As Figuras 3 e 4 mostram matrizes de confusão para os dados de treino e teste de classificação para detecção de diabetes. A matriz de confusão dos dados de teste mostra uma performance menor em comparação ao treino, com uma

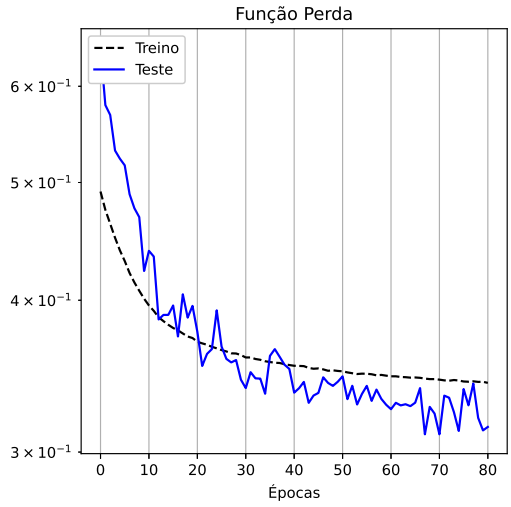

Figure 2: Função perda durante as épocas.

quantidade maior de falsos negativos e falsos positivos. O número de falsos negativos no conjunto de teste ainda é significativo, o que pode ser problemático, especialmente em um cenário clínico onde a não detecção de diabetes pode ter consequências graves.

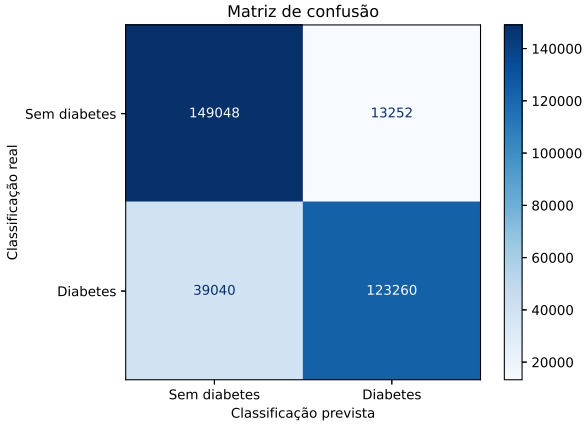

Figure 3: Matriz de confusão dos dados de treino.

A Tabela2 mostra as métricas de desempenho do modelo. Ele consegue equilibrar bem a identificação correta dos casos positivos (diabetes) e evitar falsos positivos, o que é essencial em contextos médicos, onde ambos os erros (falsos positivos e falsos negativos) podem ter consequências importantes.

A Figura 5 mostra o impacto das variáveis de entrada no diagnóstico de diabetes no modelo de rede neural proposto. As variáveis relacionadas à saúde geral, idade, IMC, colesterol e pressão arterial são as que mais contribuem para a previsão de diabetes no modelo, com um impacto significativo de estilos de vida saudáveis (atividade física, consumo de frutas) e fatores socioeconômicos.

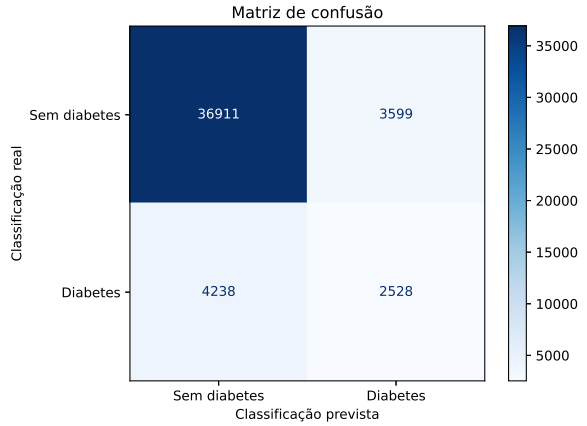

Figure 4: Matriz de confusão dos dados de teste.

Table 2: Resultados das métricas de desempenho do modelo

Métrica	Valor
Acurácia	0.8342
Precisão	0.8277
Sensibilidade	0.8342
F1-Score	0.8308

Já a Figura 6 mostra como os valores de cada variável impacta na previsão de diabetes. As cores indicam os valores das variáveis: azul representa valores baixos da variável e vermelho representa valores altos. Neste caso, saúde geral, idade e IMC (Índice de Massa Corporal), colesterol alto, pressão alta são as variáveis que mais impactam o modelo. Valores mais altos (em vermelho) dessas variáveis estão fortemente associados a um aumento na probabilidade de diagnóstico de diabetes, enquanto valores mais baixos (azul) têm impacto negativo, ou seja, diminuem a probabilidade de diagnóstico. O gráfico de SHAP torna claro que o modelo identifica corretamente fatores de risco conhecidos para diabetes, e suas previsões são fortemente influenciadas por esses fatores.

A Figura 7 mostra a dependência da idade com a saúde geral dos indivíduos consultados.

A Figura 8 mostra uma explicação local usando valores SHAP para um caso específico no diagnóstico de diabetes. As variáveis em vermelho estão puxando o valor de saída para mais próximo de um diagnóstico de diabetes. Já as variáveis em azul estão puxando o valor de saída para mais longe de um diagnóstico de diabetes. As variáveis mais importantes que aumentaram a chance de diagnóstico foram a pressão alta, o colesterol elevado e a idade avançada, enquanto fatores como um IMC mais baixo, uma boa saúde geral e o sexo contribuíram para reduzir a probabilidade de diabetes. O gráfico mostra como o modelo balanceia essas influências individuais para chegar a uma previsão final.

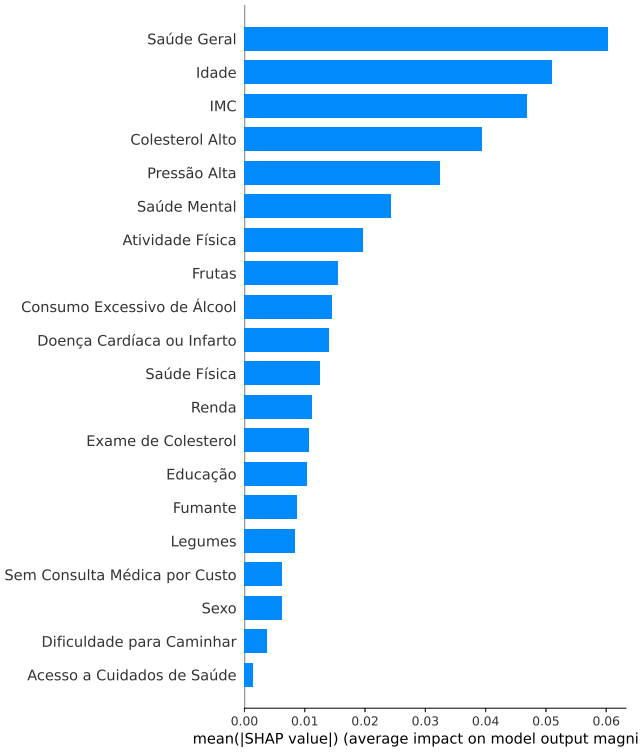

Figure 5: Variáveis que mais impactam no diagnóstico de positivo ou negativo de diabetes.

6 Conclusão

Utilizando uma rede MLP e explicabilidade do modelo, este trabalho permitiu uma compreensão mais profunda dos fatores que mais impactam as previsões de diabetes do modelo MLP, como saúde geral, idade, IMC e pressão alta. As métricas obtidas, como acurácia de 83,42%, precisão de 82,77%, sensibilidade de 83,42% e F1-Score de 83,08%, indicam que o modelo apresenta um desempenho equilibrado, sendo capaz de identificar corretamente a maioria dos casos de diabetes. Portanto, o uso de redes neurais, combinado com técnicas de explicabilidade, pode melhorar o diagnóstico precoce de diabetes, auxiliando profissionais de saúde na identificação de pacientes em risco. Futuras melhorias podem ser exploradas através de ajustes adicionais no modelo, maior balanceamento de dados e a inclusão de novos fatores de risco para tornar o sistema ainda mais preciso e confiável.

References

[1] Vivekanand Aelgani, Suneet K Gupta, and VA Narayana. 2023. Local agnostic interpretable model for diabetes prediction with explanations using xai. In *Proceedings of Fourth International Conference on Computer and Communication Technologies: IC3T 2022*. Springer, 417–425.

[2] Ahmed F Ashour, Mostafa M Fouda, Zubair Md Fadlullah, and Mohamed I Ibrahem. 2024. Optimized Neural Networks for Diabetes Classification Using Pima Indians Diabetes Database. In *2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI)*. IEEE, 1–7.

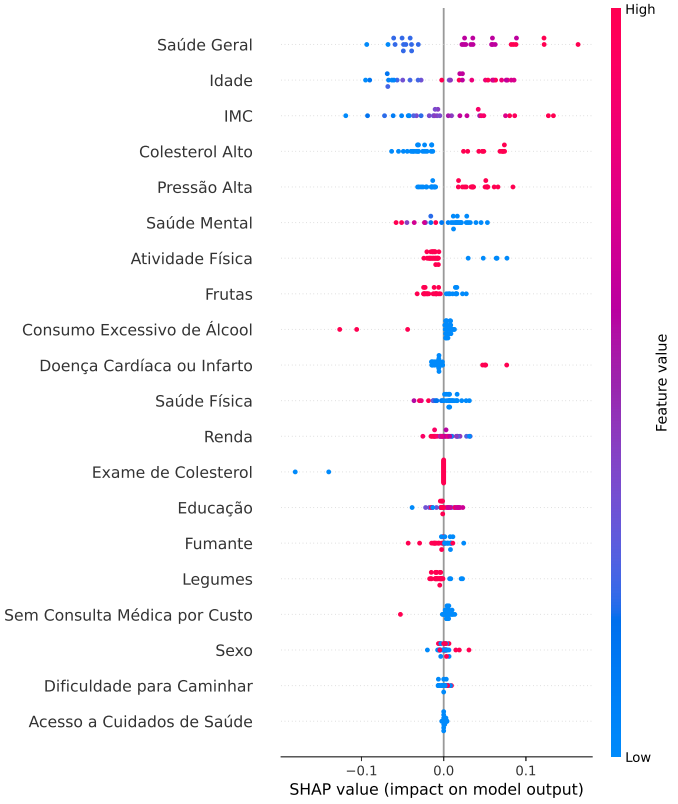

Figure 6: Variáveis que mais impactam no diagnóstico de diabetes.

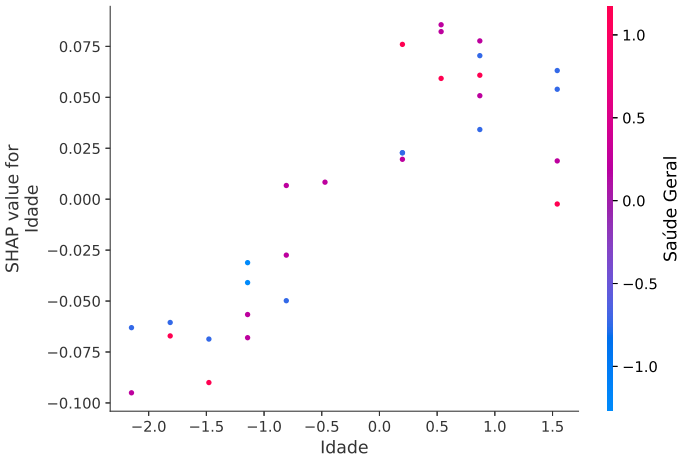

Figure 7: Relação de dependência entre a idade e a saúde geral dos indivíduos consultados.

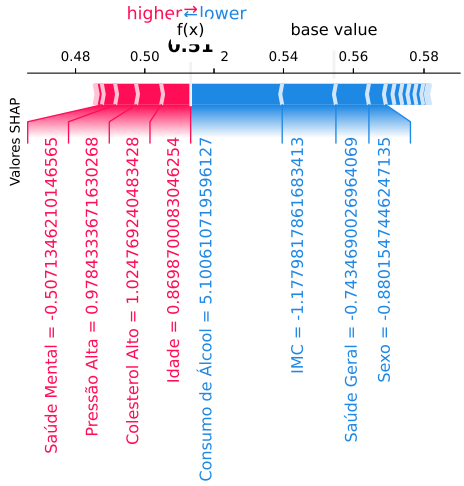

Figure 8: Explicação local para diagnóstico de diabetes.

[3] Caio MV Cavalcante and Rosana CB Rego. 2024. Early prediction of hypothyroidism based on feature selection and explainable artificial intelligence. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*. SBC, 49–60.

[4] CDC - Centers for Disease Control and Prevention. [n. d.]. Disponível em: https://www.cdc.gov/brfss/. Acessado em: [data de acesso].

[5] Tin-Chih Toly Chen, Hsin-Chieh Wu, and Min-Chi Chiu. 2024. A deep neural network with modified random forest incremental interpretation approach for diagnosing diabetes in smart healthcare. *Applied Soft Computing* 152 (2024), 111183.

[6] Chun-Yang Chou, Ding-Yang Hsu, and Chun-Hung Chou. 2023. Predicting the onset of diabetes with machine learning methods. *Journal of Personalized Medicine* 13, 3 (2023), 406.

[7] Ambika Choudhury and Deepak Gupta. 2019. A survey on medical diagnosis of diabetes using machine learning techniques. In *Recent Developments in Machine Learning and Data Analytics: IC3 2018*. Springer, 67–78.

[8] Martin Cooper. 2022. Turing Talk 2022. *ITNOW* 64, 2 (2022), 35.

[9] Francesco Curia. 2023. Explainable and transparency machine learning approach to predict diabetes develop. *Health and Technology* 13, 5 (2023), 769–780.

[10] Ofelia Cizela da Costa Tavares and Abdullah Zainal Abidin. 2024. ARTIFICIAL NEURAL NETWORK MULTI-LAYER PERCEPTRON FOR DIAGNOSIS OF DIABETES MELLITUS. *JIKO (Jurnal Informatika dan Komputer)* 7, 1 (2024), 19–23.

[11] Arianna Dagliati, Simone Marini, Lucia Sacchi, Giulia Cogni, Marsida Teliti, Valentina Tibollo, Pasquale De Cata, Luca Chiovato, and Riccardo Bellazzi. 2018. Machine learning methods to predict diabetes complications. *Journal of diabetes science and technology* 12, 2 (2018), 295–302.

[12] Data Science Academy. 2022. *Deep Learning Book*.

[13] Henock M Deberneh and Intaek Kim. 2021. Prediction of type 2 diabetes based on machine learning algorithm. *International journal of environmental research and public health* 18, 6 (2021), 3317.

[14] Diabetes 365. [n. d.]. Insulinoterapia: quando e como. Disponível em: https://www.diabetes365.pt/cuidar/insulinoterapia-quando-como-e-onde/. Acessado em: [data de acesso].

[15] Encyclopedia Britannica. [n. d.]. Diabetes mellitus desordem médica. Acessado em: [data de acesso].

[16] Muhammad Exell Febrian, Fransiskus Xaverius Ferdinan, Gustian Paul Sendani, Kristien Margi Suryanigrum, and Rezki Yunanda. 2023. Diabetes prediction using supervised machine learning. *Procedia Computer Science* 216 (2023), 21–30.

[17] Centers for Disease Control, Prevention, et al. 2015. BRFSS: Behavioral Risk Factor Surveillance System. *Prevalence and trends data, Michigan 2000 health status* (2015).

[18] Rita Ganguly and Dharmpal Singh. 2023. Explainable artificial intelligence (xai) for the prediction of diabetes management: An ensemble approach. *International Journal of Advanced Computer*

Science and Applications 14, 7 (2023).

[19] Yazan Jian, Michel Pasquier, Assim Sagahyroon, and Fadi Aloul. 2021. A machine learning approach to predicting diabetes complications. In *Healthcare*, Vol. 9. MDPI, 1712.

[20] R. Johnson and M. Lee. 2022. Predicative Diabete Using Supervised Machine Learning Techniques. *International Journal of Data Science* 18, 2 (2022), 89–102. https://www.example.com/predicative-diabete

[21] Ioannis Kavakiotis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, and Ioanna Chouvarda. 2017. Machine learning and data mining methods in diabetes research. *Computational and structural biotechnology journal* 15 (2017), 104–116.

[22] Jobeda Jamal Khanam and Simon Y Foo. 2021. A comparison of machine learning algorithms for diabetes prediction. *Ict Express* 7, 4 (2021), 432–439.

[23] Md Maniruzzaman, Md Jahanur Rahman, Benojir Ahammed, and Md Menhazul Abedin. 2020. Classification and prediction of diabetes disease using machine learning paradigm. *Health information science and systems* 8 (2020), 1–14.

[24] P. Martinez and L. Zhang. 2023. Rondon Forest Algorithm for the Classification of Health Data. *Computational Health Informatics* 9, 4 (2023), 130–145. https://www.example.com/rondon-forest-algorithm

[25] Aishwarya Mujumdar and Vb Vaidehi. 2019. Diabetes prediction using machine learning algorithms. *Procedia Computer Science* 165 (2019), 292–299.

[26] Nasim Mahmud Nayan, Ashraful Islam, Muhammad Usama Islam, Eshtiak Ahmed, Mohammad Mobarak Hossain, and Md Zahangir Alam. 2023. SMOTE Oversampling and Near Miss Undersampling Based Diabetes Diagnosis from Imbalanced Dataset with XAI Visualization. In *2023 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, 1–6.

[27] Yasunobu Nohara, Koutarou Matsumoto, Hidehisa Soejima, and Naoki Nakashima. 2022. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Computer Methods and Programs in Biomedicine* 214 (2022), 106584.

[28] Muhammad Azeem Sarwar, Nasir Kamal, Wajeeha Hamid, and Munam Ali Shah. 2018. Prediction of diabetes using machine learning algorithms in healthcare. In *2018 24th international conference on automation and computing (ICAC)*. IEEE, 1–6.

[29] M Scott, Lee Su-In, et al. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017), 4765–4774.

[30] J. Smith and A. Doe. 2021. Diabete Health Indicators Dataset. *Journal of Machine Learning Applications* 14, 3 (2021), 45–60. https://www.example.com/diabete-health-indicators-dataset

[31] Priyanka Sonar and K JayaMalini. 2019. Diabetes prediction using different machine learning approaches. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 367–371.

[32] Varad Vishwarupe, Prachi M Joshi, Nicole Mathias, Shrey Maheshwari, Shweta Mhaisalkar, and Vishal Pawar. 2022. Explainable AI and interpretable machine learning: A case study in perspective. *Procedia Computer Science* 204 (2022), 869–876.