



Técnicas de aprendizado de máquinas para predição dos níveis de glicose em pacientes com Diabetes Tipo 1

Machine learning techniques for predicting glucose levels in patients with Type 1 Diabetes

João Pedro Costa ¹, Arthur Bertachi ²

RESUMO

O gerenciamento da glicemia representa um desafio significativo para aqueles que convivem com Diabetes Mellitus Tipo 1 (DMT1). Manter os níveis de glicose sanguínea dentro da faixa normal é uma tarefa complexa que exige dedicação contínua ao longo da vida. Felizmente, avanços tecnológicos como o monitoramento contínuo da glicose, o uso de bombas de insulina e aplicativos possibilitaram um acompanhamento mais próximo de diversos indicadores relacionados ao controle do diabetes. Isso resultou no crescente uso de algoritmos de aprendizado de máquina para padrões sobre o comportamento dos níveis glicêmicos dos pacientes. Neste estudo, avaliou-se a eficácia de duas técnicas de aprendizado de máquina, a Multilayer Perceptron e o Gradient Boosting, na previsão dos níveis futuros de glicose sanguínea. Foi utilizado o *dataset* "OHIO T1DM", que é composto de dados de doze pacientes distintos. Os resultados numéricos demonstram a capacidade de ambos os métodos em fornecer informações sobre os futuros níveis de glicose sanguínea aos pacientes, realçando seu potencial para auxiliar as pessoas com DMT1 no controle de sua condição.

PALAVRAS-CHAVE: Aprendizado de Máquinas; Diabetes Mellitus Tipo 1; Hipoclicemia; Predição de Glicose Sanguínea.

ABSTRACT

Managing blood glucose presents a significant challenge for individuals living with Type 1 Diabetes Mellitus (T1DM). Maintaining blood glucose levels within the normal range is a complex task that requires ongoing dedication throughout life. Fortunately, technological advances such as continuous glucose monitoring, insulin pumps, and mobile applications have enabled closer tracking of various indicators related to diabetes management. This has led to the growing use of machine learning algorithms to discern patterns in patients' blood glucose behavior. In this study, we examined the effectiveness of two machine learning techniques, Multilayer Perceptron and Gradient Boosting, in predicting future blood glucose levels. We considered the "OHIO T1DM" dataset, which incorporates data from twelve different patients. Numeric results demonstrate the ability of both methods to provide insights into future blood glucose levels for patients, highlighting their potential to assist individuals with T1DM in managing their condition.

KEYWORDS: Machine Learning; Type 1 Diabetes Mellitus; Hypoglycemia; Blood Glucose Prediction.

INTRODUÇÃO

A Diabetes Mellitus Tipo 1 (DMT1) é uma doença autoimune que impacta diretamente as células β do pâncreas, responsáveis pela liberação de insulina na corrente sanguínea. Indivíduos

¹ Bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq. Universidade Tecnológica Federal do Paraná, Curitiba, Paraná, Brasil. E-mail: joapedrocosta@alunos.utfpr.edu.br. ID Lattes: 4917071341762406.

² Docente no Curso de Engenharia Mecatrônica. Universidade Tecnológica Federal do Paraná, Guarapuava, Paraná, Brasil. E-mail: abertachi@utfpr.edu.br. ID Lattes: 2283634850801378.



XIII Seminário de Extensão e Inovação XXVIII Seminário de Iniciação Científica e Tecnológica da UTFPR

Ciência e Tecnologia na era da Inteligência Artificial: Desdobramentos no Ensino Pesquisa e Extensão
20 a 23 de novembro de 2023 - Campus Ponta Grossa, PR



SEI-SICITE
2023

com essa condição devem realizar a reposição de insulina de origem externa para evitar o aumento excessivo da concentração de glicose no sangue, conhecido como hiperglicemia. Entretanto, caso a quantidade de insulina injetada seja maior do que a necessária, os níveis de glicose podem diminuir para valores abaixo do normal, levando à hipoglicemia. Ambas as situações são indesejáveis e podem resultar em complicações a curto e longo prazo (LIND et al., 2014).

Pacientes com DMT1 utilizam doses de insulina para manter os níveis de glicose dentro de uma faixa desejável. A dosagem de insulina pode ser realizada utilizando múltiplas injeções diárias (*multiple daily injections* - MDI) ou utilizando uma bomba de insulina. Entretanto, a grande variabilidade observada na ação da insulina faz com que o cálculo da dosagem ótima de insulina seja bastante impreciso (HOVORKA, 2006).

Para estimar a quantidade de insulina a ser injetada, é necessário também ter conhecimento do nível glicêmico do paciente. A tecnologia envolvida na medição de glicose teve uma grande evolução durante os últimos 20 anos. Atualmente, as duas maneiras mais utilizadas para mensurar a glicose sanguínea são: 1) medições realizadas manualmente através de glicosímetros; 2) utilizando o monitoramento contínuo de glicose (*Continuous glucose monitoring* - CGM) através de sensores eletroquímicos que são inseridos na camada subcutânea da pele (BERTACHI et al., 2018).

A integração dos dispositivos tecnológicos utilizados na gestão da DMT1 tornou capaz a armazenagem dos dados de diversas variáveis envolvidas na gestão da DMT1, tais como: quantidade de insulina injetada, leituras de glicose sanguínea, carboidratos consumidos, níveis de atividade física, etc. Considerando todo o avanço tecnológico disponível nos dispositivos digitais atualmente, é viável a aplicação de sistemas inteligentes embarcados tanto no *firmware* do PA ou até mesmo em celulares, através do desenvolvimento de aplicativos voltados para a gestão da DMT1. Em diversos campos da ciência, algoritmos inteligentes orientados a dados (*data driven*) são aplicados para auxiliar os usuários nas tomadas de decisão, inclusive na área de saúde (WOLFF et al., 2021; TRIANTAFYLLIDIS; TSANAS, 2019).

A aplicação de técnicas de inteligência artificial (*artificial intelligence* - AI) para problemas relacionados a DMT1 vem aumentando nos últimos anos (CONTRERAS; VEHI, 2018). Dentre as diversas possibilidades de aplicação, a maioria das abordagens utilizam modelos preditivos individualizados, seja para regressão ou classificação. Entretanto, uma grande desvantagem deste tipo de abordagem é a necessidade da coleta de uma grande quantidade de dados para ser capaz de treinar e validar os modelos. Como se sabe, a etapa de coleta de dados possui um custo financeiro elevado, o que pode ser uma barreira para a transferência de tecnologia entre a pesquisa científica e a aplicação final no dia a dia dos pacientes.

O objetivo deste estudo é avaliar o desempenho de técnicas de *machine learning* na previsão dos níveis de glicose em pacientes com DMT1, considerando um horizonte de predição de 30 minutos. Para essa finalidade, é utilizado um banco de dados que contém informações de 12 pacientes diagnosticados com essa condição, coletados durante a rotina de vida tradicional de cada paciente.



MATERIAIS E MÉTODOS

BANCO DE DADOS OHIO T1DM

O conjunto de dados Ohio T1DM está disponível para pesquisadores interessados investigar novas técnicas que possam trazer benefícios para a saúde e o bem-estar de pessoas com DMT1. Ele contém dados de 8 semanas para cada um dos 12 indivíduos com DMT1 (MARLING; BUNESCU, 2020). Todos esses indivíduos estavam em terapia com bomba de insulina, com monitoramento contínuo de glicose (*Continuous Glucose Monitoring* - CGM). Estão disponíveis dados de glicose no sangue, dados de insulina, dados de eventos da vida autorrelatados e dados de pulseiras de atividades físicas. O conjunto de dados inclui: níveis de glicose no sangue a cada 5 minutos pelo CGM; níveis de glicose no sangue a partir de monitoramento periódico de glicose no sangue (coletados a partir de um glicosímetro); doses de insulina: tanto insulina de bolus quanto insulina basal; horários de refeições autorrelatados com estimativas de carboidratos consumidos; horários de atividades físicas, períodos de sono, trabalho, estresse e doença autorrelatados; além dos dados fornecidos pela pulseira de atividade física como: quantidade de passos, ritmo cardíaco, resposta galvânica da pele e outros. Um total de 20 atributos estão disponíveis para cada paciente com um intervalo amostral de 5 minutos.

PROCESSAMENTO DOS DADOS

Para preparar os dados brutos e torná-los adequados para o treinamento de algoritmos de *machine learning*, é necessário realizar o processamento dos dados. Toda a etapa de processamento dos dados, incluindo o treinamento e a validação dos modelos de *machine learning*, foi executada utilizando a linguagem de programação Python. No que diz respeito ao treinamento dos modelos, foi utilizada a biblioteca *Scikit-learn* (PEDREGOSA et al., 2011) que é amplamente utilizada para esse objetivo entre desenvolvedores de modelos preditivos.

O processamento dos dados envolve tarefas como limpeza, normalização e transformação dos dados para torná-los adequados para análise. Como o trabalho visa realizar previsões de valores de glicose no sangue no futuro, deve-se também informar ao algoritmo qual o valor desejado. Considera-se que t_0 o instante atual, e deseja-se prever o valor de glicose 30 minutos a frente (t_{0+30}). Os seguintes atributos foram considerados como entrada para os modelos: Valor de glicose 25 minutos antes de t_0 ; 4 variações de glicose em intervalos de 5 minutos, iniciando-se em t_{0-25} até t_0 ; Hora do dia em t_0 (*timestamp*); Intervalo de tempo entre a última refeição e t_0 ; Estimativa de carboidratos da última refeição; Taxa de insulina injetada durante os 25 minutos antes de t_0 ; Valor médio do nível de glicose no intervalo 90 minutos antes de t_0 . Para gerar modelos de previsão individualizados, todo o processo de processamento de dados foi realizado de forma individualizada para cada paciente.



ALGORITMOS DE PREDIÇÃO E HIPERPARÂMETROS

Foram selecionados dois algoritmos para obter as predições de glicose e validar a proposta do trabalho: 1) *Multi-layer Perceptron* (MLP) e 2) *Gradient Boosting Regressor* (GBR). O MLP é uma arquitetura de rede neural artificial composta por várias camadas de neurônios. Cada neurônio na camada recebe informações ponderadas das camadas anteriores e passa os resultados por uma função de ativação. O MLP é amplamente utilizado para tarefas de regressão, onde busca aprender padrões e relações complexas entre os atributos selecionados para serem as entradas do algoritmo e o valor desejado na saída (HAYKIN, 2009). Já o GBR é um algoritmo de aprendizado de máquina baseado em árvores de decisão que utiliza o conceito de *boosting*. Para melhorar gradualmente a precisão do modelo, ele combina várias árvores de decisão fracas em um modelo mais forte. Cada nova árvore de decisão é ajustada aos resíduos do modelo anterior (FRIEDMAN, 2001).

A seleção dos hiper parâmetros de ambos os algoritmos foi realizada através de um *grid search*. Inicialmente foi criada uma combinação dos possíveis hiperparâmetros para cada um dos algoritmos. Foram criados um total de 2000 e 3000 possíveis combinações de hiperparâmetros para o GBR e MLP, respectivamente. Por questões de processamento computacional, foram escolhidas de maneira aleatório 300 combinações para cada algoritmo. Para evitar problemas de *overfitting*, o método *k-fold cross validation* ($k=5$) foi aplicado. A escolha do melhor modelo foi feita baseado no RMSE alcançado durante a fase de validação. Este modelo foi utilizado nos dados previamente separados para serem utilizados na fase de teste, ou seja, dados que em nenhum momento foram apresentados ao algoritmo durante as fases de treinamento ou validação.

RESULTADOS

A Tabela 1 apresenta os resultados individualizados para cada um dos 12 pacientes, através da métrica RMSE. Esta métrica permite uma comparação direta com os resultados medidos, uma vez que apresenta a mesma unidade de medida da glicose no sangue. Para ilustrar o comportamento das predições, apresenta-se na Figura 1 a comparação dos valores medidos pelo CGM com os métodos de predição. Foi selecionado um dia do conjunto de teste do paciente #544 que apresentou os melhores resultados de RMSE.

Os resultados obtidos neste estudo estão consistentes com outras publicações que também utilizaram a mesma base de dados considerando o mesmo horizonte de predição. O trabalho apresentado por Martinsson et al. (2018) utilizou uma rede neural conhecida como *Long Short Term Memory* e alcançou um erro médio quadrático de 20,1 mg/dl. Já o trabalho de Zhu et al. (2018) utilizou técnicas de *Deep Learning*, alcançando um RMSE médio de 22,17 mg/dl.

CONSIDERAÇÕES FINAIS

É importante ressaltar que o uso de modelos de *machine learning* para predição de glicose no sangue ainda está em constante evolução e requer validação cuidadosa e supervisão médica antes de serem disponibilizados para o uso por parte dos pacientes. No entanto, essa abordagem promissora pode melhorar



XIII Seminário de Extensão e Inovação XXVIII Seminário de Iniciação Científica e Tecnológica da UTFPR

Ciência e Tecnologia na era da Inteligência Artificial: Desdobramentos no Ensino Pesquisa e Extensão
20 a 23 de novembro de 2023 - Campus Ponta Grossa, PR



SEI-SICITE
2023

Tabela 1 – Resultados da métrica RMSE para todos os pacientes

Paciente ID	MLP Val.	MLP Test	GBR Val.	GBR Test
540	20.14	22.86	17.87	27.38
544	16.06	17.24	13.62	16.42
552	18.84	18.06	15.66	18.35
559	23.87	19.53	20.61	19.10
563	15.86	19.00	14.46	21.36
567	20.68	20.57	20.67	20.90
570	17.00	17.31	14.81	17.57
575	17.14	28.87	17.10	23.61
584	21.81	27.00	23.49	23.71
588	17.19	22.45	18.45	19.08
591	23.25	23.03	19.61	21.05
596	15.36	16.98	14.71	16.81
Média	18.78	21.12	17.31	20.41

Fonte: Autoria própria

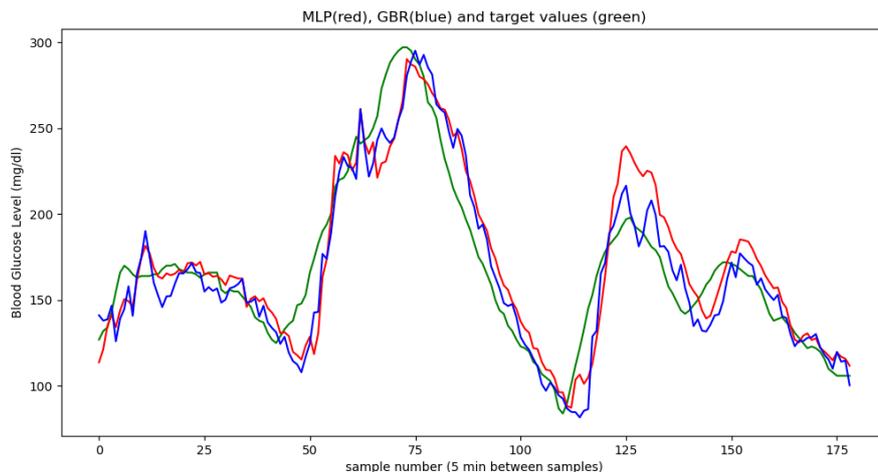


Figura 1 – Comparação entre os modelos preditivos para o paciente #544: MLP (vermelho), GBR (azul) e valores medidos pelo CGM (Verde).

significativamente a qualidade de vida e o auto gerenciamento de pessoas com diabetes.

A utilização de modelos individualizados para previsão de glicose no sangue apresenta alguns pontos negativos a serem considerados. Embora esses modelos sejam capazes de levar em conta características e dados específicos de cada indivíduo. Porém existe a necessidade de coletar uma quantidade significativa de dados pessoais de cada paciente, incluindo histórico glicêmico detalhado, informações sobre alimentação, exercícios e administração de insulina. Este processo de coleta de dados acarreta em um alto custo financeiro e demanda de tempo considerável. Futuros trabalhos a serem conduzidos irão abordar a obtenção de modelos populacionais e comparar os resultados com os modelos individualizados obtidos no presente trabalho.



AGRADECIMENTOS

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq pela bolsa de iniciação científica. Os autores também agradecem a UTFPR pelo apoio à pesquisa.

REFERÊNCIAS

- BERTACHI, Arthur et al. Automated blood glucose control in type 1 diabetes: A review of progress and challenges. **Endocrinología, Diabetes y Nutrición**, v. 65, n. 3, p. 172–181, 2018. DOI: [10.1016/j.endinu.2017.10.011](https://doi.org/10.1016/j.endinu.2017.10.011). Disponível em: [↗](#).
- CONTRERAS, Ivan; VEHI, Josep. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. **Journal of Medical Internet Research**, JMIR Publications Inc., v. 20, n. 5, e10775, mai. 2018. DOI: [10.2196/10775](https://doi.org/10.2196/10775). Disponível em: [↗](#).
- FRIEDMAN, Jerome H. Greedy function approximation: A gradient boosting machine. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 29, n. 5, out. 2001. DOI: [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451). Disponível em: [↗](#).
- HAYKIN, S.S. **Neural Networks and Learning Machines**. [S.l.]: Pearson, 2009. (Pearson International Edition). ISBN 9780131293762. Disponível em: [↗](#).
- HOVORKA, R. Continuous glucose monitoring and closed-loop systems. **Diabetic Medicine**, Blackwell Science Ltd, v. 23, n. 1, p. 1–12, 2006. ISSN 1464-5491. DOI: [10.1111/j.1464-5491.2005.01672.x](https://doi.org/10.1111/j.1464-5491.2005.01672.x). Disponível em: [↗](#).
- LIND, Marcus et al. Glycemic Control and Excess Mortality in Type 1 Diabetes. **New England Journal of Medicine**, v. 371, n. 21, p. 1972–1982, 2014. PMID: 25409370. DOI: [10.1056/NEJMoa1408214](https://doi.org/10.1056/NEJMoa1408214). eprint: <http://dx.doi.org/10.1056/NEJMoa1408214>. Disponível em: [↗](#).
- MARLING, Cindy; BUNESCU, Razvan. The OhioT1DM dataset for Blood Glucose Level Prediction: Update 2020. en. **CEUR Workshop Proc.**, v. 2675, p. 71–74, set. 2020.
- MARTINSSON, John et al. Automatic blood glucose prediction with confidence using recurrent neural networks. en. **CEUR Workshop Proc.**, v. 2148, p. 64–68, set. 2018.
- PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.
- TRIANAFYLLIDIS, Andreas K; TSANAS, Athanasios. Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature. **Journal of Medical Internet Research**, JMIR Publications Inc., v. 21, n. 4, e12286, abr. 2019. DOI: [10.2196/12286](https://doi.org/10.2196/12286). Disponível em: [↗](#).
- WOLFF, Justus et al. Success Factors of Artificial Intelligence Implementation in Healthcare. **Frontiers in Digital Health**, Frontiers Media SA, v. 3, jun. 2021. DOI: [10.3389/fdgth.2021.594971](https://doi.org/10.3389/fdgth.2021.594971). Disponível em: [↗](#).
- ZHU, Taiyu et al. A Deep Learning Algorithm For Personalized Blood Glucose Prediction. en. **CEUR Workshop Proc.**, v. 2148, p. 74–78, set. 2018.