

Sistema inteligente baseado em conhecimento para acompanhamento de pré-diabéticos

Jacques Amorim¹, Leandro Dias da Silva¹,
Evandro de Barros Costa¹, Álvaro Sobrinho²

¹Instituto de Computação – Universidade Federal de Alagoas (UFAL)
Av. Lourival Melo Mota, S/N, – 57072-970 – Maceió – AL – Brasil

²NITE – Universidade Federal do Agreste de Pernambuco (UFape)
Av. Bom Pastor, s/n, – 55292-270 – Garanhuns – PE – Brasil

{jwsmsga, leandrodias, evandro}@ic.ufal.br, alvaro.alvares@ufape.edu.br

Abstract. *Metabolic diseases have gained more visibility for health researchers because they compromise humanity's lifestyle. Type 2 Diabetes Mellitus is one of those chronic diseases that has become more present in the lives of human beings. Although there is no cure, it is possible to reduce the progression of the disease through the adoption of preventive health care in phases before its full development. This work is an intelligent system to help people who do not have developed conditions or who already have pre-diabetes. The objective is to monitor disease indicators to mitigate, predict, or reverse pre-diabetes. When someone reaches the Diabetes stage, only controlling the disease becomes a solution to avoid possible complications. This intelligent mobile system can mainly benefit people who have difficulty accessing professionals to monitor their health continuously.*

Resumo. *As doenças metabólicas têm ganhado mais visibilidade para pesquisadores da saúde porque comprometem o estilo de vida da humanidade. O Diabetes Mellitus tipo 2 é uma dessas doenças crônicas que se tornou mais presente na vida dos seres humanos. Embora não haja cura, é possível reduzir a progressão da doença por meio da adoção de cuidados de saúde preventivos. Este trabalho apresenta um sistema inteligente para ajudar pessoas que não possuem condições desenvolvidas ou que já possuem pré-diabetes. O objetivo é monitorar indicadores da doença para mitigar, prever ou reverter a pré-diabetes. Quando alguém atinge o estágio de Diabetes, apenas controlar a doença se torna uma solução para evitar possíveis complicações. Este sistema móvel inteligente pode beneficiar principalmente pessoas que possuem dificuldades em acessar profissionais para monitorar sua saúde continuamente.*

1. Introdução

O Diabetes *Mellitus* tipo 2 é uma doença silenciosa, an qual os níveis de açúcar no sangue estão em condições elevadas (hiperglicemia) [Khan et al. 2019]. O pré-diabetes representa um estágio intermediário entre a capacidade normal de tolerância à glicose e o Diabetes de fato [Beulens et al. 2020]. No entanto, diferente da condição de diabético, para qual nenhuma cura completa foi descoberta até o momento, a condição de pré-diabético

possui tratamentos que demonstram uma taxa de sucesso significativa na prevenção de progressão da doença [Khan et al. 2019].

Segundo [Atlas 2021], na América Central e Sul, 32 milhões de adultos vivem com Diabetes, e dentre esses, 33% não são diagnosticados. O Diabetes *mellitus* tipo 2, domina com 90% os casos da doença [Sociedade Brasileira de Diabetes 2023], e por possuir um caráter silencioso, oferece dificuldade para combate prematuro. A presença da doença na vida da população mundial é crescente, ano após ano, e os casos do Diabetes entre adultos na faixa de 20-79 anos podem chegar a mais de 23 milhões até 2045 no Brasil [Internacional Diabetes Federation 2021]. O acompanhamento contínuo por um profissional é essencial para um combate eficiente a doença, no entanto, existem empecilhos no caminho que afetam os pacientes. Estes obstáculos podem ser impostos pela falta de informações sobre a doença e sua real gravidade, dificuldade de acesso ao acompanhamento de um especialista ou até mesmo a negligência do paciente em relação ao seu estado de saúde. Isso conduz um paciente, que ainda esteja no estágio de pré-diabetes, a perder a oportunidade de retardar ou reverter a evolução para o Diabetes.

Muitos sistemas se propõem a auxiliar pacientes no controle do Diabetes, monitorando taxas e fornecendo dicas sobre como manter uma condição de saúde estável, prevenindo assim possíveis complicações. Entretanto, sistemas que envolvem o pré-diabetes ainda foram pouco explorados, indicando a necessidade de soluções para esse campo.

Este trabalho tem o propósito de reimplementar e aprimorar o *Assistente Inteligente para Auxílio na Prevenção do Diabetes tipo 2* [Marinho et al. 2019], visando a adição de novas ferramentas que facilitarão o acompanhamento não só do Diabetes tipo 2, como também o seu estágio anterior, o pré-diabetes, tendo como base a análise dos dados que foram coletados do *CDC – 2015 BRFSS Survey Data and Documentation* [CDC 2015] para extrair novas percepções sobre a doença.

O objetivo principal neste trabalho foi desenvolver um sistema inteligente, que por meio de um aplicativo (*app*) seja capaz de monitorar, avaliar e recomendar dicas de saúde a um paciente, afim de prevenir ou regredir um quadro de pré-diabetes. Além disso, fornecer suporte para evitar o avanço da doença para estágios mais avançados (Diabetes).

As contribuições deste trabalho são: pré-processamento e análise dos dados; implementação do código utilizando algoritmos tradicionais de Aprendizado de Máquina (AM), com o objetivo de criar um modelo preditivo de classificação; desenvolvimento de uma API para integrar o modelo preditivo com o *app*; aprimorar o *app* existente, responsável por captar os dados dos pacientes e partir destes dados, realizar o acompanhamento, indicando quais condições de saúde requerem atenção.

2. Trabalhos Relacionados

Estudos na área de sistemas inteligentes utilizam algoritmos de AM para beneficiar profissionais da medicina e melhorar eficiência dos sistemas de saúde. Por exemplo, [Dinh et al. 2019] utilizou uma abordagem baseado em dados para prever o Diabetes e doenças cardiovasculares. São apresentados resultados ao utilizar modelos de AM supervisionados em um conjunto de dados, onde para o Diabetes, por exemplo, foi possível identificar os cinco melhores preditores: comprimento abdominal, idade, maior peso auto-relatado, comprimento da perna e ingestão de sódio. Isso demonstra o potencial de siste-

mas, nos quais, pela análise de um conjunto de dados, é possível obter percepções sobre características que pacientes diabéticos apresentam.

Outros estudos como [Yamaguchi et al. 2019] e [Waki et al. 2015], aliam as tecnologias de captação de dados presentes em *smartphones* para o desenvolvimento de *apps* capazes de prover suporte aos pacientes portadores do Diabetes tipo 2. As soluções propostas envolvem o auto monitoramento de peso corporal, níveis de açúcar no sangue, atividades físicas representadas por contagem de passos e dieta. Dessa forma, por meio de dispositivos móveis, o acompanhamento de doenças pode ser realizado.

[Kriventsov et al. 2020] coleta dados de dispositivos de medição de glicose. O *app* utiliza modelos preditivos para gerar resultados ao paciente, indicando se é necessário corrigir a ingestão de insulina, alimentos ou outros comportamentos, a fim de evitar possíveis casos de hipoglicemia ou hiperglicemia.

Este trabalho concentrando-se na condição do pré-diabetes pois, como mostrado, os estudos visam, principalmente, o monitoramento e controle de pacientes já diabéticos. Por meio da análise dos dados de pacientes pré-diabéticos, busca-se extrair informações que indicam quais características são mais dominantes. O objetivo é permitir o acompanhamento e diagnóstico aprimorado, contribuindo com a redução na evolução da doença.

3. Funcionalidades do Sistema

Ao ser realizado o primeiro acesso, o *app* conduz um questionário com o objetivo de obter respostas referentes aos atributos selecionados durante fase de análise dos dados. Essas respostas são relevantes para que o sistema, posteriormente, comunique-se com a *API* e gere o relatório do paciente, exibindo detalhes sobre possíveis condições críticas de saúde associadas ao pré-diabetes. A versão anterior do sistema já incluía a captação de resultados de exames *Hemoglobina Glicada*, *Glicose em Jejum*, *Glicose após 75g*, além de informações relacionadas ao peso e à circunferência abdominal, que são suficientes para indicar se o paciente possui o Diabetes tipo 2.

Ao pressionar o ícone do *app*, a comunicação com a *API* é iniciada e os dados previamente coletados são enviados. Assim, o *app* é capaz de gerar um relatório contendo informações sobre a situação atual do paciente em relação ao pré-diabetes, que estão de acordo com os dados fornecidos anteriormente.

A condição de saúde apresentada no relatório é criada por meio de uma árvore de decisão implementada no código do *app*, que leva em consideração a resposta obtida através da *API*, que indica se há indícios de pré-diabetes, assim como os dados coletados no questionário, taxas e medidas. Caso o relatório assinale que o paciente apresenta pré-diabetes ou uma condição normal, porém com indícios de progressão para pré-diabetes, será criada uma seção de detalhes. Nessa seção, serão fornecidas recomendações para melhorar a condição de saúde do paciente, que não esteja dentro das condições normais e que influencie diretamente no possível desenvolvimento ou agravamento da doença.

4. Resultados

Durante a análise dos dados, foram gerados diversos gráficos¹ com base nos 21 atributos do conjunto de dados. O objetivo foi avaliar quais atributos apresentavam correlação

¹https://drive.google.com/drive/folders/1GDw5H4oIfiqmbXNKII-NHyRSsvvMp5Nz?usp=drive_link

Tabela 1. Tabela de métricas de avaliação do primeiro teste

model	accuracy	precision	recall	f1	roc auc
Random Forest	0.858075	0.841145	0.883867	0.861977	0.858002
Decision Tree	0.803915	0.774886	0.858262	0.814446	0.803762
KNN	0.749825	0.718914	0.822723	0.767324	0.749619
SVM	0.727585	0.714624	0.760319	0.736764	0.727493
Gradient Boosting	0.713484	0.705598	0.735413	0.720197	0.713422
Logistic Regression	0.692928	0.697914	0.683364	0.690562	0.692955

relevante entre si e correlação com a variável alvo (não pré-diabético ou pré-diabético). Esses resultados foram relevantes para identificar quais atributos possuem menor impacto no modelo de classificação e também não seriam adequados para inclusão no *app*.

Portanto, foi possível obter percepções iniciais sobre o conjunto de dados. Foi observada uma correlação significativa entre alguns atributos, o que indica que alguns deles podem ser redundantes. Os atributos *Education* e *Income* foram identificados como tais, em que o primeiro representa o nível de escolaridade do paciente e o segundo a escala de renda familiar. Assim, foi analisado qual dos atributos poderia ser mais benéfico ao *app*, pois não seria necessário manter ambos para a implementação do modelo. Como resultado, optou-se por manter o atributo *Education*, visto que seria mais simples para o *app* coletar as informações relacionadas a esse atributo.

Outros atributos com uma correlação significativa foram *PhysHlth* e *DiffWalk*. O primeiro indica quantos dias o paciente apresentou problemas físicos ou lesões no último mês, enquanto o segundo indica se o paciente tinha dificuldades para caminhar. Neste caso, escolheu-se *PhysHlth*, pois as informações sobre o paciente obtidas por esse atributo eram mais úteis para o módulo de dicas do *app*. Além disso, *PhysHlth* possui uma alta correlação com *GenHlth*. Após testes, foi detectado que ambos os atributos contribuíam bastante para o modelo de classificação, então, ambos foram mantidos. Portanto, o número de atributos foi reduzido de 21 para 19.

Outras observações pertinentes estão relacionadas à correlação dos atributos com a variável alvo. As correlações apresentaram índices baixos, no entanto, foi possível identificar quais atributos são mais relevantes para a implementação do modelo. Os cinco que mais se destacaram foram: *GenHlth*, *HighBp*, *HighChol*, *BMI* e *Age*.

Após a conclusão das fases de pré-processamento e análise dos dados, foi executada a etapa de treinamento. Nesta etapa, foi realizada a primeira tentativa de criar um modelo preditivo com desempenho adequado. Essa tentativa levou em consideração os 19 atributos. Os resultados para cada modelo são apresentados na Tabela 1.

Por meio das métricas de avaliação, é possível observar que o modelo que obteve melhores resultados foi o *Random Forest*. A partir desse modelo, foi possível também identificar atributos mais relevantes.

No segundo experimento, foram removidos do teste os atributos com importância abaixo de 0.025. Um atributo de importância elevada que também foi removido foi o "*MentHlth*", pois o valor atribuído indica o nível de saúde mental do paciente durante um determinado mês. No entanto, para o contexto geral do *app*, não seria interessante incluí-lo, já que envolve uma área muito ampla e não haveria uma forma eficiente de coletar

Tabela 2. Tabela de métricas de avaliação do segundo teste

model	accuracy	precision	recall	f1	roc auc
Random Forest	0.853585	0.827805	0.893942	0.859603	0.853471
Decision Tree	0.808545	0.779126	0.862740	0.818804	0.808392
KNN	0.742318	0.706786	0.830698	0.763749	0.742069
SVM	0.696647	0.689539	0.718483	0.703714	0.696585
Gradient Boosting	0.688438	0.679253	0.717364	0.697788	0.688357
Logistic Regression	0.658903	0.663471	0.648804	0.656055	0.658931

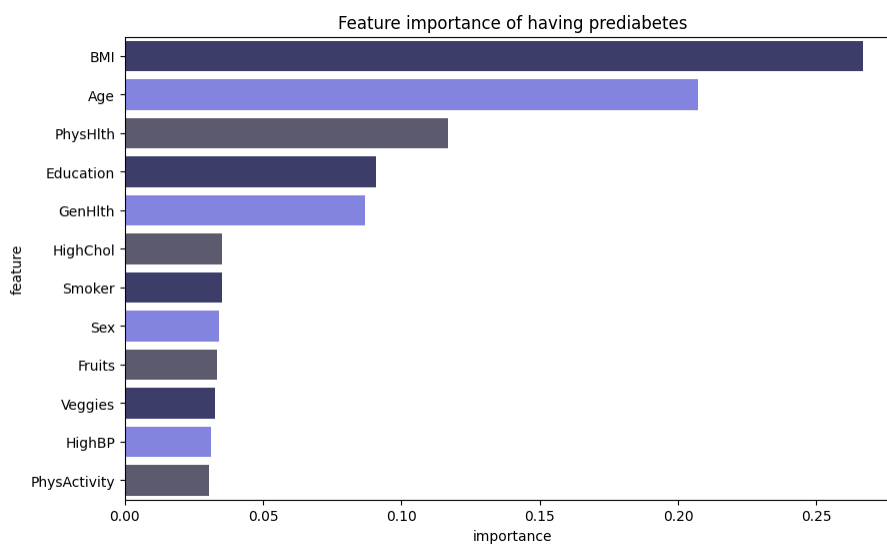


Figura 1. Nível de importância dos atributos para o modelo gerado

esses dados. Logo, o número de atributos foi reduzido de 19 para 12. Na Tabela 2 são apresentados os resultados de avaliação obtidos.

Percebe-se que nesta segunda tentativa, o modelo baseado em *Random Forest* continuou sendo o melhor. No entanto, a métrica de acurácia apresentou uma queda de aproximadamente 0.52%. Essa redução ocorreu devido à remoção do atributo *MentHlth*, mas o impacto para o modelo foi baixo.

A redução de atributos ocasionou facilitou a integração com o *app*, já que as informações relacionadas aos atributos restantes podem ser facilmente coletadas. Além disso, o tempo gasto no processamento dos modelos diminuiu, pois haviam menos dados para se testar.

Na Figura 1 são ilustrados os 12 atributos mais relevantes para diagnosticar o pré-diabetes, com base nos dados usados neste trabalho. Foi possível determinar quais características do paciente são determinantes para indicar a presença da doença ou sinais de um eventual desenvolvimento, auxiliando no monitoramento aprimorado dos pacientes.

5. Conclusão

Neste trabalho, foi apresentado um sistema inteligente baseado em conhecimento, cujo objetivo é realizar o acompanhamento de pacientes em situações de risco de diabetes e pré-diabetes. Com base nos estudos sobre o conjunto de dados apresentado, foi possível desenvolver funcionalidades que aprimoraram a aplicação, permitindo avaliar, fornecer

dicas e recomendações para monitorar e melhorar a saúde do paciente.

Como trabalhos futuros, pretende-se desenvolver uma versão do *app* que também funcione sem conexão com a internet. Isso permitirá que o sistema utilize tanto armazenamento local quanto em nuvem. Em relação às recomendações, propõe-se realizar um estudo específico com profissionais de saúde para ampliar a gama de recomendações geradas pelo *app*. Por fim, é interessante trabalhar em uma possível integração do *app* com dispositivos capazes de medir a glicemia do paciente, recebendo esses dados e possibilitando um diagnóstico em tempo real de uma condição problemática do paciente.

Reconhecimento

Os autores gostariam de agradecer ao apoio financeiro concedido pela Fundação de Amparo à Pesquisa do Estado de Alagoas - FAPEAL.

Referências

- Atlas, I. D. (2021). *IDF Diabetes Atlas 2021*. Internacional Diabetes Federation, 10 edition.
- Beulens, J., Rutters, F., Rydén, L., Schnell, O., Mellbin, L., Hart, H., and Vos, R. (2020). Risk and management of pre-diabetes. *European Journal of Preventive Cardiology*, 26(2):47–54.
- CDC (2015). Behavioral risk factor surveillance system. Disponível em: https://www.cdc.gov/brfss/annual_data/annual_2015.html. Acesso em 20 de setembro de 2023.
- Dinh, A., Miertschin, S., Young, A., and Mohanty, S. D. (2019). A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC Medical Informatics and Decision Making*, 19(1).
- Internacional Diabetes Federation (2021). Brazil: Diabetes report 2000-2045. <https://diabetesatlas.org/data/en/country/27/br.htmlr>.
- Khan, R. M. M., Chua, Z. J. Y., Tan, J. C., Yang, Y., Liao, Z., and Zhao, Y. (2019). From pre-diabetes to diabetes: Diagnosis, treatments and translational research. *Medicina*, 55(9).
- Kriventsov, S., Lindsey, A., and Hayeri, A. (2020). The diabits app for smartphone-assisted predictive monitoring of glycemia in patients with diabetes: Retrospective observational study. *JMIR Diabetes*, 5(3):e18660.
- Marinho, T., Filho, W. F., Freire, G., Costa, M., da Silva, L., Álvaro C. Sobrinho, and Costa, E. (2019). Assistente inteligente para auxiliar na prevenção do diabetes tipo 2. In *Anais Estendidos do XIX Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 121–126, Porto Alegre, RS, Brasil. SBC.
- Sociedade Brasileira de Diabetes (2023). Diabetes. <https://diabetes.org.br>.
- Waki, K., Aizawa, K., Kato, S., Fujita, H., Lee, H., Kobayashi, H., Ogawa, M., Mouri, K., Kadowaki, T., and Ohe, K. (2015). Dialbetics with a multimedia food recording tool, foodlog: Smartphone-based self-management for type 2 diabetes. *Journal of diabetes science and technology*, 9(3):534—540.
- Yamaguchi, S., Waki, K., Nannya, Y., Nangaku, M., Kadowaki, T., and Ohe, K. (2019). Usage patterns of gluconote, a self-management smartphone app, based on researchkit for patients with type 2 diabetes and prediabetes. *JMIR Mhealth Uhealth*, 7(4):e13204.