Artigo Original

v. 6 n. 4 (2025)
ISSN: 2675-8008
Revista Multidisciplinar em Saúde

ACESSO ABERTO

Data de Recebimento:
03/10/2025

Data de Aceite:
21/10/2025

Data de Publicação:
22/10/2025

****Autor correspondente:***
Sílvio Fernando Alves Xavier Júnior, Doutor Biometria, Departamento de Estatística, UEPB, Rua das Baraúnas, S/N, Campina Grande – Paraiba, Brasil.
Telefone: 55 (83) 99616-4333
E-mail: silvio@servidor.uepb.edu.br

Citação:
SANTOS, J.V; *et al*, A predictive analysis system for type 2 diabetes using deep learning. **Revista Multidisciplinar em Saúde**, v. 6, n. 4, 2025. https://doi org/10.51161/integrar/rems/4695

DOI: 10.51161/integrar/rems/4695

A PREDICTIVE ANALYSIS SYSTEM FOR TYPE 2 DIABETES USING DEEP LEARNING

Jefferson Vieira dos Santos[a], Igor Barbosa Negreiros[b], Érika Fialho[b], Tiago Almeida de Oliveira[b], Sílvio Fernando Alves Xavier Júnior[b*], Jader da Silva[a], Vladimir Costa de Alencar[c]

[a] Departamento de Estatística, Universidade Federal Rural de Pernambuco (UFRPE). Rua Dom Manoel de Medeiros, s/n – Dois Irmãos, Recife– Pernambuco, Brasil.
[b] Departamento de Estatística, Universidade Estadual da Paraíba (UEPB). Rua das Baraúnas s/n, Campina Grande – Paraíba, Brasil.
[c] Departamento de Computação, Universidade Estadual da Paraíba (UEPB). Rua das Baraúnas s/n, Campina Grande – Paraíba, Brasil.

RESUMO

Introdução: A Organização Mundial da Saúde publicou seu primeiro relatório global sobre diabetes, concluindo que, de 1980 até os dias atuais, o número de diabéticos no mundo quadruplicou em uma única geração. O diabetes, descoberto há milhares de anos, tornou-se uma doença crônica do século XXI, aumentando o risco de acidente vascular cerebral, insuficiência renal, doença vascular periférica, doenças cardíacas e morte. **Objetivo:** O objetivo deste estudo foi desenvolver um sistema preditivo capaz de classificar o risco de diabetes tipo 2 por meio de técnicas de Deep Learning, contribuindo para a detecção precoce e prevenção da doença. **Metodologia:** Este estudo analisou dados públicos de diabetes coletados pelos *Centers for Disease Control and Prevention* (CDC, EUA) e desenvolveu um modelo preditivo para classificar o risco de diabetes tipo 2 utilizando *Deep Learning.* O modelo de classificação foi implementado utilizando a arquitetura de Rede *Perceptron Multicamadas* (MLP). **Resultados:** O modelo proposto alcançou uma acurácia geral de 73,8%. Ao analisar a Matriz de Confusão, a acurácia obtida foi de 68,78% para a classe Normal e 78,72% para a classe Diabetes. O modelo desenvolvido foi disponibilizado por meio de uma aplicação web, permitindo que os usuários insiram dados como idade, sexo, índice de massa corporal (IMC) e hábitos de tabagismo. **Conclusões:** O sistema classifica o risco de diabetes tipo 2 e fornece a probabilidade de desenvolvimento da doença, representando uma ferramenta útil para estratégias de prevenção em saúde pública.

Palavras-chave: Deep Learning; Machine Learning; predição de risco; Diabetes Mellitus, Tipo 2

ABSTRACT

Introduction: The World Health Organization has published its first global report on diabetes, concluding that from 1980 to the present day, the number of diabetics worldwide has quadrupled in a single generation. Diabetes, discovered thousands of years ago, has become a chronic disease of the 21st century, increasing the risk of stroke, kidney failure, peripheral vascular disease, heart disease, and death. **Objective:** The objective of this study was to develop a predictive system capable of classifying the risk of type 2 diabetes using Deep Learning techniques, contributing to early detection and disease prevention. **Methodology:** This study analyzed public diabetes data collected by the Centers for Disease Control and Prevention (CDC, USA) and developed a predictive model to classify the risk of type 2 diabetes using Deep Learning. The classification model was implemented using the Multilayer Perceptron Network (MLP) architecture. **Results:** The proposed model achieved an overall accuracy of 73.8%. When analyzing the Confusion Matrix, the accuracy obtained was 68.78% for the Normal class and 78.72% for the Diabetes class. **Conclusions:** The developed model was made available via web application, allowing users to enter data such as age, sex, body mass index (BMI), and smoking habits. The system then classifies the risk of type 2 diabetes and provides the probability of developing the disease, representing a useful tool for public health prevention strategies.

Keywords: Deep Learning; Machine Learning; Risk Prediction; Diabetes Mellitus, Type 2.

INTRODUCTION

Diabetes is a chronic disease that increases the risk of stroke, kidney failure, peripheral vascular disease, heart disease, and death. People with diabetes lose the ability to effectively regulate their blood glucose levels, which can lead to a decreased quality of life and life expectancy. After different foods are broken down into sugars during digestion, these are released into the bloodstream. In this way, it causes the pancreas to release insulin. Insulin helps the body's cells use the sugars in the bloodstream for energy. Diabetes is generally characterized by the body not producing enough insulin or not being able to use the insulin that is produced effectively enough (American Diabetes Association, 2018; XIE *et al*., 2019).

Complications such as heart disease, vision loss, lower limb amputation, and kidney disease are associated with recurrent high blood sugar levels in people with diabetes. Strategies such as losing weight, eating a healthy diet, being active, and receiving medical treatments can mitigate the damage of this disease in many patients. Early diagnosis can lead to behavior changes and more efficient treatment, making diabetes risk prediction models an important tool for healthcare professionals (American Diabetes Association, 2018; XIE *et al*., 2019).

As stated by Cho *et al*. (2018), The International Diabetes Federation projects that 693 million people will have diabetes worldwide, at the current growth rate, by 2045. Studies by The Centers for Disease Control and Prevention (CDC) pointed out that 29.1 million people in the United States of America were diagnosed with diabetes in 2012. This result makes diabetes the seventh most prominent cause of death in the country. Diabetes represents a high financial burden on the global economy. Studies show that the estimated total cost of diagnosed diabetes increased to $327 million in 2017, including $237 million in direct medical costs and $90 million in reduced productivity in the United States (American Diabetes Association, 2018; XIE *et al*., 2019).

In Brazil, recent data indicate that approximately 10.2% of adults report having been diagnosed with diabetes, according to the 2023 Vigitel survey conducted by the Brazilian Ministry of Health. The disease affects both sexes and all age groups, but its prevalence increases substantially after 45 years of age.

Epidemiological studies also show a prevalence of around 9.2% for type 2 diabetes, with marked regional variation — from 6.3% in the North to 12.8% in the Southeast (Santos, Lima & Corrêa, 2024). Together, these findings highlight the increasing burden of diabetes in Brazil and reinforce the need for predictive and preventive approaches capable of supporting early detection and healthcare planning.

However, despite these advances in surveillance and awareness, a significant proportion of type 2 diabetes cases remain undetected or are identified only when complications have already developed. This delayed diagnosis highlights the need for innovative approaches capable of identifying individuals at early stages of risk. The central research problem addressed in this study is whether computational models based on Deep Learning can accurately predict the likelihood of type 2 diabetes using large-scale population data.

There are three main varieties of diabetes: type 1, type 2, and the gestational one. Of these 3, type 2 diabetes is the most frequent and accounts for 90% to 95% of all cases. Type 2 diabetes is a predictable and avoidable disease since it usually develops later in life (over 30 years of age), resulting in the association of risk factors related to lifestyle (low physical activity, obesity) and others (age, sex, race, family history) (NOBLE *et al*., 2011).

The justification for this research lies in the growing importance of artificial intelligence as a decision-support tool in healthcare. Predictive systems based on Deep Learning can help identify at-risk individuals earlier, guiding preventive strategies and optimizing public health resources. Moreover, the integration of epidemiological data with computational models represents an innovative contribution to the scientific community and society, enabling scalable, data-driven health interventions.

Specifically, the objective of this study was to develop a predictive system capable of classifying the risk of type 2 diabetes using Deep Learning techniques, contributing to early detection and disease prevention.

MATERIAL AND METHODS

2.1. Data Collection and Pre-processing

When collecting the database, data cleaning was performed (handling and filling in missing data, reducing noise, identifying and removing outliers, and resolving inconsistencies), and noisy data was treated (meaningless data that cannot be interpreted by a machine, which may have been generated due to collection errors).

2.2. Building the model

2.2.1. Neural Networks

The human brain is a naturally occurring model for constructing intelligent machines. Everyday tasks, such as walking, picking up an object, or recognizing a person, involve the action of several components, such as memory, Learning, and motor coordination. The performance of these tasks is due to our complex biological structure, and the human brain is mainly responsible for processing information and generating responses (HAYKIN, 2000; RASCHKA & MIRJALILI, 2017).

Based on these motivations, Artificial Neural Networks (ANNs) were inspired by the structure and functioning of the nervous system. ANNs aim to simulate the human brain's learning capacity in acquiring knowledge. With the advent of distributed computing and parallel processing, neural networks have been gaining prominence through Deep Learning—deep Neural Networks with thousands of internal layers (HAYKIN, 2000; FACELI *et al*., 2011).

2.2.2. Artificial Neural Networks (ANNs)

ANNs are distributed computing systems composed of simple, densely interconnected computational units. These units are known as artificial neurons, which compute mathematical functions. The units are arranged in one or more layers and interconnected by many connections. These connections simulate biological synapses, which have associated weights that weigh the input received by each neuron in the network. The weights have their values adjusted by a learning process that encodes the acquired knowledge (HAYDIN, 2000).

In the architecture of artificial neural networks (Figure 1), the neuron is the fundamental processing unit of an ANN. Each input terminal of the neuron receives a value. The received values are weighted and combined by a mathematical function f(a).

The input values are multiplied by their synaptic weights and added together. The total value of the sum is then presented to an activation function, which will return the neuron's response (HAYKIN, 2000).

The input signals to the neuron are described by the vector x=[x_1,x_2,x_3,…,x_N] and may correspond to the pixels of an image, risk factors for type 2 diabetes. When they reach the neuron, they are multiplied by their respective synaptic weights, which are the elements of the vector w=[w_1,w_2,w_3,…,w_N], generating the value z, commonly called activation potential (HAYKIN, 2000), according to the expression:

$$z = \sum_{i=1}^{N} x_i w_i + b \qquad (1)$$

The additional term b provides an additional degree of freedom, which is not affected by the input to this expression, typically corresponding to the "bias." The value z then passes through a mathematical activation function σ, with the characteristic of being nonlinear, responsible for limiting this value to a specific interval, producing the final output value y of the neuron. Some activation functions σ(z) used are the step, sigmoid, hyperbolic tangent, softmax, and ReLU (Rectified Linear Unit) (HAYKIN, 2000; MOHRI, 2016).

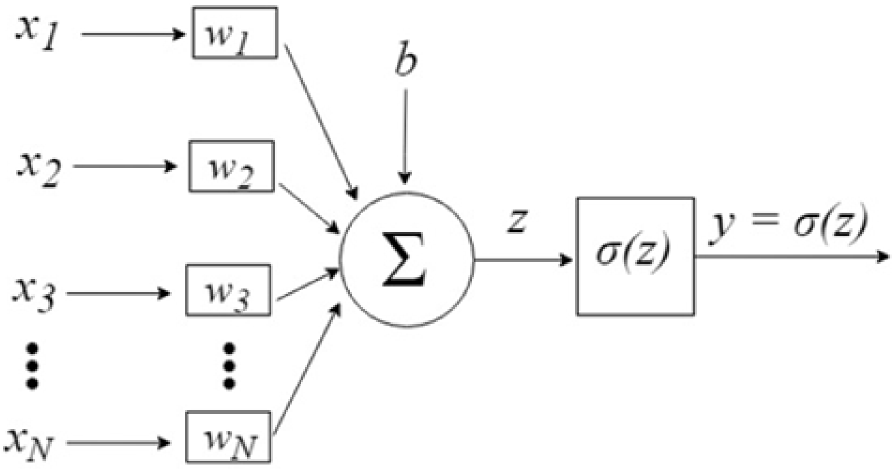

Figure 1: The artificial neuron. Source: Adapted from MOHRI, 2016.

2.2.3. Backpropagation

The backpropagation algorithm is based on calculating the error that occurs in the output layer of a neural network (NN). This error is used to prepare the values of the weights in the vector w of the last layer of neurons. The algorithm then works backward through the network, updating all weights in the layers from the last one until meeting the input layer. This process involves backpropagating the error throughout the network to optimize weight values. In other words, the error between what the network thought it was and what it was (it was diabetes, and it thought it was not - we have an error there), then we recalculate the value of all the weights, starting from the last layer and going to the first, always intending to reduce this error (HAYKIN, 2000; MOHRI *et al*., 2016).

2.2.4. Deep Learning

The deep learning technique is a subset of Machine Learning that applies algorithms to process data and mimic the processing done by the human brain. It uses layers of mathematical neurons to process data, understand human speech, and recognize objects visually. Information is passed through each layer, with the previous layer's output providing input to the next layer. The first layer is denominated as the input layer, while the last is named the output layer. All layers in between are attributed to be hidden layers. Each layer is commonly a simple, uniform algorithm containing a type of activation function (GOODFELLOW *et al*., 2017; Raschka & Mirjalili, 2017).

Deep Learning (DL) architectures are usually complex and require many data for training. Therefore, it is inevitable that they will rely on much computational power to apply these techniques. Although some classical methods use much computational power, such as memory or CPU, DL techniques are on another scale. If it were not for research related to parallel computing, the use of GPUs with CUDA, and the existence of Big Data, Deep Learning would not exist because it is unfeasible to use only the CPU. We see in Figure 2 that intensive (parallelizable) computing uses the GPU (graphics processing unit), and the CPU executes the sequential part of the code (GOODFELLOW *et al*., 2017; DETTMERS, 2015).

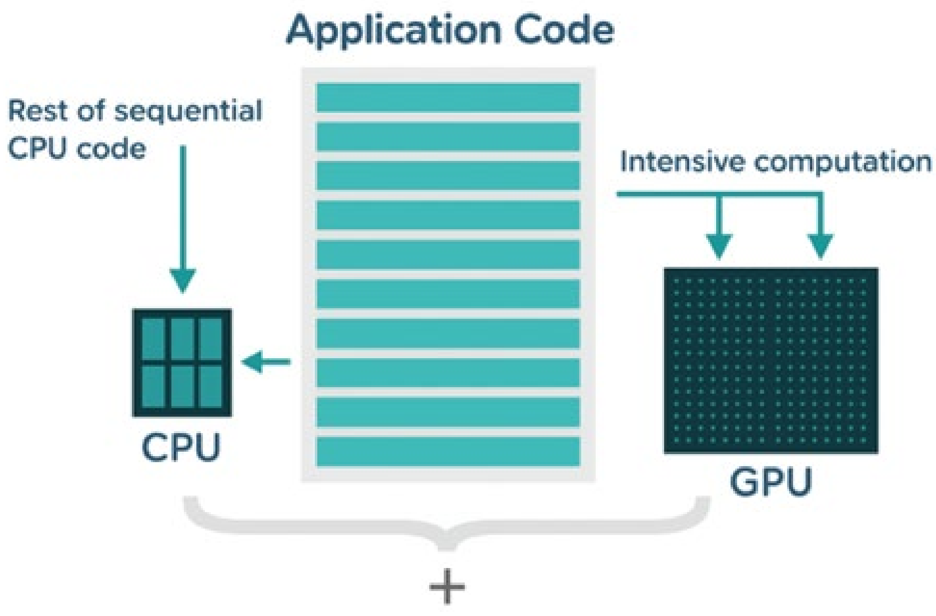

Figure 2: CPU usage vs. GPU usage - Source: Adapted from MENDES *et al*., 2023

The great differentiator of Deep Learning architectures is the flexibility of interconnecting multiple neurons in more complex structures to solve problems. Deep Learning can be an MLP (Multilayer Perceptron Neural Network) with tens or even thousands of layers; it can also be a CNN (Convolutional Neural Network - specialized in image processing) or an LSTM (Long Short-Term Memory - specialized in speech processing and time series), among others. Of the various possible Deep Learning architectures, we will use the MLP architecture with multiple layers in our project (GOODFELLOW *et al*., 2017).

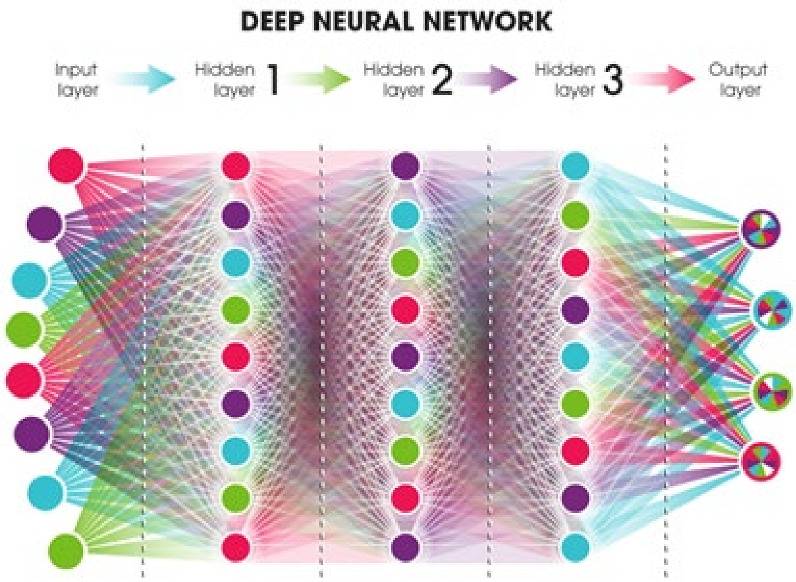

Figure 3: Multilayer Perceptron (MLP) Network –
Source: neuralnetworksanddeeplearning.com

Most MLP neural networks are organized into groups of units called layers. This architecture is organized with the layers in a chain structure (or in series), each receiving a function from the previous layer. In this structure, the first layer is given by (Raschka & Mirjalili, 2017):

$$h^{(1)} = g^{(1)}(W^{(1)T} + b^{(1)}) \qquad (2),$$

In which g is the activation function, Ws are the synaptic weights, b is the bias; x is the input data, h(1) is the 1st layer.

The second layer is given by:

$$h^{(2)} = g^{(2)}(W^{(2)T} + b^{(2)}) \qquad (3).$$

RESULTS E DISCUSSION

3.1 Dataset

The data were collected from the Behavioral Risk Factor Surveillance System (BRFSS), a public database available by the CDC (Centers for Disease Control and Prevention) from 415,185 individuals containing 303 variables, including blood pressure, cholesterol, smoking, obesity, age, sex, race, diet, physical exercise, alcohol consumption, family income, marital status, sleep quality, time since last check-up, academic background, mental health (CDC, 2021). Of these records, 54,635 are individuals diagnosed with type 2 diabetes, 9,404 were diagnosed with prediabetes, 3,559 had gestational diabetes, and 346,659 had neither diabetes nor prediabetes. The 21 most relevant attributes were selected using the RFE (Recursive Attribute Elimination) method (DARST *et al*., 2018).

The constructed MLP model was composed of:

1 dense layer with 128 neurons, ReLU activation
1 dense layer with 128 neurons, ReLU activation, and dropout of 0.5
1 dense layer with 128 neurons, ReLU activation, and dropout of 0.5
1 dense layer with 1 neuron and sigmoid activation
The hyperparameters used were:
Epochs: 100
batch_size: 20
Optimizer: AdamW,
Leaning_rate_Inicial= 0.001

3.2 The Statistical metrics

In this paper, we used Accuracy and the Confusion Matrix (Rascka *et al*., 2017) as classification metrics. Several models were trained with dozens of different hyperparameters. The best model had a training accuracy of 78.63%, validation accuracy of 73.9%, and testing accuracy of 73.8% (Figures 4 and 5).

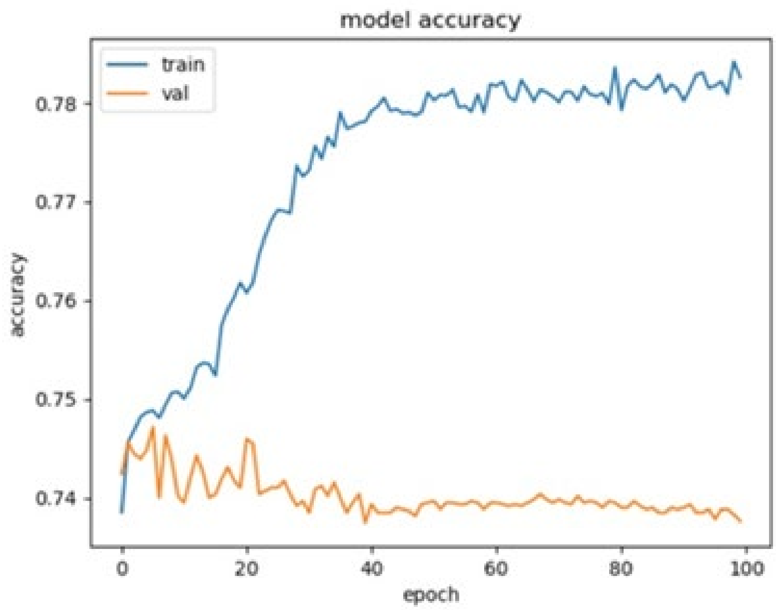

Figure 4: Model accuracy in training and validation - Source: Research data.

Analyzing the Confusion Matrix, we obtained an accuracy of 68.78% for the Normal class (the class with the most data obtained) and 78.72% for the class with diabetes.

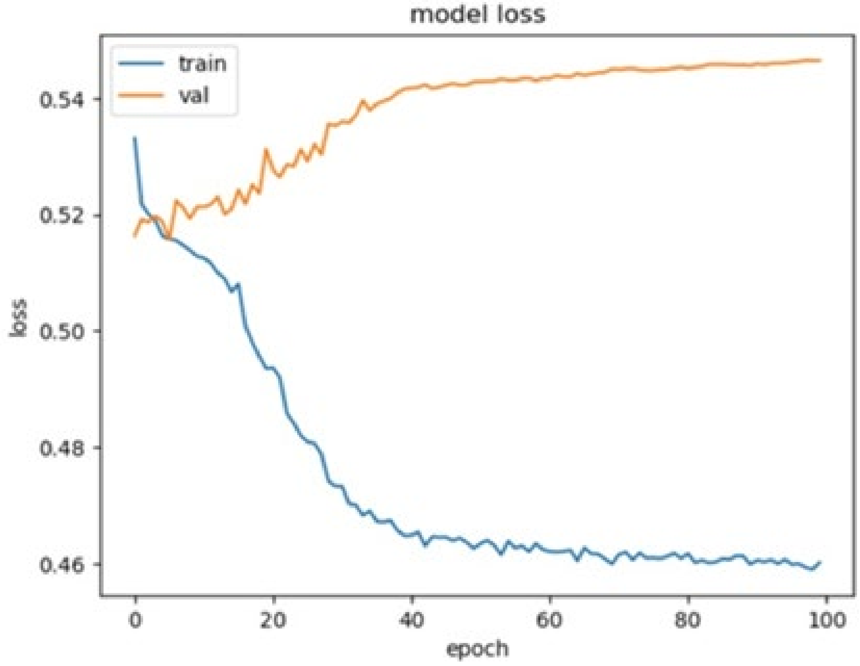

Figure 5: Model error in training and validation - Source: Research data

The MLP (Multilayer Perceptron) model proposed in this study achieved an overall accuracy of 73.8%, with notable performance in classifying individuals with diabetes (78.72%). These results are consistent with previous reports in which traditional machine learning methods, such as Random Forest, XGBoost, Support Vector Machines, and Logistic Regression, achieved accuracies typically ranging from 70% to 80%, depending on data size and quality (Noble *et al*., 2011; Xie *et al*., 2019; Han *et al*., 2022; Kopitar *et al*., 2020). While decision tree–based approaches are valued for their robustness in imbalanced data and ease of interpretability, neural networks stand out for their capacity to capture complex nonlinear dependencies and explore interdependent features (Shahid *et al*., 2019; Zhou, 2021).

Nevertheless, the slightly lower accuracy of the MLP model for the Normal class (68.78%) reflects the challenge of class imbalance, a well-documented limitation in healthcare predictive modeling (Japkowicz & Stephen, 2002; Beam & Kohane, 2018). Strategies such as synthetic oversampling, cost-sensitive learning, or ensemble techniques have been proposed to address this issue and could enhance performance in future implementations (Han *et al*., 2022).

Another central challenge relates to model interpretability. Despite their strong predictive capacity, deep learning models are often described as "black boxes." Explainable Artificial Intelligence (XAI) methods, such as LIME (Ribeiro *et al*., 2016) and SHAP (Lundberg & Lee, 2017), have shown potential to increase clinicians' trust and to transform predictions into actionable insights, particularly in chronic disease management (Miotto *et al*., 2018).

Furthermore, generalizability remains a critical concern. Since the BRFSS dataset reflects the U.S. population, the applicability of these results to other regions and populations may be limited due to genetic, lifestyle, and socioeconomic differences (Pierannunzi *et al*., 2013). Future research should employ multicenter and international datasets, as well as incorporate social and environmental determinants of health, to validate predictive models across diverse contexts.

Finally, the deployment of this predictive system as a web-based application expands its potential for practical use, supporting both healthcare professionals and the general public in early risk assessment. However, the tool should be positioned as a complement rather than a substitute for medical diagnosis, ensuring safe and responsible integration into healthcare systems (Beam & Kohane, 2018).

CONCLUSION

The results obtained in this paper have prominent clinical implications regarding type 2 diabetes prevention and management. Implementing a predictive model such as MLP can aid in the early identification of individuals at risk of developing diabetes, enabling targeted interventions that can significantly decrease the disease's impact on public health systems and individual well-being. Early detection is particularly critical in mitigating the progression of prediabetic conditions into full-blown diabetes, reducing associated complications such as cardiovascular disease, kidney failure, and neuropathy. Our results accentuate the model's ability as an interconnected tool in clinical decision-making processes.

The model was further carried out as web-based software for diabetes risk classification. It is also available to the general public and healthcare professionals. This accessibility links the gap between machine learning models and real-world applications.

Future works could focus on these limitations by enhancing the diversity and representativeness of the training data and consolidating advanced techniques to handle class imbalance. Nonetheless, the results obtained in this study demonstrate the transformative potential of integrating machine learning models in healthcare, particularly in the proactive management of chronic diseases such as type 2 diabetes

CONFLICTS OF INTEREST

Os autores declaram não haver conflitos de interesse.

REFERENCES

AMERICAN DIABETES ASSOCIATION. Economic costs of diabetes in the US in 2017. *Diabetes Care,* v. 41, n. 5, p. 917–928, 2018.

BEAM, A. L.; KOHANE, I. S. Big data and machine learning in health care. *JAMA,* v. 319, n. 13, p. 1317–1318, 2018. doi:10.1001/jama.2017.18391.

CDC – CENTERS FOR DISEASE CONTROL AND PREVENTION. 2021 BRFSS Survey Data and Documentation [Internet]. Disponível em: https://www.cdc.gov/brfss/annual_data/annual_2021.html. Acesso em: 01 maio 2023.

CHO, N. H.; SHAW, J. E.; KARURANGA, S.; HUANG, Y.; DA ROCHA FERNANDES, J. D.; OHLROGGE, A. W. *et al.* IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Research and Clinical Practice,* v. 138, p. 271–281, 2018.

DARST, B. F.; MALECKI, K. C.; ENGELMAN, C. D. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genetics,* v. 19, n. 1, p. 1–6, 2018.

DETTMERS, T. Deep Learning in a Nutshell: Core Concepts [Internet]. *NVIDIA Developer Blog,* 2015. Disponível em: https://devblogs.nvidia.com/deep-learning-nutshell-core-concepts/. Acesso em: 02 maio

2019.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. *Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina.* Rio de Janeiro: Editora LTC, 2011. 394 p.

FUNG, J. *O Código do Diabetes.* São Paulo: Editora nVersos, 2020.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning.* Cambridge, MA: MIT Press, 2017. 776 p.

HAYKIN, S. *Neural Networks: A Comprehensive Foundation.* 2. ed. New Jersey: Prentice Hall, 2000. 842 p.

JAPKOWICZ, N.; STEPHEN, S. The class imbalance problem: A systematic study. *Intelligent Data Analysis,* v. 6, n. 5, p. 429–449, 2002.

KOPITAR, L. *et al*. Early detection of type 2 diabetes mellitus using machine learning-based prediction models. *Scientific Reports*, v. 10, n. 1, p. 11981, 2020.

LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*. p. 4765–4774, 2017.

MENDES, D. *et al*. *Deep Learning Book* [Internet]. Disponível em: http://deeplearningbook.com.br/o-que-sao-redes-neurais-artificiais-profundas/. Acesso em: 01 maio 2023.
MINISTÉRIO DA SAÚDE. VIGITEL BRASIL 2023: Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico. Brasília: Secretaria de Vigilância em Saúde e Ambiente. 2023.
Disponível em: https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes

MIOTTO, R. *et al*. Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics,* v. 19, n. 6, p. 1236–1246, 2018.

MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. *Foundations of Machine Learning.* Cambridge, MA: MIT Press, 2016. 412 p.

NOBLE, D.; MATHUR, R.; DENT, T.; MEADS, C.; GREENHALGH, T. Risk models and scores for type 2 diabetes: systematic review. *BMJ*, v. 343, d7163, 2011.

PIERANNUNZI, C.; HU, S. S.; BALLUZ, L. A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004–2011. *BMC Medical Research Methodology,* v. 13, n. 49, p. 1–14, 2013.

RASCHKA, S.; MIRJALILI, V. *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow.* 2. ed. Birmingham, UK: Packt Publishing, 2017. 622 p.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. "Why should I trust you?": Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. p. 1135–1144, 2016. doi:10.1145/2939672.2939778.

SANTOS, A. L., LIMA, D. R., & CORRÊA, L. A. Regional differences in type 2 diabetes prevalence and risk factors in Brazil: Insights from population-based studies. *Frontiers in Public Health,* 12, 1275167. 2024.
Disponível em: https://www.frontiersin.org/articles/10.3389/fpubh.2024.1275167/full

SAHA, P.; MAROUF, Y.; POZZEBON, H.; GUERGACHI, A.; KESHAVJEE, K.; NOAEEN, M. *et al*. Predicting time to diabetes diagnosis using random survival forests. *medRxiv,* 2024. DOI: 10.1101/2024.02.

SHAHID, N.; RAPPON, T.; BERTA, W. Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PLoS One,* v. 14, n. 2, p. e0212356, 2019.

XIE, Z.; NIKOLAYEVA, O.; LUO, J.; LI, D. Building risk prediction models for type 2 diabetes using machine learning techniques. *Preventing Chronic Disease*, v. 16, p. 190109, 2019. DOI: http://dx.doi.org/10.5888/pcd16.190109.

XIE, P.; XU, J. Prediction of diabetes mellitus using XGBoost model. *Applied Computing Engineering,* v. 67, p. 131–141, 2024.

ZHOU, Z.-H. *Machine Learning.* Springer, 2021.